Stefan Bechtold
Indra Spiecker gen. Döhmann [Hrsg.]

Law, Behavior, and Decision
Recht, Verhalten und Entscheidung

Festschrift zum 70. Geburtstag von Christoph Engel



Nomos

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access - (cc) BY

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Stefan Bechtold
Indra Spiecker gen. Döhmann [Hrsg.]

Law, Behavior, and Decision
Recht, Verhalten und Entscheidung

Festschrift zum 70. Geburtstag von Christoph Engel

Nomos

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

1. Auflage 2026

© Die Autor:innen

Publiziert von
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

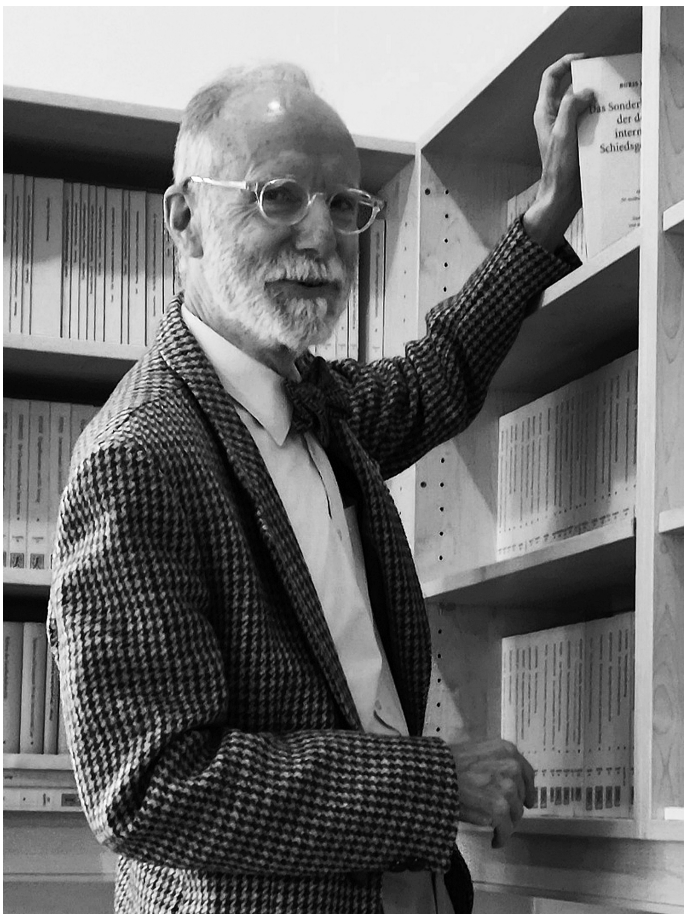
ISBN (Print): 978-3-7560-0117-0
ISBN (ePDF): 978-3-7489-4725-7

DOI: https://doi.org/10.5771/9783748947257

Onlineversion
Inlibra

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -



https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table of Contents

Preface 13

I. On Christoph Engel

Stefan Bechtold & Indra Spiecker gen. Döhmann
Christoph Engel, Collective Goods, and the Institute 17

Hanjo Hamann
„Der schreibt auch so."
Die persönliche Handschrift des Ge(l)ehrten im Spiegel seiner Auftaktsätze 31

Svenja Hippel & Dorothee Mischkowski
Cooperation in Social Dilemmas.
28 Years of Research on Collective Goods 37

Hanjo Hamann
Of Festschrifts and Preprints.
Exploring Understudied Academic Genres 49

Jeffrey J. Rachlinski
The First 67

Renate Buijze & Sigrid Hemels
On Generosity.
Our Wonderful Journey with Christoph Engel 79

II. Experiments

Daniel S. Nagin
Using Experiments to Test Criminological Theory 93

Crystal Reeck, Hale A. Forster & Elke U. Weber
Assessing How We Decide.
Psychometric Development of the Decision Modes Scale 101

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Anna Baumert & Julia Sasse
The Roles of Anger in Moral Courage 119

Marina Chugunova, Andreas Nicklisch & Kai-Uwe Schnapp
On the Effects of Transparency on Labour Supply
in Redistribution Systems 133

Sebastian J. Goerg, Sebastian Kube & Michael Kurschilgen
The Engel Effect?
Experimental Evidence on Rule Compliance 157

Heike Hennig-Schmidt & Gari Walkowitz
The (Im)Moral Man?
An Experimental Study on Hypocrisy and Gender
in Dictator-Game Giving 177

Claudia Moellers & Hans-Theo Normann
Vertical Mergers and Price Leadership.
An Experiment 193

Azzura Morreale & Luigi Mittone
Institutions, Experiments, and Tax Morale.
An Integrated Perspective on Tax Compliance 221

Peter G. Moffatt
An Application of the Panel Hurdle Model to Experimental Data 227

Lilia Wasserka-Zhurakhovska
Anticipated Third-Party Rewards.
Experimental Evidence on Social Control and Cooperation 243

Oliver Kirchkamp
Public Goods, Teams and Institutions 261

III. Judges and Decision-Making

Markus Englerth
„Catch Me If You Can".
Die Flucht als Straftat im Auslieferungsverfahren 281

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Daniel M. Klerman
Are Trump Judges Different?
Evidence from Immigration Cases 299

Keren Weinshall
Judging Theories.
A Survey in Honor of Christoph Engel 311

Patrick C. Wu, Chang-Ching Lin & Yun-chien Chang
Familiarity Breeds Contempt or Deference?
Homage to Christoph Engel 325

Niels Petersen
Die Gleichheitsrechtsprechung des Bundesverfassungsgerichts.
Eine empirische Analyse 341

Emanuel V. Towfigh
Die „gewagte Formel".
Plädoyer für eine statistisch informierte Gleichheitsdogmatik 355

IV. Institutions

Florian Becker
Das „Netz der Behörden" als Netzwerk? 373

Konstantin Chatziathanasiou & Christian Traxler
Governing the Asphalt Commons.
Recht und Akzeptanz der Parkraumbepreisung 397

Stefanie Egidy
Staatliche Opportunität.
Kontrollmodi der Wirtschaftsaufsicht 411

Alexander Morell
„Heterogenität als Problem für Steuerung", Einzelfallgerechtigkeit
und das Referenzklassenproblem im Kartellschadensersatzrecht 427

Stefan Okruch
Rechtsinnovationen über Grenzen.
Der lange Weg zum Europäischen Verbund für Territoriale
Zusammenarbeit 441

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Jörn Lüdemann
Der Rundfunkbeitrag.
Rechts- und verhaltenswissenschaftliche Perspektiven auf ein umstrittenes Finanzierungsmodell 453

Morten Broberg
Using Behavioural Insights When Designing Regulatory Measures 469

Klaus Heine
Datenökonomie.
Zwischen ordnungspolitischer Inversion und staatlicher Willkür 481

Armin Steinbach
Verantwortung in Politik und Beratung 497

V. Machine Learning and the Law

Helena Cecilie Lyng-Olsen & Daniel E. Ho
AI in the Judiciary.
Practice and Policy 509

Elliott Ash
AI Alignment and Legal Reasoning 525

Oren Bar-Gill & Florencia Marotta-Wurgler
Price Discrimination by LLMs 541

Jens Frankenreiter & Michael A. Livermore
Large Language Models in Legal Analysis 561

Johannes Kruse & Pascal Langenbach
LLM-gestützte Simulationen in der rechtsempirischen Forschung.
Ein Werkstattbericht 573

Jonah B. Gelbach
Can LLMs Replicate Surveys in Empirical Legal Interpretation? 601

Yoan Hermstrüwer & Alexander Stremitzer
Verfahrensfairness ohne Menschen?
Evidenz aus Human- und LLM-Experimenten 617

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

VI. Theory

Eric Helland
The Coase Theorem and Divorce 637

Laurence O'Hara
Establishing the Individualistic Viewpoint within Instrumental Administrative Analysis 671

Frederike Zufall
Data as Facts 683

Werner Güth & Hartmut Kliemt
Rational Choice Modeling of Sustainable Corporate Social Responsibility 691

Anne van Aaken
Bridging Minds and Practices.
Integrating Cognitive Psychology and Practice Theory in Legal Reasoning 711

Eyal Zamir
Towards an Integrative Legal Research 727

Axel Ockenfels & Carl Christian von Weizsäcker
Democracy and Paternalism 745

Susann Fiedler & Andreas Glöckner
The Engel Method.
Psychological Insights and Practical Guidelines for Interdisciplinary Research Collaboration 757

VII. Appendices

Christoph Engel: List of Publications 777

List of Authors 801

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Preface

This Festschrift – a format particular to German legal academia – unites forty-seven chapters by seventy researchers whose academic careers have been influenced or shaped by a single scholar. It is organized around five core themes that have accompanied him throughout his academic journey: experimental law & economics, judges & judicial decision-making, institutions, machine learning & the law, and legal theory. The chapters in this volume draw on doctrinal, theoretical, experimental, empirical, and natural-language processing methods. They address topics in constitutional law, administrative law, tax law, competition law, media law, AI & law, European law, procedural law, criminal law, immigration law, corporate law, and family law, among others. The Symposium held in Bonn in honor of this scholar on May 29, 2026, similarly reflects the profound impact he has had on scholarship in law, economics, psychology, and beyond.[1] In its range of subjects and methods, the Festschrift – and the accompanying Symposium – come thereby close to matching the breadth of the scholar to whom it is dedicated: Christoph Engel.

The authors have contributed generously and shown great patience throughout the process. Johannes Rux, Edward Schwarzenberger, and their team at Nomos provided a smooth and fitting venue for publication. Christina von Wintzingerode brought structural clarity to the editing and organization of the papers. Thank you all for this remarkable book!

Stefan Bechtold & Indra Spiecker gen. Döhmann

1 At the Symposium on "Law, Behavior, and Decision", fellow Max Planck Co-director Martin Hellwig gave a keynote lecture on "What Does it Mean to Say 'Jurisprudence is Concerned with the Study of Legal Norms'?"; co-author Oren Bar-Gill (NYU) gave a keynote lecture on "Price Discrimination by LLMs: Following in the Footsteps of the Great Christoph Engel"; and co-author Elke Weber (Princeton) gave a keynote lecture on "'A Man for all Seasons': Christoph Engel's Contributions to Understanding and Utilizing the Complex Connections between Institutional Design and Human Modes of Decision Making." The following colleagues, co-authors and "Schüler" participated in a panel discussion on Law, Behavior, and Decision: Yun-chien Chang, Cornell University; Stefanie Egidy, University of Mannheim; Dan Klerman, University of Southern California; Florencia Marotta-Wurgler, New York University; Daniel Nagin, Carnegie Mellon University; Emanuel Towfigh, EBS Law School; Keren Weinshall, Hebrew University Jerusalem; Eyal Zamir, Hebrew University Jerusalem.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I.
On Christoph Engel

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Christoph Engel, Collective Goods, and the Institute

Stefan Bechtold & Indra Spiecker gen. Döhmann

Christoph Engel's sentences are short. The same cannot be said of the range of his interests. His intellectual horizon seems almost boundless. He embodies the ideal of a Max Planck Institute Director in the fullest sense of the term. Christoph Engel has been writing as a scholar for more than forty years. This offers an opportunity to approach the phenomenon that is Christoph Engel in four dimensions: The Researcher. The Institution Builder. The Mentor. The Original.

I. The Researcher

Nowadays, Christoph Engel is best known for his contributions to experimental law & economics, empirical legal studies, law & psychology, and – more recently – artificial intelligence & the law. But one should not forget where Christoph Engel comes from as a scholar. Christoph Engel was born into an academic household. His father, who died at the age of 56, was a history professor at the University of Tübingen. Christoph Engel started his academic career along the traditional paths of German legal scholarship, including a year of study at the University of Geneva.

Christoph Engel started his publishing career in public law. In 1985, he analyzed the rules of public international law that govern cross-border telecommunications, using direct satellite television as his point of departure. A central concern of the article was to reconcile territorial sovereignty, freedom of communication and information, and international law in an age of emerging satellite broadcasting.[1] It marked the start of Christoph Engel's more than decade-long journey into the world of public international law (also the topic of his doctoral dissertation),[2]

1 Christoph Engel, Das Völkerrecht des Telekommunikationsvorgangs, Rabels Zeitschrift 1985, 90.

2 Christoph Engel, Völkerrecht als Tatbestandsmerkmal deutscher Normen, Berlin 1989.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

broadcasting law (also the topic of his Habilitation),[3] and telecommunications law (the topic of a major research project led by his academic mentor, Ernst-Joachim Mestmäcker). When Christoph Engel became Head of the Max Planck Project Group for "Law of Common Goods" in the mid 1990s, he shifted his focus from broadcasting and telecommunications law to environmental and waste-management law, which would become a major focus of his research for a number of years.[4]

This is not to say that Christoph Engel did not also explore issues well beyond what a typical German public law scholar would do. He wrote about buyout clauses in corporate shareholder agreements, which serve to preserve liquidity and to discipline shareholders, but may threaten the interests of the departing shareholders or their creditors.[5] He analyzed constitutional, legislative, and referendum requirements for changes to German territory – less than a year before the Berlin Wall fell.[6] And he did write about, well, interim measures against offenders who commit traffic offences without a valid driver's license.[7] Looking at Christoph Engel's entire *œuvre*, there are probably fewer areas of law he has *not* written about than there are areas he has written about. Just consider that, in addition to the areas mentioned above, Christoph Engel has also written about constitutional law,[8]

3 Christoph Engel, Privater Rundfunk vor der Europäischen Menschenrechtskonvention, Baden-Baden 1993.

4 See, e.g., Christoph Engel, Abfallrecht und Abfallpolitik, Baden-Baden 2002; Gemischtwirtschaftliche Abfallentsorgung. Ein Lehrstück zur Verschränkung des Abfallrechts mit Kartellrecht, Kommunalrecht, dem Recht der öffentlichen Aufträge und Wirtschaftsverfassungsrecht, Cologne 1995; Marco Verweij et al., Clumsy Solutions for a Complex World. The Case of Climate Change, in: Public Administration 84 (2006) 817.

5 Christoph Engel, Abfindungsklauseln – Eine systematische Übersicht, Neue Juristische Wochenschrift 1986, 345.

6 Christoph Engel, Verfassungs-, Gesetzes- und Referendumsvorbehalt für Änderungen des Bundesgebiets und andere gebietsbezogene Akte, Archiv des öffentlichen Rechts 114 (1989), 46.

7 Christoph Engel, Vorläufige Maßnahmen gegen Täter von Verkehrsdelikten, die nicht im Besitz einer gültigen Fahrerlaubnis sind, Deutsches Autorecht 1984, 108.

8 See, e.g., Christoph Engel, Delineating the Proper Scope of Government: A Proper Task for a Constitutional Court?, in: Journal of Institutional and Theoretical Economics 157 (2001), 187; Christoph Engel, The European Charter of Fundamental Rights. A Changed Political Opportunity Structure and its Normative Consequences, in: European Law Journal 7 (2001), 151; Christoph Engel, The Constitutional Court – Applying the Proportionality Principle – as a Subsidiary Authority for the Assessment of Political Outcomes, in: Christoph Engel & Adrienne Héritier (ed.), Linking

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

administrative law,[9] European law,[10] Internet law,[11] AI & law,[12] competition law,[13] intellectual property law,[14] privacy law,[15] criminal law and crim-

Politics and Law, Baden-Baden 2003, p. 285; Christoph Engel, Is the German Constitutional Court Partisan?, Journal of Legal Studies, forthcoming 2026.

9 See, e.g., Christoph Engel, Planungssicherheit für Unternehmen durch Verwaltungsakt, Tübingen 1992; Christoph Engel, Legal Responses to Bounded Rationality in German Administration, Journal of Institutional and Theoretical Economics 150 (1994), 145; Christoph Engel, Allgemeines Verwaltungsrecht, in: Christoph Engel, Lawrence O'Hara, Stefanie Egidy, Yoan Hermstrüwer, Leonard Hoeft & Pascal Langenbach (eds.), Öffentliches Recht als Verhaltensordnung, Tübingen 2025, p. 135.

10 See, e.g., Christoph Engel, Der Einfluß des europäischen Gemeinschaftsrechts auf die deutsche Rundfunkordnung, Baden-Baden 1990; Christoph Engel, The European Charter of Fundamental Rights. A Changed Political Opportunity Structure and its Normative Consequences, European Law Journal 7 (2001), 151; Christoph Engel, Imposed Liberty and its Limits. The EC Treaty as an Economic Constitution for the Member States, in: Talia Einhorn (ed.), Spontaneous Order, Organization and the Law, The Hague 2003, 159.

11 See, e.g., Christoph Engel, Das Internet und der Nationalstaat, in: Berichte der Deutschen Gesellschaft fur Volkerrecht 39 (2000), 353; Christoph Engel, Governing the Egalitarian Core of the Internet, International Journal of Communications Law and Policy 10 (2005), 1; Christoph Engel, The Role of Law in the Governance of the Internet, International Review of Law, Computers & Technology 20 (2006), 201.

12 See, e.g., Christoph Engel & Johannes Kruse, Kommentar ohne Autor. Können Sprachmodelle das Kommentieren übernehmen?, Juristenzeitung 79 (2024), 997; Christoph Engel & Richard McAdams, Asking GPT for the Ordinary Meaning of Statutory Terms, Journal of Law, Technology & Policy 2024, 235; Christoph Engel & Johannes Kruse, LLM as a Law Professor. Having a Large Language Model Write a Commentary on Freedom of Assembly, Artificial Intelligence and Law, forthcoming 2026.

13 See, e.g., Christoph Engel, Die verfassungsrechtliche Zulässigkeit eines Entflechtungstatbestands im Kartellgesetz, Baden-Baden 2008; Christoph Engel, How Much Collusion? A Meta-Analysis of Oligopoly Experiments, Journal of Competition Law and Economics 3 (2007), 491; Christoph Engel, Tacit Collusion. The Neglected Experimental Evidence, Journal of Empirical Legal Studies 12 (2015), 537.

14 See, e.g., Christoph Engel, When is Intellectual Property Needed as a Carrot for Innovators?, Journal of Competition Law and Economics 7 (2011), 277; Christoph Engel, Geistiges Eigentum als Anreiz zur Innovation – Die Grenzen des Arguments, in: Martin Eifert & Wolfgang Hoffmann-Riem (eds.), Geistiges Eigentum und Innovation, Berlin 2008, 43; Christoph Engel & Marco Kleine, Who is Afraid of Pirates? An Experiment on the Deterrence of Innovation by Imitation, Research Policy 44 (2015), 20; Stefan Bechtold & Christoph Engel, The Valuation of Moral Rights. A Field Experiment, American Law and Economics Review, forthcoming at some point.

15 See, e.g., Christoph Engel & Joshua Fairfield, Privacy as a Public Good, Duke Law Journal 65 (2015), 385; Christoph Engel, Organizing Co-Existence in Cyberspace: Content Regulation and Privacy Compared, in: Christoph Engel & Adrienne Héritier (eds.), Linking Politics and Law, Baden-Baden 2003, 219.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

inology,[16] judicial process and evidence law,[17] property law,[18] tort law,[19] labor law,[20] landlord-tenant law,[21] and corporate law.[22]

Remarkably, Christoph Engel is one of the very few German scholars whose publications do not follow the traditional grouping into public, civil, and criminal law pillars, and he has always seen himself and was regarded as a true "Schüler" of the late and legendary Max-Planck Director Ernst-Joachim Mestmäcker, a renowned private law scholar. Several honorary degrees and distinguished positions have been awarded to Christoph Engel, with the accompanying laudationes often emphasizing

16 See, e.g., Christoph Engel, Experimental Criminal Law. A Survey of Contributions from Law, Economics, and Criminology, in: Willem H. van Boom et al. (eds.), Empirical Legal Research in Action, Cheltenham 2018, 57; Christoph Engel, Heike Hennig-Schmidt, Bernd Irlenbusch & Sebastian Kube, On Probation. An Experimental Analysis, Journal of Empirical Legal Studies 12 (2015), 252; Christoph Engel & Daniel Nagin, Who is Afraid of the Stick? Experimentally Testing the Deterrent Effect of Sanction Certainty, Review of Behavioral Economics 2 (2015), 405.

17 See, e.g., Christoph Engel & Lilia Zhurakhovska, You Are In Charge. Experimentally Testing the Motivating Power of Holding a Judicial Office, Journal of Legal Studies 46 (2017), 1; Christoph Engel & Andreas Glöckner, Can We Trust Intuitive Jurors? Standards of Proof and the Probative Value of Evidence in Coherence Based Reasoning, Journal of Empirical Legal Studies 10 (2013), 230; Christoph Engel & Keren Weinshall, Lucky You: Your Case is Heard by a Seasoned Panel: Panel Effects in the German Constitutional Court, Journal of Empirical Legal Studies 19 (2022), 1179.

18 See, e.g., Christoph Engel & Oren Bar-Gill, Bargaining in the Absence of Property Rights. An Experiment, Journal of Law and Economics 59 (2016), 477; Christoph Engel & Oren Bar-Gill, How to Protect Entitlements: An Experiment, Journal of Law and Economics 61 (2018), 525.

19 Christoph Engel & Theodore Eisenberg, Unpacking Negligence Liability. Experimentally Testing the Governance Effect, Journal of Empirical Legal Studies 13 (2016), 116.

20 See, e.g., Christoph Engel, Arbeitsmarkt und staatliche Lenkung, in: Veröffentlichungen der Vereinigung der Deutschen Staatsrechtslehrer 59 (2000), 56; Christoph Engel, Guido Bunstorf, Sven Fischer & Werner Güth, Non-Compete Clauses, Employee Effort and Spin-off Entrepreneurship: A Laboratory Experiment, Research Policy 45 (2016), 2113.

21 Christoph Engel, REITs ante portas – Die Anpassung des deutschen Rechts an institutionelle Investoren in den Grundstücks- und Mietmärken, Juristenzeitung 63 (2008), 1027.

22 See, e.g., Christoph Engel, The Behaviour of Corporate Actors. How Much Can We Learn from the Experimental Literature?, Journal of Institutional Economics 6 (2010), 445; Christoph Engel, Corporate Design for Regulability. A Principal-Agent-Supervisor Model, Journal of Institutional and Theoretical Economics 162 (2006), 104–124; Christoph Engel, Versetzt die Europäische Menschenrechtskonvention der Sitztheorie des deutschen internationalen Gesellschaftsrechts den Todesstoß?, Zeitschrift für Europäisches Privatrecht 1993, 150.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

this point: the non-rivalry of his interests and methods, the truly interdisciplinary approach, and, finally, the integrity with which he uses his toolkit.

While Christoph Engel began his publishing career entirely in German, he gradually integrated English-language publications into his production cycles. His first English-language publication dates from 1987.[23] International scholars and debates were frequent guests even in the early years of the Max Planck Project Group. In the 2000s, English became Christoph Engel's dominant language for scientific publications, thereby exposing him increasingly to international debate and competition. It is noteworthy, though, that Christoph Engel has always insisted on maintaining his imprint on German legal academia by continuing to contribute German-language publications, and being a regular attendant at the annual meetings of the renowned German "Staatsrechtslehrervereinigung" where he presented – not very surprisingly – on the border between civil law and public regulation in the field of labor-market regulation.[24] As a result, a fair amount of his scholarship appears in German up to this day (see Figure).

It was already mentioned how broad Christoph Engel's interests are. But some common threads can nevertheless be identified. Decision-making is certainly one of these: Christoph Engel has always been interested, from different perspectives, in the instruments and underlying assumptions of legal decision-making. Institutional design has therefore been a core interest. And although very few legal scholars would name him a doctrinal scholar, he nevertheless is in awe of the thousand-year-old fine-tuning of this legal method, which allows, as he once put it, to "react in a predictable way to unpredictable events." Interdisciplinary knowledge on decision making – be it from political science, economics, behavioral, cognitive sciences, or even from computer science – has given him the informed ground to re-establish, re-[formulate, and re-recognize legal decision principles. And as the decision-making process in law results frequently in court decisions, his interest in courtroom practices and their institutional design falls neatly into place.

23 Christoph Engel, The Position of Public Monopolies on Broadcasting under the European Convention on Human Rights, in: Ernst-Joachim Mestmäcker (ed.), The Law and Economics of Transborder Telecommunications, Baden-Baden 1987, 55.

24 Christoph Engel, Arbeitsmarkt und staatliche Lenkung, in: Veröffentlichung der Vereinigung der Deutschen Staatsrechtslehrer 59 (2000), 56.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

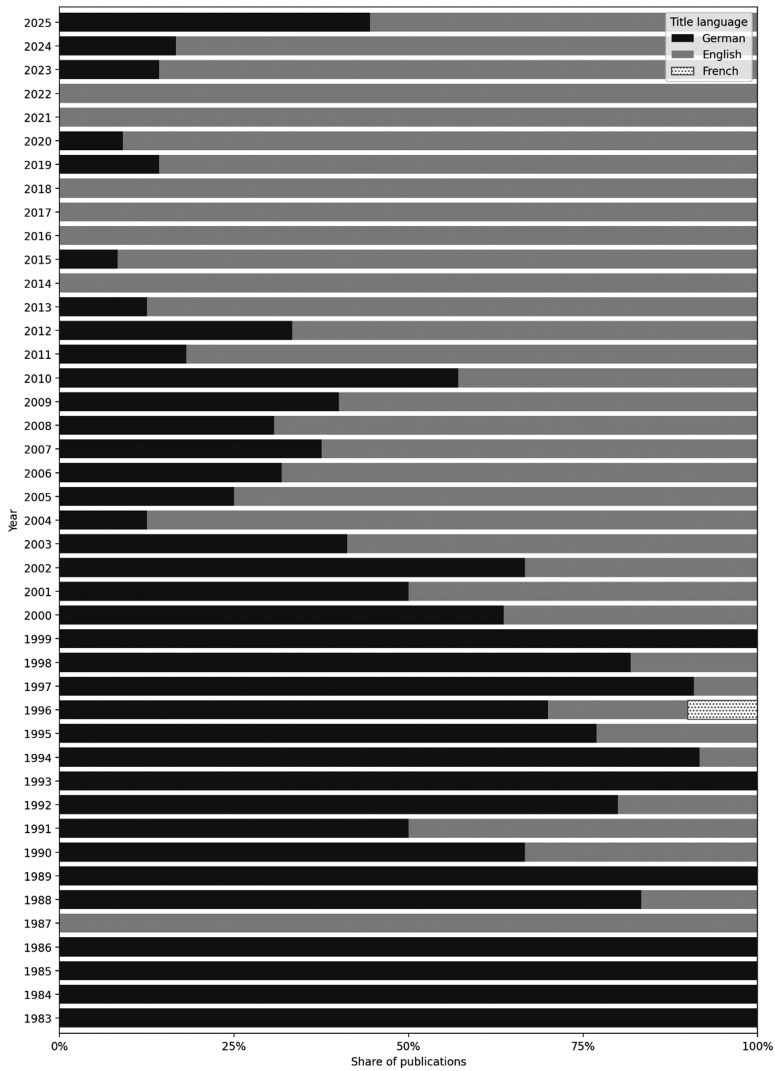


Figure: Publications by Christoph Engel, Language Distribution by Year (Normalized)

Note: A ChatGPT pipeline was used to determine the language of the publication titles listed in the Appendix to this Festschrift. Any errors are not ours, but ChatGPT's. The only French publication by Christoph Engel is L'accès du publique et l'accès des opérateurs aux technologies avancées de communication, in Fundamental Rights and New Information Technologies in the Audiovisual Sector, Munich 1996, 45.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

To put it mildly, Christoph Engel's extensive "legal area hopping" and his shift towards English-publication venues would already have set him apart among German law professors. But this is even before considering that Christoph Engel reinvented himself as a law-and-social-science scholar at a point in his career when other scholars would start thinking about how to solidify their legacy rather than learning mathematics and statistics. Being at ease with the legal methodology as his home turf, his hunger for intellectual development and fresh intellectual nourishment has consistently driven him into new interdisciplinary approaches. During his time at the Hamburg Max Planck Institute for Comparative and International Private Law in the 1980s and early 1990s, Christoph Engel had already been exposed to close exchanges between legal scholars and economists. And so it is not surprising that early signs of law & economics scholarship popped up in the early 1990s: Christoph Engel wrote about efficient breach in contract law in 1994,[25] functional approaches towards tort law in 1995,[26] and law as an applied social science in 1998.[27] This quickly established his reputation as a fresh mind and a communicator between fields.

But only when Christoph Engel had convinced the Max Planck Society that the concept of an interdisciplinary legal institute could work and was then vested, jointly with Martin Hellwig, with building a new Max Planck Institute from scratch, did he take his interdisciplinary interests to another level. He began focusing on law & psychology and behavioral approaches in the early 2000s.[28] The late 2000s and early 2010s were the period of meta-studies in Christoph Engel's scholarship.[29] The 2010s were almost entirely devoted to experimental law & economics. Over the last

25 Christoph Engel, "Effizienter Vertragsbruch": eine juristische Antwort, Homo Oeconomicus 1994, 143.

26 Christoph Engel, Zivilrecht als Fortsetzung des Wirtschaftsrechts mit anderen Mitteln. Rechtspolitische und verfassungsrechtliche Überlegungen am Beispiel des Haftungsrechts, Juristenzeitung 1995, 213.

27 Christoph Engel, Rechtswissenschaft als angewandte Sozialwissenschaft. Die Aufgabe der Rechtswissenschaft nach der Öffnung der Rechtsordnung für sozialwissenschaftliche Theorie, in: Christoph Engel (ed.), Methodische Zugänge zu einem Recht der Gemeinschaftsgüter, Baden-Baden 1998, 11.

28 See, e.g., Gerd Gigerenzer & Christoph Engel (eds.), Heuristics and the Law, Cambridge 2006.

29 Christoph Engel, How Much Collusion? A Meta-Analysis of Oligopoly Experiments, Journal of Competition Law and Economics 3 (2007), 491; Christoph Engel, Dictator Games. A Meta-Study, Experimental Economics 14 (2011), 583; Christoph Engel,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

few years, he has primarily turned to machine learning and the law.[30] And, not altogether unexpectedly, Christoph Engel has recently begun to integrate his experimental and machine-learning interests by exploring the extent to which behavioral experiments can be conducted on large language models[31] – thereby, in a sense, rendering obsolete the experimental infrastructure he built over the course of his career at his Max Planck Institute and relocating experimental social science onto Nvidia chips. Who knows: perhaps Christoph Engel will turn to post-quantum empirical legal scholarship in the early 2030s.

II. The Institution Builder

Christoph Engel had already experienced the world of Max Planck as a Research Associate in Ernst-Joachim Mestmäcker's group at the Max Planck Institute for Comparative Law and Private International Law in Hamburg between 1983 and 1992. Following his immediate appointment after habilitation as a full professor of law in Osnabrück at the age of 36, he was shortly thereafter given the once-in-a-lifetime opportunity to build a new Max Planck Institute from scratch – at that time, one of the youngest Max Planck directors ever. And while Christoph Engel's scholarly achievements are truly impressive and will gain him a place in the legal profession's "Walhalla", his creation of the Max Planck Institute for

Tacit Collusion. The Neglected Experimental Evidence, Journal of Empirical Legal Studies 12 (2015), 537.

30 Christoph Engel & Nina Grgić-Hlača, Machine Advice with a Warning about Machine Limitations. Experimentally Testing the Solution Mandated by the Wisconsin Supreme Court, Journal of Legal Analysis 13 (2021) 284; Christoph Engel & Richard McAdams, Asking GPT for the Ordinary Meaning of Statutory Terms, Journal of Law, Technology & Policy 2024, 235; Ricardo Dominguez-Olmedo, Vedant Nanda, Rediet Abebe, Stefan Bechtold, Christoph Engel, Jens Frankenreiter, Krishna P. Gummadi, Moritz Hardt & Michael Livermore, Lawma: The Power of Specialization for Legal Annotation, International Conference on Learning Representations (ICLR), 2025; Christoph Engel & Johannes Kruse, LLM as a Law Professor. Having a Large Language Model Write a Commentary on Freedom of Assembly, Artificial Intelligence and Law, forthcoming 2026.

31 Christoph Engel, Max Grossmann & Axel Ockenfels, Integrating Machine Behavior into Human Subject Experiments: A User-friendly Toolkit and an Application to Framed Prisoner's Dilemmas, Experimental Economics, forthcoming 2026; Christoph Engel, The Negotiation Trap: An Experiment on a Large Language Model, Max Planck Institute Discussion Paper 2024/19.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Research on Collective Goods has arguably become his greatest masterpiece.

The starting point of this new Max Planck Institute (which initially existed as a Max Planck Project Group for five years) arose on the initiative of Hans Zacher, a law professor and President of the Max Planck Society until 1996. Although Zacher had, at first, thought more along the lines of common goods with respect to public services, coming from the area of social law, he was very open to a broader, more open understanding during the Institute's founding process. Initially, the idea had been to start with two directors in law and (empirical) political sciences and then, once the Institute was formally established, to enlarge it with a third group focusing on economics.

Although the Institute became a well-oiled machine for research and intellectual exchange over the years under the wise leadership of Christoph Engel and Martin Hellwig, this did not fall into their laps. It was the result of the hard work of both directors, many researchers, and the Institute's devoted permanent employees, including – but not limited to – his assistant Anja Schäfer and her predecessor Anja Moosmann, the Institute's head of administration Heidi Morgenstern, and the Institute's head librarian, the late Regina Goldschmitt. Most importantly, it was Christoph Engel's persuasiveness and credibility as a motor for a new research direction in law, his ability to find wise, good-willed, and influential mentors in the Advisory Board of the Institute and within the Max Planck Society, as well as his talent to line up innovative, challenging, and reliable companions in academia more broadly. The names on the Institute's Advisory Board were impressive. But it was even more impressive to observe how active these board members were in their interactions with the Institute and its team, and how much they helped the Institute position itself.

An episode from the Institute's founding process may illustrate Christoph Engel's role in all of this. Shortly before the decisive evaluation of the temporary Max Planck Project Group on "Law of Common Goods" was concluded, which would decide whether the Max Planck Society would invest in the creation of a legal interdisciplinary institute, Christoph Engel's then political science co-director had surprisingly decided to accept a position abroad, thus throwing the Institute into an existential crisis before it even existed. It was Christoph Engel who created a convincing new research agenda focusing on the integration of law and economics, brought Martin Hellwig into the game, and turned the threat into a promising and flourishing scientific enterprise that would

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

develop its own stamp over the years. Similarly, when the newly founded Max Planck Institute needed a new home in Bonn, it took all the strategic foresight and negotiating skills of Christoph Engel to secure not only North Rhine Westphalia as the hosting state (against opposition from other states), but also a beautiful neo-baroque villa that had been built for the wife of the Brazilian Consul to Germany in the 1920s, and later became the residence of the Egyptian Ambassador to Germany, as the new home of the Max Planck Institute for Research on Collective Goods.

Collective Goods. It was a wise choice to give the Institute a distinct name that was open to interpretation, rather than rely on more standard terminology such as "public goods" or "common goods," which had been dominant in the title of the preceding project group, "Law of Common Goods". When Christoph Engel's long-time co-director of the Max Planck Institute for Research on Collective Goods, Martin Hellwig, was once asked what a collective good is, he replied dryly: "That still has to be researched." It was little ingredients like these that turned the Institute into an institution where intellectual curiosity and creativity were the drivers and only real limits to research ambitions. It certainly also helped that the Institute was led by two congenial directors who not only excelled in their own research but also shared a similar cast of mind and values.

Christoph Engel's institution-building activities reached beyond his Max Planck Institute in many regards. Let us look at some of the most revealing ones. Since 2002, he has, for a long time together with his University of Bonn colleague, Urs Schweizer, organized the seminar series on "New Institutional Economics", whose conference proceedings have always been published in the Journal of Institutional and Theoretical Economics. The "JITE" conferences, whose first iteration in 1984 featured speakers such as Ronald Coase, Douglas North, and Oliver Williamson, became the home of Christoph Engel's international research network for more than two decades.

Christoph Engel has been a member – and former Vice Chairman – of the Academic Advisory Council to the German Ministry of Economics for nearly three decades. Economic policy advice in Germany has benefited greatly from Christoph Engel's broad legal knowledge, tireless work ethic, and willingness to engage with economists. However, he has arguably benefited even more from his access to the ministerial bureaucracy and to excellent economists through the Advisory Council. Furthermore, the former and present Pope rely on his expertise, as he accepted the call to join the *Pontificia Academia Scientiarum Socialium*. Christoph Engel

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

is a modest man. But getting appointed to the Pope's Academy of Social Sciences made even him proud. His innovative drive, as well as his amicable, charming, yet clear and forceful personality, paved the way for him to serve on the prestigious Board of the German "Staatsrechtslehrervereinigung" from 2008 to 2009. For several years, he served as head of the social sciences section within the Max Planck Society, guiding Max Planck's direction across the entire field of social science.

III. The Mentor

While the Max Planck Institute provided Christoph Engel with unparalleled resources and infrastructure, as well as an administrative team that understood its mission as one of genuine enablement, he ultimately considered encouraging young researchers to be one of the most important tasks. While at first surrounding himself with doctoral students – after all, he was one of the first to establish an international Max Planck research school for doctoral students at the Max Planck Institute (IMPRS) – he focused on postdoctoral scholars as well, and not only in law, as the list of his former postdocs in this volume demonstrates impressively.

Unlike many other legal scholars, Christoph Engel never tried to establish a "Schule" (in the philosophical sense) in which his "Schüler" and consequently their "Schüler" again all share a common approach to particular legal questions. He was never interested in hearing his own thinking repeated. Rather, he was looking for new input on his own research areas from others. As he put it, there is genuine happiness in seeing that his mentees become what they want to become. He exerted influence as an academic role model, and colleagues and "Schüler" have often wondered how he manages not to yawn even once after a long day of discussions, workshops, or group meetings, yet remains focused. It is his great pleasure in intellectual endeavors that motivates him, and, as he was privileged to be a Max Planck Institute director, he pursues his research interests at his own pace, always researching, composing, and formulating his texts himself.

Like his own mentor, Ernst-Joachim Mestmäcker, Christoph Engel cherishes relationships with his former "Schüler" through a regular yearly reunion, and it is their success as the next generation of scholars that makes him truly proud. By now, well over a dozen of his former mentees are professors of law, economics, psychology, and other social sciences at

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

universities across Europe and the United States, all with their very own profiles and independent interests, and another considerable group has pursued remarkably successful careers in the judiciary, law firms, government, the private sector, and beyond.

Given Christoph Engel's breadth of interests, it is not surprising that the typical boundaries of the German legal system have not held him back in his mentorship activities. Early in his career, Christoph Engel – the public law scholar – co-founded the German Society for Young Private Law Scholars.[32] There were periods at the Institute when he simultaneously mentored doctoral students and postdocs in public, private, and criminal law. This was probably as close as Christoph Engel ever came to heresy in the eyes of German legal academia.

IV. The Original

Christoph Engel is a true original in legal academia. He has overcome many of the boundaries that his field had set for itself and explored so many paths over the course of his career that even he must, at times, be astonished by where his curiosity has taken him.

For outsiders, some of Christoph Engel's characteristics are exactly that: typical of him. There is his style of beginning articles with a catchphrase (e.g., "It can be wrong to say the right thing. Even taboos serve a social function").[33] When he comments on a presentation, he often starts with a general statement that seems unrelated to the topic. Still, in the more detailed analysis, he offers a comprehensive alternative approach to the topic. He is always friendly, and even his closest companions have almost never been able to discern a "bad day" or a "bad mood". Coming from Tübingen in the state of Baden-Württemberg, and therefore from Swabia, luxury is not something he ventures into, neither for himself nor for his guests – so do not expect a dinner invitation to Bonn's famous 37-year-in-a-row Michelin-star restaurant around the corner from his house, but rather a down-to-earth meal without champagne. Staying in a pretty run-

32 Harald Baum, Christoph Engel, Olivier Riem & Manfred Wenckstern, Jahrbuch Junger Zivilrechtswissenschaftler 1 (1991), Stuttgart 1992.

33 Christoph Engel, Arbeitsmarkt und staatliche Lenkung, in Veröffentlichung der Vereinigung der Deutschen Staatsrechtslehrer 59 (2000), 56. See also the chapter "Der schreibt auch so" by Hanjo Hamann in this Festschrift.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

down youth hostel in Walle-Rullenhorst, somewhere in Westphalia, therefore, is no matter to him as long as the topic is worth it (at that time: Luhmann's systems theory). He is interested in arguments, not in status, and thus he is independent in his judgement of who is an inspiring academic.

As a U.S. colleague once put it, Christoph Engel's appearance reflects the very epitome of a scholar: fully intellectual, indifferent to food and comfort if not food for thought, never wearing jeans, let alone jogging pants, but always a shirt, jacket and – a bowtie.

All of us have our own images of Christoph Engel – a Festschrift is an honor bestowed by colleagues who have become academic friends. Thus, we have all experienced him in different places, times, situations, roles, and circumstances over more than forty years of academic endeavor. But as Christoph Engel is now actively working to replace experimental social science labs and legal commentaries with large language models, it seems appropriate to conclude this chapter by making ourselves superfluous and asking ChatGPT 5.4 himself (or herself? Or itself?) to write a laudatio for Christoph Engel.

Christoph Engel's scientific œuvre is marked by a rare combination of intellectual range, analytical discipline, and institutional imagination. What begins in his earlier work as a jurist's sustained engagement with doctrine, institutions, and the structure of legal ordering develops, over time, into something much more ambitious: a research program that treats law not merely as a system of norms, but as a mode of organizing behavior under conditions of bounded rationality, social interdependence, and institutional constraint. The official profile of his work now captures this development in three terms – Empirical Legal Studies, Experimental Law and Economics, AI and the Law – but the formula understates the coherence of the whole. The unifying thread is Christoph Engel's persistent insistence that legal scholarship should not be content with internal reconstruction alone. It should also ask how legal institutions actually work, how they shape motives and expectations, and under what conditions they succeed or fail.

In substantive terms, Engel has made major contributions across an astonishingly broad field: judicial behavior, sanctions and deterrence, corruption, class actions, privacy, property, innovation, taxation, customary law, and, more recently, the implications of artificial intelligence for legal reasoning. Yet this breadth is not eclecticism. Again and again, his work returns to a central question: how does law interact with the behavioral

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dispositions of those who make, apply, and live under it? That question animates, among many others, his work on the motivating power of judicial office, role-induced bias in adjudication, the governance effects of negligence liability, transparency and punishment, privacy as a public good, and the diffusion of legal innovations.

Methodologically, Christoph Engel's importance can hardly be overstated. He is among the very small number of European legal scholars who did not simply borrow empirical methods from neighboring disciplines, but helped establish the conditions under which those methods could become integral to legal research itself. Just as important, his reception in the United States shows that his work has never been merely national or regional; it has entered the core transatlantic conversation. Christoph Engel has not only been a participant in, but one of the builders of, the field's international infrastructure. In that fuller sense, his œuvre is both scholarly and institutional: it has expanded the imagination of legal research and helped create the community in which that expanded imagination could flourish.

Christoph Engel's work has been foundational in legitimizing experiments, behavioral analysis, and empirically disciplined causal inference as tools for answering genuinely legal questions. Just as importantly, he has been reflective about method: he has not treated empirics as a badge of modernity, but as a demanding craft that requires conceptual clarity, institutional sensitivity, and constant vigilance about inference. In this lies, perhaps, his most enduring contribution. He has shown that legal scholarship can be simultaneously more rigorous and more realistic; more open to economics, psychology, and now computational tools; and yet remain unmistakably legal in its questions, categories, and ambitions. Few scholars have done more to shape empirical legal studies in Europe, and fewer still have done so with comparable breadth, depth, and lasting influence. Chapeau!

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

„Der schreibt auch so."
Die persönliche Handschrift des Ge(l)ehrten im Spiegel seiner Auftaktsätze

Hanjo Hamann

War ja klar, dass Christoph Engel sich für Künstliche Intelligenz begeistern würde.[1] Und jetzt ist sie sowieso überall. Eines ihrer vielen Potentiale ist die Autorschaftsattribution. Damit kann man Autoren daran erkennen, *wie* sie schreiben. Das ist bei Christoph Engel ganz leicht. Schon zu Beginn eines seiner wichtigsten Bücher. Gemeint ist der Band über das „Proprium der Rechtswissenschaft". Dessen Vorwort schrieb Engel nicht alleine. Doch der Auftakt stammt unübersehbar von ihm:

> „Juristen schreiben nicht auf Englisch. Juristen veröffentlichen keine *discussion papers*. Juristen stellen ihre Texte nicht im Internet zur Verfügung. Juristische Aufsätze sind nicht *peer reviewed*. Juristen achten nicht auf den *impact factor*. Juristen finanzieren ihre Forschung nicht aus Drittmitteln. Juristen haben keine Sonderforschungsbereiche. Juristen schreiben keine Modelle. Juristen verwenden keine Mathematik. Juristen falsifizieren keine Hypothesen. Juristen nutzen keine Statistiken. Juristen führen keine Interviews. Juristen machen keine Experimente. Zu jeder dieser Aussagen gibt es Ausnahmen."[2]

In der Sache hatte er Recht. Damals. Heute etwas weniger. Dank ihm. Darum geht es ja in dieser Festschrift.

Aufsehen erregte aber sein Stil. Stakkato. Kein Satz mehr als neun Worte. Das war ungewöhnlich. In Bonn gab es damals einen emeritierten Rechtshistoriker. Der hieß Jakobs. Er schrieb ein Jahr später:

> „Kurze Sätze schreiben. Subjekt Prädikat Objekt usw. Oder noch nicht einmal das. Ein einziges Wort und Punkt geht auch. So schreiben heute selbst emeritierte Aka-

1 Bspw. *Engel*, Discussion Paper 2023/10; *Engel/Kruse*, JZ 2024, 997; *Engel/McAdams*, J. Law Tech. & Pol'y. 2024, 235; *Engel*, JITE 2024, 365; *Engel*, Discussion Paper 2024/10; *Engel*, Discussion Paper 2024/12; *Engel/Hermstrüwer/Kim*, Discussion Paper 2025/3; *Engel*, Discussion Paper 2025/8.

2 *Engel/Schön*, Vorwort, S. IX in: Engel/Schön (Hrsg.), Das Proprium der Rechtswissenschaft, 2007.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

demiepräsidenten. In Bonn gibt es ein Max-Planck-Institut mit einem Juristen als Direktor. Er heißt Engel. Der schreibt auch so."[3]

Stimmt. Und Uwe Wesel auch. Noch ein emeritierter Rechtshistoriker. Aber einer, dessen Bücher man vergnügt und mit Leichtigkeit las. Emeritierte Rechtshistoriker sind eben nicht alle gleich.[4] Aber zurück zu Engel. Der ist zwar kein Rechtshistoriker. Aber nun immerhin pensioniert.

Im Gegensatz zu seinem Stil. Dazu eine Anekdote. Vor zehn Jahren wurde Engel sechzig. Was wünscht sich einer zum Geburtstag, der schon alles erreicht hat? Ein kleines Privatsymposium am eigenen Institut. Da sprachen viele Wegbegleiter. Einer schoss den Vogel ab. Alle amüsierten sich köstlich. Was hatte er getan? Auftaktsätze vorgelesen. Auftaktsätze vieler Aufsätze von Christoph Engel. Alle erkannten den ausgefallenen Stil. Die Idee war grandios.

Das Manuskript von damals ist leider verschollen. Aber nun haben wir ja Künstliche Intelligenz. Die könnte alle Auftaktsätze aus Engels Preprints extrahieren. Dachte ich. Konnte sie aber nicht. Aber bei 137 Texten geht das auch von Hand. Es zeigt sich: War Engel Alleinautor (81 Mal), umfasste sein erster Satz im Schnitt acht Worte. Schrieben andere mit (56 Mal), waren es dreizehn. Engel allein eröffnete 91 % seiner Texte in unter 15 Worten. Schrieben andere mit, waren 27 % der Eröffnungssätze *mindestens* so lang. Klarer Fall von *smoking gun*.[5]

Nehmen wir also nur Auftaktsätze unter 15 Worten.[6] Nur ein einziger beginnt ganz bieder mit „In this paper, we …" (2014). Wetten, dass er von Engels Koautorin stammt? Autorschaftsattribution geht eben auch ohne KI. Die 115 weiteren Auftakte sind ganz anders. Der kürzeste kommt mit zehn Zeichen aus: „Es reicht." (2005). Was für ein Beginn. Daneben finden sich mindestens sieben engeltypische Varianten:

3 *Jakobs*, Der demokratische Schreibstil oder Abschied vom Komma, myops 2008, Heft 3, 74.

4 Ihre Lebenszeiten dagegen schon: Jakobs 1934–2023, Wesel 1933–2023. Beide fehlen uns heute.

5 Wer Engel kennt, weiß: Er liebt diese Metapher. Vgl. Preprint 2017/7, S. 16; Preprint 2017/20, Abstract, S. 3, S. 18; Discussion Paper 2024/4, Abstract, S. 16, 37.

6 Fundstellen dokumentiert im Zenodo-Datensatz Preprints & Discussion Papers of the Max Planck Institute for Research on Collective Goods Bonn (1998–2024) (doi.org/10.5281/zenodo.15479677).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Fragender Auftakt: „Is competition good for innovation?" (2007), „Experimental public international law?" (2010),[7] „Does damages liability promote socially desirable behavior?" (2012), „Is there anything hogs and law professors have in common?" (2012), „Why does a case go to court?" (2015),[8] „How should entitlements be protected?" (2017) und „2 + 2 = ? What a stupid question!" (2007).

Antithetischer Auftakt: „Institutions are constraints, rational choice institutionalists say. Institutions are social constructions, holistic institutionalists say." (2006), „European law is irrational. American law is irresponsible." (2008),[9] „You are facing a dilemma – what a mess! Your opponent is facing a dilemma – what a relief!" (2009),[10] „One man's meat is another man's poison, as they say." (2013),[11] und „Boilerplate is good for suppliers and bad for consumers." (2024)

Lateinischer Auftakt: „Iudex non calculat." (2006),[12] „Credo quia absurdum." (2007), „Jurisprudentia, non juris scientia." (2018)[13] und „Impossibilium nulla est obligatio. Wenn es ernst wird, flüchten wir Juristen uns ins Lateinische." (2008)

Zitierender Auftakt: Engel zitiert ganz ökumenisch. 1999 Lenin („Vertrauen ist gut, Kontrolle ist besser."), 2005 das Matthäus-Evangelium („Whoever humbles himself will be exalted.") und Max Weber („Politik gehört nicht in der Hörsaal."), 2016 Richard Posner („What Do Judges and Justices Maximize? The Same Thing Everybody Else Does.") und 2017 die Menschenrechte („The Universal Declaration of Human Rights stipulates: [...]").

Apodiktischer Auftakt: Die apodiktische Kopula bildet wohl die häufigste Engel-Eröffnung. Er apodizierte: „Good constitutions are short and

7 Und weiter: „Isn't it patent that such an endeavour is doomed to failure?"

8 Und weiter: „Because there is a dispute. Why is there a dispute? Usually because some facts are disputed.".

9 Und weiter: „These beliefs are the essence of one of the few true conflicts between (US) common law and (European) continental law."

10 Und weiter: „This is the situation of cartel members vs. cartel authorities." Siehe nächste Fußnote.

11 Und weiter: „Antitrust is a field of application. For those forming a cartel, or coordinating tacitly, collusion is a dilemma."

12 Und weiter: „For most lawyers, the proverb settles it. Formal methods are none of their business."

13 Und 2020 noch: „Es gibt einen Grund, warum das Fach Jurisprudenz heißt und nicht Jurisscienz."

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

enigmatic." (2001), „Governance by law is not fashionable." (2002), „Richness can be embarrassing." (2006), „Words are a source of power." (2007), „Das deutsche Kartellrecht ist unvollständig." (2007), „Nothing is as uncertain as the future." (2007) „Nichts ist so wirksam wie das Selbstverständliche." (2009), „Religion is universal." (2011), „Corruption is among the most vexing social evils." (2012), „Eventually, all law is behavioral." (2013), „Sanctions are frequently imperfect." (2013), „Life is fraught with risk." (2015), „Academia is a worthy topic for anthropological research." (2016), „Individual preferences are the backbone of economic theory." (2018), „Judges are no machines." (2020), „Interdisciplinarity is akin to foreign trade." (2020), „Experiments are artificial by design." (2022), „The law is a methodological omnivore." (2022), „The German Constitutional Court is an active policymaker." (2024), und „Superintelligence is on the horizon." (2024)

Provokativer Auftakt: Provozieren können die Auftaktsätze alle. Gleich ob sie auf rhetorische Fragen, Juristenlatein, Fremdzitate, Antithesen oder Apodiktion setzen. Mancher Auftakt jedoch suchte geradezu die Provokation: „Der Wolf hat Kreide gefressen." (1998), „Ein Fach verweigert sich seinem Gegenstand." (2002), „Regulators are not good at multi-tasking." (2003), „Beware of the dismal science!" (2017) oder „The legislator speaks and nobody listens." (2017).

Philosophischer Auftakt: Man müsste fast T-Shirts drucken mit Engeliaden wie

- „Kaum jemand kennt die Grammatik seiner eigenen Sprache." (2000)
- „Lawyers are the engineers of the social sciences, and their doctors." (2001)
- „Wer zu lang im Ausland lebt, kann schließlich heimatlos werden." (2001)
- „Good politicians are experts in the art of insulting." (2002)
- „Es gibt kaum ein säkulareres Volk als das deutsche." (2002)
- „When God expelled Adam and Eve from paradise, he created rationality." (2004)
- „A healthy dose of illusions is a necessary ingredient of success." (2004)
- „Angekündigte Revolutionen bleiben meist aus." (2005)
- „Times Square, Manhattan, 1990: clearly not the place to be." (2009)
- „Not many scientific achievements are cast in stone." (2010)

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

– „Ultimately, reality trumps philosophy." (2017)
– „Most deep insights are simple – in retrospect." (2020)
– „A social dilemma is called a dilemma for a reason." (2024)

Nun kennen wir die sieben Basisvariationen der Engel-Eröffnung. Fast noch interessanter aber ist die Dialektik seiner Auftaktsätze über mehrere Texte hinweg:

– 2002 eröffnete Engel dystopisch: „Reality is dirty. People murder and injure each other, they steel and lie, they rape and insult their neighbours."[14] Viel philantropischer schon 2013: „Most people do not steal, trespass, blackmail, assault, rape or kill." Ganz sokratisch dann 2021: „Why do individuals commit crime?"
– 2002 meinte Engel: „Katzenjammer ist ausgebrochen in der Wettbewerbspolitik." 2006 dann martialisch: „Im Wettbewerb hat man die Lizenz zum Töten."
– 2003 erklang sein Kassandraruf: „Begriffe können blind machen."[15] 2011 kündete er kämpferisch: „Conceptual rigour can be a dangerous weapon."
– Sogar der Lebenszyklus einer Ehe zieht sich durch die Einstiegssätze. Ihr Beginn 2008 in „Wenn man einmal im Leben einen Ehering braucht, lernt man nicht Goldschmied." Ihr Ende 2013 in „Divorce is a messy affair, not only for couples."[16]
– Und als dialektischer Höhepunkt schließlich Tod und Wiederauferstehung gleich zu Beginn: „Otto Mayer ist tot. Es lebe Otto Mayer." (2008)

14 Man bedenke die Reihung: „murder and injure", „rape and insult".

15 Und weiter: „Was ein Markt ist, weiß doch jeder. Jedenfalls jeder leidlich gebildete Kartellrechtler."

16 Wie „messy", das lernte ich nicht von Engel. Sondern von seinen Kollegen. Immerhin schon bei meinem ersten Kontakt mit Engels Welt: Auf der Sommerschule 2010 seines Graduiertenprogramms (IMPRS Uncertainty). Dort notierte ich wörtlich etwa: „Sometimes you think ‚Oh, it was not a good idea to marry that person. The payoff matrix is really bad, all outcomes are negative, my situation is really miserable.' You may decide to get a divorce, but then the payoff is also negative [...] So some economists come to the decision to kill their wife..." (*Ido Erev*, 29.7.2010) – „You know what the current affiliation of Rafael Robb is? Pennsylvania State Prison. He is a game theorist, game theorists are dangerous. So when he faced the divorce settlement with his wife, he did some calculations and found it didn't pay – so he killed her." (*Carlos Alós-Ferrer*, 10.8.2010).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Das waren 70 Engel-Auftakte. So viele wie Engels Lenze. Und doch nur gut die Hälfte unserer Stichprobe. Da bleibt noch manches zu entdecken. Als Jakobs einst sarkastisch monierte: „Ein einziges Wort und Punkt geht auch", da hatte ein Schalk schon acht Jahre zuvor paradox präpliziert: „Geht nicht." Aber damit widersprach er nicht Jakobs. Sondern kam gerade erst in Fahrt:

> „Geht nicht. Ist gescheitert. Apolitisch. Ahistorisch. Naiv. Vormodern. Unwissenschaftlich. Schädlich. Solche wenig schmückenden Beiworte bekommt angeheftet, wer heute noch aufbricht, das Gemeinwohl zu definieren." (2000)

Zu Unrecht. Geht wohl! War erfolgreich. Glückwunsch dazu! Und damit schließt sich der Kreis.[17]

17 Nämlich zu Engels Auftaktsatz im allerersten Preprint: „Auch ein Glückwunsch kann befangen machen." (1998)

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Cooperation in Social Dilemmas
28 Years of Research on Collective Goods

*Svenja Hippel & Dorothee Mischkowski**

"When God expelled Adam and Eve from paradise, he created rationality. Not only would humans have to cope with scarcity, and hence have to trade some desired goods for others. They would even be pitted against each other. Individually rational behaviour would lead to social dilemmas." (Engel, 2004, p. 2)

A. Introduction

Cooperation, understood as the willingness to forgo immediate self-interest for the sake of collective benefit, forms a cornerstone of functioning societies. Yet in social dilemmas, where individual payoff maximization conflicts with group welfare, cooperation becomes both crucial and inherently puzzling. Christoph Engel has fundamentally advanced our understanding of this tension through a genuinely interdisciplinary research program that unites law, economics, and experimental psychology. His work shows how institutional structures – legal rules, social norms, and economic incentives – shape the conditions under which cooperation can emerge or fail, and how psychological factors such as expectations about others and individual differences in social preferences guide behavior in these environments. The resulting research represents a level of cross-disciplinary synthesis that remains rare in the academic landscape.

Understanding the mechanisms that foster or hinder cooperation has far-reaching implications. Social dilemmas underlie a wide range of real-world challenges. "In fact, people's selfishness is often the biggest obstacle to reaching better social outcomes. From environmental pollution over tax evasion, corruption and misappropriation, to doping, queuing, and bank runs, the list of situations is sheer endless in which the collective interests of society and the selfish interests of its individual members are

* Both authors contributed equally to this article.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

at odds" (Engel & Kurschilgen, 2015, p. 2). Even scientific integrity is not immune to adverse dilemma structures: Engel (2015) discussed various mechanisms for ensuring scientific integrity nonetheless (e.g., sanctions and internalized norms) to prevent personal incentives from dominating, showing, however, that none of these solutions are perfect. Adding further complexity, there are situations such as competition in markets, where social dilemmas might even be a socially desirable feature (Engel, 2009; Engel & Zhurakhovska, 2011).

Earlier research already identified key determinants of human cooperation (e.g., Camera & Casari, 2009; Fischbacher & Gächter, 2010; Friedman & Oprea, 2012). Yet important questions remained about why cooperation sometimes thrives despite misaligned incentives – and why, in other instances, it ultimately collapses. The work of Christoph Engel and his group at the Max Planck Institute for Research on Collective Goods has provided critical new insights into these dynamics.

B. Determinants of Cooperation

In what follows, we organize our contribution around the key determinants of cooperation that Christoph Engel has investigated. We begin with contextual factors – such as institutional mechanisms (e.g., sanctions and punishment), the role of authorities, social norms, and negative externalities – that shape the extent of cooperative behavior. A second section highlights individual differences and the dynamics of conditional cooperation, with particular attention to the fragility of cooperation following negative experiences (e.g., exploitation by interaction partners). Finally, we turn to the psychological reactions to insufficient cooperation, including Christoph Engel's pioneering work on deliberate ignorance.

I. Sanctions and Punishment

Peer punishment provides contributors to public goods with an additional tool for enforcing cooperation within the group. Fehr and Gächter (2000) demonstrated that individuals are indeed willing to punish defectors despite bearing personal costs, and that such punishment can help sustain cooperation.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In Engel (2014), the central puzzle is why even weak sanctions – those too mild to deter a payoff-maximizing free rider – nonetheless stabilize cooperation. Here, Christoph Engel integrates formal theorizing with experimental evidence, demonstrating that the solution to this puzzle is rooted in social preferences – specifically, advantageous inequity aversion. A parsimonious model illustrates how even non-deterrent sanctions can amplify weak inequity aversion, shifting the incentive balance in favor of cooperation. Experimentally, participants increased their contributions following punishment, even when sanctions were too small to deter a strictly payoff-maximizing individual. This response was markedly stronger among participants with higher inequity aversion, indicating that the stabilizing effect of sanctions depends on social preferences, not deterrence alone. Together, these findings reveal that imperfect sanctions complement underlying distributive concerns, relying on preference-based cooperation motives alongside – not instead of – material incentives.

Engel, Kube, and Kurschilgen (2021) turn to the effect of selective information on cooperation in the presence of sanctions. In their experiments, subjects received favorable or unfavorable histories of cooperation from unrelated groups before playing public good games with and without peer- and counter-punishment. Unfavorable information sharply lowered initial contributions through influencing initial beliefs and substantially reduced the effectiveness of punishment, preventing groups from recovering once cooperation started at a low level. Favorable information, by contrast, helped stabilize contributions in less cooperative environments but had little effect where cooperation was already high. Thus, while punishment can in principle enforce norms, its effectiveness is fragile and highly dependent on the informational environment in which expectations are formed.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Authorities

While peer punishment allows contributors to stabilize cooperation on their own, an authority – a person with elevated status and decision-making power – may also be present, shaping the situation through their actions. Such an authority figure may be either a member of the group or an external third party.

Engel and Hippel (2017) demonstrate both the promise and limits of such an institutionalized authority in solving social dilemmas. They bridge mechanism design and experimental public good research by studying the behavior of experimental social planners. In their design, a randomly assigned authority could impose a contribution rule on group members with heterogeneous valuations for a public good. Across treatments, the authors varied whether valuations were common knowledge or private information (with scope for misrepresentation), and whether the authority was personally affected by the rule.

Disinterested planners predominantly opted for rules equalizing final payoffs across participants, even when efficiency would have dictated otherwise. Strikingly, they were overly optimistic about truth telling: Most expected active participants to report valuations honestly, despite strong incentives to misrepresent. Once the planner's own interests came into play, optimism gave way to self-serving behavior. When deciding under a veil of ignorance, planners balanced efficiency and fairness, but they still overestimated others' willingness to reveal valuations truthfully. When they knew their own valuation, decisions became straightforwardly selfish. While many experimental planners acted in line with fairness norms when being personally unaffected, the combination of optimism about others' honesty and the temptation of self-interest meant that even institutional solutions remained precarious. This cautions against assuming that the mere presence of a (hopefully) benevolent authority guarantees efficient public good provision.

III. Externalities and Voting

While punishment expands the players' action space by adding behavioral response options, authorities are introduced as additional actors who constrain or enforce the boundaries of choice. However, public good provision is typically embedded within a broader social context. Externalities –

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

positive or negative effects on bystanders – are a defining and ubiquitous feature of such environments and may meaningfully shape contributions to public goods.

Engel and Rockenbach (2009) examine in an initial study how externalities influence cooperation when passive outsiders are affected. They find that the presence of either negative or positive externalities in a group's public good provision had almost no impact on contribution levels, provided that no explicit provision norms exist. This was the case in the first and last phase of their three-phase experiment. However, when norms emerged endogenously – implemented through a non-binding vote on the desired contribution level before the beginning of the second phase – negative externalities became internalized and contributions declined. Somewhat surprisingly, positive externalities also reduced provision. The authors attribute this to conditional cooperators who face a twofold risk: Earning lower payoffs not only than free riders within the group but also than passive outsiders. Additionally, there was no spillover of normativity after the contribution norm was abolished again in the last phase. It continued to only matter indirectly since it has induced some favorable experiences.

The experimental design of Engel and Rockenbach (2009) was subsequently refined to investigate two aspects in greater depth. Both follow-up studies continued to examine the effect of externalities on cooperation, but they do so from different angles. Engel and Zhurakhovska (2014) focused on the apparent lack of restraint in imposing harm on passive bystanders, whereas Engel and Rockenbach (2014) took a closer look at the role of giving people a voice through voting.

Engel and Zhurakhovska (2014) simplified the original design to clarify what drives conditional cooperation under negative externalities. A passive third party was harmed whenever at least one of two active players cooperated. By varying the intensity of this harm and eliciting participants' expectations about others' cooperation, they separated cognitive from motivational influences. Their results show that cooperation declined as outsider harm intensified, and this decline could not be explained by inequity aversion or reciprocity. Instead, guilt aversion was the central mechanism: Insiders experience discomfort from jointly harming outsiders, reducing their willingness to cooperate. The findings also suggest that greater strategic uncertainty heightens insiders' sensitivity to harming outsiders.

Engel and Rockenbach (2014) use a design similar to the first two

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

phases of Engel and Rockenbach (2009), but add a control treatment without the possibility of voting. This allows them to distinguish voting effects from mere restart effects. Their design enables them to study to what extent giving people a voice helps resolve the normative conflict that arises when a social dilemma is embedded in a broader social context. The results illustrate both the potential and the limits of such institutional voice. When no externalities were present, voting raised cooperation, showing that even a mere recommendation could coordinate insiders. When externalities were negative, voting lead to lower contributions, thereby successfully internalizing outsider harm. Yet when externalities were positive, voting failed to produce higher contributions. Although groups overwhelmingly voted for high provision, insiders complied only weakly, largely due to adverse payoff comparisons similar to those found in Engel and Rockenbach (2009): Group members were reluctant to contribute when outsiders received windfall gains.

Taken together, these findings on the role of passive outsiders underscore the importance of institutional environments that protect contributors from being exploited, particularly when public-good provision generates positive externalities that allow outsiders to benefit more than insiders. Conversely, when policymakers aim to protect global rather than local public goods, heavy-handed intervention may not be required to shield those harmed by negative externalities: Guilt aversion already helps minimizing harmful spillovers. However, in larger groups, institutions that give outsiders a voice and enable coordination on lower contribution rates can help reduce harmful effects.

IV. Nudges

Nudges do not formally alter the actions or consequences available to contributors, but are interventions that aim to steer behavior in a particular direction while preserving freedom of choice (Hertwig & Grüne-Yanoff, 2017). Therefore, policymakers and researchers have been drawn to exploit their potential. Although critiques and caveats have accumulated over time, nudges remain a widely studied tool because they may offer, among other benefits, a way to encourage contributions to public goods without requiring drastic institutional intervention.

Engel and Kurschilgen (2020) investigate whether asking individuals to articulate their own normative goals can help increase contributions in

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

social dilemmas. In their experiments, they contrast two framings of this nudge: Participants were asked either what every group member should optimally contribute or what they should contribute at minimum. In their first study, the "minimum" framing significantly increased cooperation and delayed the typical decline of cooperation across rounds. This effect was traced to a dual mechanism: self-set standards shaped participants' own willingness to contribute (a self-image channel) and influenced their expectations about others' choices (an indirect belief channel).

However, two later replication studies reveal the limits of articulating own normative goals. With nearly identical designs, but in contexts where baseline cooperation was already higher, neither the "minimum" nor the "optimal" framing produced a measurable effect on contributions. Importantly, the psychological process remained intact and stated norms still predicted individual behavior.

Engel and Kurschilgen (2020) thus caution that nudges built on self-set normative standards are highly context dependent: They may work when cooperative behavior is fragile, but lose traction when norms of cooperation are already entrenched. Their findings nuance the optimism surrounding behavioral policy tools, showing that although self-image can be a powerful motivator, the effectiveness of such nudges depends critically on whether the social dilemma was sufficiently pronounced prior to the intervention.

V. Heterogeneity in Cooperativeness: Individual Differences and Underlying Motives

Even in similar environments, cooperation is rarely uniform: Individuals differ systematically in their cooperativeness. A research line by Christoph Engel and colleagues has provided key insights into the dispositional foundations and underlying motives of cooperation behavior, highlighting how cooperation arises from the psychologically meaningful interplay of situational cues and individual characteristics.

A particularly dominant motive underlying (lacking) cooperation is the aversion to being taken advantage of: People are willing to cooperate, but conditional on mutual cooperation. Negative experiences, in turn, can trigger a disproportionate decline in cooperation (Engel & Rockenbach, 2024). Engel and Zhurakhovska (2014) argue that, based on defining environmental features of the prisoner's dilemma – most notably the risk of

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

exploitation alongside the possibility of maximizing joint gains – cooperation is best understood as the joint outcome of multiple motives, not of a single one alone. By taking efficiency concerns, conditional cooperation, fear, and greed into account, they show that these motives predict cooperation most reliably when estimated together, as some only become predictive when controlling for others. Through this lens, the prisoner's dilemma reflects a dynamic combination of (pro-)social preferences and interpersonal risk calculus that explain when cooperation becomes, for the individual, worth taking.

Importantly, these motives help explain not only when people choose to cooperate in single (i.e., one-shot) interactions, but more importantly so across repeated interactions. Engel and Rockenbach (2024) revisit one of the most robust findings in experimental economics: The gradual decline of contributions in repeated public good games. Building on Fischbacher and Gächter's (2010) influential claim that this decay reflects an inherent imperfection in conditional cooperation, they reanalyze the original data and introduce a more nuanced interpretation. Classifying participants into selfish free-riders (short- or far-sighted) and conditional cooperators, they show that the latter group actually tracked its own beliefs about others' contributions remarkably closely. The apparent imperfection was largely attributable to free-riders. Conditional cooperators only deviated from their beliefs after experiencing disappointment – when others contributed less than expected.

A series of simulations confirmed that disappointment asymmetry, rather than inherent imperfection, explains much of the observed decline. Conditional cooperators adjusted more strongly to negative than to positive experiences, and their behavior was further destabilized when even a single free-rider was present in the group. The analysis implies that cooperation is precarious not because conditional cooperators are insincere, but because they are vulnerable to the corrosive impact of free-riding and noisy behavior. Policy implications follow: Broad-based nudges may be less effective than targeted interventions aimed at preventing free-riding or mitigating its visibility. By highlighting the asymmetry of reactions, Engel and Rockenbach refine our understanding of conditional cooperation and its limits, and underscore how fragile collective action becomes once negative experiences enter the social learning process.

Desmet and Engel (2021) follow up on the motivation to conditionally cooperate in the context of rule compliance: They examine how compliance is influenced by the behavior of others. Using arbitrary, unen-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

forced upper bounds in a real-effort task, they show that rule compliance became stronger when group members adhered to the rule, creating a robust form of behavioral dependence that resembles conditional cooperation. Unlike cooperation behavior in social dilemmas, however, conditional rule adherence appeared to be buffered against breakdowns, as each additional rule-abiding group member reinforced rule adherence. Conditionality was strongest among marginal rule followers, whose compliance was most strongly promoted by positive social information when their basic propensity to follow rules was low.

The findings highlight a deontological component in cooperative behavior: Individuals consider compliance to be inherently appropriate, but express this more consistently in their behavior when others also adhere to the rules. This dynamic complements multi-motive models of cooperation by indicating that rule following draws on motives similar to those underlying cooperation, albeit in a different and partly attenuated manner.

VI. Low-cost Cooperation and Deliberate Ignorance

People are generally willing to share parts of their resources (Engel, 2011), but cooperation particularly flourishes when costs are low, as shown in research on social mindfulness and everyday prosocial behavior (Van Doesum et al., 2025).

Engel and Van Lange (2021), for instance, demonstrated that people are generally socially mindful – that is, considerate of others' choice options and outcomes – when the costs are minimal. Yet even small increases in cost lead to a rapid decline in socially mindful behavior. Importantly, cost sensitivity already becomes visible in patterns of (reduced) information search and attention allocation: Offer et al. (2024) found that a substantial share of participants deliberately ignored information about unfairness when doing so spared them the burden of inflicting costly punishment on norm transgressors. This phenomenon illustrates a broader principle: Deliberate ignorance can serve as a functional strategy that protects individuals from motivational conflict, cognitive overload, or normative obligations (Hertwig & Engel, 2016). In this sense, choosing *not* to know can be understood as a form of resource management – even though actively seeking and integrating information is only minimally costly (Engel & Hertwig, 2021). Moreover, deliberate ignorance is not confined to the individual level. Institutional and legal settings often create or regulate

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

opportunities for ignorance to balance competing demands of efficiency, accountability, and fairness, thereby shaping when knowledge is required and when ignorance is tolerated or even promoted (Teichman et al., 2021).

Taken together, these findings indicate that the costliness of action shapes cooperation behavior itself, while deliberate ignorance can serve as a way to avoid the additional costs that may arise from responding to missing or insufficient cooperation – for instance, the psychological and material burdens of punishing norm violations. In this sense, choosing not to know can function as a protective strategy in the face of potentially costly reactions to others' non-cooperation.

C. Conclusion and Outlook

Across nearly three decades of research, Christoph Engel has shown that cooperation reflects a finely balanced, sensitive interplay of psychological motives and situational cues. And while no single mechanism can fully solve the problem of collective action, Christoph Engel's research has profoundly refined our understanding of when cooperation is possible, when it fails, and which levers we can use to strengthen the common good.

Naturally, the question remains whether the puzzle of cooperation will ever be completely solved. Earlier in his career – then still deeply immersed in the challenges of waste management – Christoph Engel addressed a very tangible social dilemma: He demonstrated that even something as mundane as separating household waste is not merely a technical matter, but a fundamentally social one. With hindsight, this early work reads almost like a prologue to his later insights: Even the most elegant system works only when individuals understand, expect, and trust that others will do their part. From the garbage bin at the curb to the repeated public-good game in the lab, a clear thread connects Christoph Engel's work: A persistent curiosity about what makes humans cooperative beings – and what gets in their way.

More recently, this curiosity has turned toward a new domain: Large language models, whose potential contributions include advancing our understanding of the persistent under-provision of public goods. If, one day, a large language model manages to solve this problem, it will almost certainly have been informed and guided by Christoph Engel's work.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

References

Camera, G., & Casari, M. (2009). Cooperation among strangers under the shadow of the future. A*merican Economic Review*, *99*(3), 979–1005. https://doi.org/10.1257/aer.99.3.979.

Desmet, P. T. M., & Engel, C. (2021). People are conditional rule followers. *Journal of Economic Psychology, 85*, 102384. https://doi.org/10.1016/j.joep.2021.102384.

Engel, C. (2004). Social dilemmas, revisited from a heuristics perspective. *Preprints of the Max Planck Institute for Research on Collective Goods,* No. 2004/4. https://dx.doi.org/10.2139/ssrn.539442.

Engel, C. (2009). Competition as a socially desirable dilemma – Theory vs. experimental evidence. *Preprints of the Max Planck Institute for Research on Collective Goods,* No. 2009/24. http://dx.doi.org/10.2139/ssrn.1444270.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7.

Engel, C. (2014). Social preferences can make imperfect sanctions work: Evidence from a public good experiment. *Journal of Economic Behavior & Organization, 108*, 343–353. https://doi.org/10.1016/j.jebo.2014.02.015.

Engel, C. (2015). Scientific disintegrity as a public bad. *Perspectives on Psychological Science, 10*(3), 361–379. https://doi.org/10.1177/1745691615577865.

Engel, C., & Hertwig, R. (2021). Deliberate ignorance: Present and future. In R. Hertwig & C. Engel (Eds.), *Deliberate ignorance: Choosing not to know* (pp. 317–332). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/13757.003.0004.

Engel, C., & Hippel, S. (2017). Experimental social planners: Good natured, but overly optimistic. *Preprints of the Max Planck Institute for Research on Collective Goods*, No. 2017/23. https://dx.doi.org/10.2139/ssrn.3074350.

Engel, C., Kube, S., & Kurschilgen, M. (2021). Managing expectations: How selective information affects cooperation and punishment in social dilemma games. *Journal of Economic Behavior & Organization, 187*, 111–136. https://doi.org/10.1016/j.jebo.2021.04.029.

Engel, C., & Kurschilgen, M. (2015). The jurisdiction of the man within: Introspection, identity, and cooperation in a public good experiment. *Preprints of the Max Planck Institute for Research on Collective Goods*, No. 2015/1. https://dx.doi.org/10.2139/ssrn.2544579.

Engel, C., & Kurschilgen, M. (2020). The fragility of a nudge: The power of self-set norms to contain a social dilemma. *Journal of Economic Psychology, 81*, 102293. https://doi.org/10.1016/j.joep.2020.102293.

Engel, C., & Rockenbach, B. (2009). We are not alone: The impact of externalities on public good provision. *Preprints of the Max Planck Institute for Research on Collective Goods*, No. 2009/29. https://dx.doi.org/10.2139/ssrn.1463259.

Engel, C., & Rockenbach, B. (2014). Give everybody a voice! The power of voting in a public goods experiment with externalities. *Preprints of the Max Planck Institute for Research on Collective Goods*, No. 2014/16. https://dx.doi.org/10.2139/ssrn.2519479.

Engel, C., & Rockenbach, B. (2024). What makes cooperation precarious?. *Journal of Economic Psychology, 101*, 102712. https://doi.org/10.1016/j.joep.2024.102712.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Engel, C., & Van Lange, P. A. (2021). Social mindfulness is normative when costs are low, but rapidly declines with increases in costs. *Judgment & Decision Making, 16*(2), 290–322. https://doi.org/10.1017/S1930297500008585.

Engel, C., & Zhurakhovska, L. (2011). Oligopoly as a socially embedded dilemma. An experiment. *Preprints of the Max Planck Institute for Research on Collective Goods*, No. 2011/1. https://dx.doi.org/10.2139/ssrn.1750094.

Engel, C., & Zhurakhovska, L. (2014). Conditional cooperation with negative externalities – An experiment. *Journal of Economic Behavior and Organization, 108*, 252–260. https://doi.org/10.1016/j.jebo.2014.10.005.

Engel, C., & Zhurakhovska, L. (2016). When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives. *Applied Economics Letters, 23*(16), 1157–1161. https://doi.org/10.1080/13504851.2016.1139672.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, *100*(1), 541–556. https://doi.org/10.1257/aer.100.1.541.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9.

Friedman, D., & Oprea, R. (2012). A continuous dilemma. *American Economic Review*, *102*(1), 337–363. https://doi.org/10.1257/aer.102.1.337.

Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science*, *11*(3), 359–372. https://doi.org/10.1177/1745691616635594.

Hertwig, R., & Engel, C. (2021). Homo ignorans: Deliberately choosing not to know. In R. Hertwig & C. Engel (Eds.), *Deliberate ignorance: Choosing not to know* (pp. 3–18). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/13757.003.0004.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12*(6), 973–986. https://doi.org/10.1177/1745691617702496.

Offer, K., Mischkowski, D., Rahwan, Z., & Engel, C. (2024). Deliberately ignoring inequality to avoid rejecting unfair offers. *Communications Psychology, 2*(48), 1–13. https://doi.org/10.1038/s44271-024-00093-6.

Teichman, D., Talley, E., Egidy, S., Engel, C., Gummadi, K. P., Hagel, K., Lewandowsky, S., MacCoun, R. J., Utz, S., & Zamir, E. (2021). Institutions promoting or countering deliberate ignorance. In R. Hertwig & C. Engel (Eds.), *Deliberate ignorance: Choosing not to know* (pp. 275–298). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/13757.003.0004.

Van Doesum, N. J., Van Lange, P. A., Mischkowski, D., & Van Dijk, E. (2025). Social mindfulness and the SoMi Paradigm: A decade of research on low-cost prosociality. *European Review of Social Psychology*, 1–51. https://doi.org/10.1080/10463283.2025.2457918.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Of Festschrifts and Preprints
Exploring Understudied Academic Genres

Hanjo Hamann

A. Background and Motivation

How better to celebrate a pioneer of quantitative legal studies than by quantitatively studying his own legacy? To be sure, the following exploration (like all exploratory statistics) can only paint one view of the cathedral that is the Engel legacy. Yet, this particular view may be of general interest beyond the professional network whose members assemble in this festschrift.

But first, a few thoughts about the festschrift as such.

I. Of Cemeteries and Promenades

Festschrifts are a German thing. There is no proper English word for them, so the English language in 1898 simply ran with the German term.[1] Even one century later though, not all German law professors were fond of *festschrifts*. A German émigré with some outside perspective called them "mass burials of erudition" and "cemeteries of research",[2] and a lawyer-and-economist emeritus beheld *festschrifts* "as graveyards on which original ideas, coherent thoughts, and masterpieces of style are laid to rest".[3]

Coincidentally it was in just the (few) years between these quotes that a new-fangled research unit on law-and-economics was established in the

1 See www.etymonline.com/word/festschrift.

2 *Stiefel*, JZ 1995, 613: "Der Festschriftler [...] hört die Kassandrarufe: Massengrab der Gelehrsamkeit. Er merkt, sein Beitrag ist im Leichenwagen auf dem Wege zum Friedhof der Wissenschaft."

3 *Wolfram Müller-Freienfels*, cited by *Löwisch* (JZ 1998, 946) and *Münch* (NJW 2000, 3253, 3256 footnote 28): "Festschriften sind wie Friedhöfe, auf denen originelle Ideen, schlüssige Gedanken und stilistische Meisterwerke ruhen."

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

former German capital (*Projektgruppe Recht der Gemeinschaftsgüter*).[4] Its director did entomb quite a few coherent thoughts in the fertile soil of *festschrifts*.[5]

Yet, even before I became his student – and long before I lowered my first coffin of ideas into a *festschrift* – I had already learned to love unearthing *festschrift* articles. Not only do they serve more general interests than most law journals: They also make room for (and almost demand) personal anecdote, creative style, and – quite simply – academic gossip. I gladly comply with tradition by releasing a personal anecdote into the realm of gossip: Throughout my years of research, whenever I requested a library to digitize some *festschrift* article, I invariably perused the volume's entire table of contents and exhausted the 10 % copyright limit on digital library reproduction.[6] There was always something else that piqued my interest. I thus perceive *festschrifts* as a strip mall of sorts, or a French shopping promenade: A quick succession of enticingly dressed windows into academia's haute-couture. Very different styles, some even artsy, but all quite intriguing.

Whence these diametrically opposed views of the *festschrift*? Why did scholars in the 1990s consider them morbid necropoles, while I see them as buzzing bazars? I can only guess that my particular position is that of a digital almost-native at a time when the table of contents of virtually any German book is easily available through the National Library (DNB) catalogue,[7] while free digitization-on-demand services allow me to obtain a digital copy of any book chapter within a working day. Even though most *festschrifts* are still not available online (unlike this one), or even

4 On the institute's origin, see *Kazemi/Henning*, Chronik der Kaiser-Wilhelm-/Max-Planck-Gesellschaft zur Förderung der Wissenschaften. 1911–2011, 2011, p. 700.

5 *Christoph Engel* contributed at least ten such celebratory contributions: For Dietrich Rothoeft (1994), Ernst-Joachim Mestmäcker (1996 and 2003), Klaus J. Hopt (2010), Wernhard Möschel (2011), Martin Bullinger (2011), Paul Kirchhof (2013), Martin Morlok (2019), Christine Windbichler (2021), and Michael Faure (2025). The first two *festschrifts* he co-edited himself.

6 Sec. 60e (5) of the German Copyright Act (UrhG): "In response to individual orders, libraries may, for non-commercial purposes, transmit reproductions of up to 10 per cent of a published work to users […]."

7 Or through dedicated DNB staff, who digitize missing TOCs for free. Most notably, *Kristina Knull-Schlomann* has been my unfailing pillar to sourcing literature (*festschrift* or otherwise) for the past five years. Let this footnote be a token of gratitude for some 150 emails in which she confirmed the retro-digitization of hundreds of TOCs since our first exchange in October 2020.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

sufficiently indexed in article databases, I go article-shopping in them way too often to think of them as burial grounds.

This was all different in the 1990s when the *festschrift-as-cemetery* theorem (as we shall call it) was prominent. It had sprung from two observations: (1) that there were "masses, an inflation of *festschrifts*",[8] while (2) "legal scholarship and legal practice pass over many a festschrift contribution because the festschrift is simply not available to them".[9] This brings us to this contribution's topic, which is dear to the hearts of both the author and the dedicatee:[10] Digital access to scholarly writing.

II. Of Electronic Bulletins and Neverprints

The greatest driver for improved access to scholarly articles has, of course, been the Internet. Scientists were among its early adopters when they started sharing drafts electronically in the 1980s – first through e-mail lists, then "electronic bulletin boards", which preceded our present-day "repositories". The most influential of these (arXiv) started as the "xxx.lanl.gov physics e-Print archive" in Los Alamos in 1991.[11] In the same year, a German economist at the University of Surrey began developing "net-based services for economists" (NetEc),[12] which would become "Research Papers in Economics" (RePEc) just as the aforementioned *Projektgruppe* was established in Bonn. Concurrently with NetEc, financial economists in the US started the "Financial Economics Network" (FEN) in 1993, which would grow into a "Social Science Research Network" (SSRN) the next year. Others followed suit, and now a Wikipedia "list of preprint repositories" lists some eighty databases with millions of arti-

8 *Stiefel*, JZ 1995, 613: "Der Festschriftler wird gewahr, daß er Teilnehmer an einer Masse, einer *Inflation* von Festschriften ist.", citing apocryphal data on "Festschrift-Hypertrophie: 1864–1944: 3 pro Jahr; 1945–1961: 10; 1987: 45; neunziger Jahre um 100.".

9 *Löwisch*, JZ 1998, 946: "Rechtswissenschaftliche Diskussion und vor allem juristische Praxis gehen vielfach an Festschriftbeiträgen vorbei, [...] weil ihnen die Festschrift einfach nicht zur Verfügung steht."

10 See, e.g., *Engel*, Give the Journals Back to the Scientists, JITE 160 (2004), 35–38.

11 See *Taubes*, Publication by Electronic Mail Takes Physics by Storm, Science 259 (1993), 1246–1248.

12 See *Krichel*, About NetEc, with special reference to WoPEc, CHEER 11 (1997).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

cles.[13] Institutional repositories have joined their ranks: Some festschrift contributions by *Christoph Engel* were pre-published as drafts in the Publication Repository of the Max Planck Society (MPG PuRe).[14]

Even though "e-prints" (as arXiv called them) or "preprints" (as SSRN calls them) are still a quite recent phenomenon – and not widely adopted (or even known) in German legal academia – they are in some sense already outdated. The term "… print" may seem anachronistic at a time when much research really *never* gets printed – or even published.[15] They might better be called "research papers" or "working papers" instead (as RePEc does), as illustrated by the subsequent development of the Bonn *Projektgruppe*: While in 1998 it started publishing its own digital series of "Preprints aus der Max-Planck-Projektgruppe", this was renamed twenty years later to "Discussion Papers of the Max Planck Institute" – reflecting not just a shift in working language and incorporation status, but also a broader development in the objective of research dissemination.

III. Of Preprint Bibliometrics and the Grey Beyond

Like any evolving social practice, "preprinting" is being studied empirically. Some have found that the practice predates the internet, with the first large-scale preprint platforms in biology existing as early as the 1960s.[16] Today, across academic disciplines, bibliometricians find that "the number of preprints has exponentially increased 63 times in 30 years, although it only accounts for 4 % of research articles."[17] A comprehensive

13 Available at en.wikipedia.org/wiki/List_of_preprint_repositories. Apart from the three already mentioned, the list contains four newer entries with at least 1 mio. articles: HAL Open Archives (est. 2001), Science Open (est. 2013), OSF Preprints (est. 2017), and Synthical (est. 2023).

14 Of the ten contributions listed in note 5, this applies to at least three of the most recent ones honoring *Hopt* (https://hdl.handle.net/11858/00-001M-0000-0028-6DA8-C), *Windbichler* (https://hdl.handle.net/21.11116/0000-0005-A6A1-9) and *Faure* (https://hdl.handle.net/21.11116/0000-000F-AA4B-E). The repository was only established in May 2009.

15 Research indicates that as few as 41 % of preprints might ever get published, see *Xie/Shen/Wang* (Fn. 17).

16 On the NIH's "Information Exchange Groups" (1961–1967) see *Cobb*, The prehistory of biology preprints, PLoS Biol 15 (2017), e2003995.

17 *Xie/Shen/Wang*, Is preprint the future of science?, IW3C2 Proceedings 2021.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

review of preprint research is impossible within the confines of this contribution, yet two general study formats seem to emerge.

One explores the *causes* (motivations) for researchers to circulate research ahead of publication: "Preprinting benefits include rapid dissemination of academic work, open access, establishing priority or concurrence, receiving feedback, and facilitating collaborations."[18] This research has revealed another important catalyst (apart from the Internet) for a recent surge in preprint dissemination: The COVID pandemic.[19]

The other strand of research explores the *effects* (reception) of preprints: Do scholars use and cite preprints like published research?[20] Is the evidence reported in preprints robust, or does it frequently get revised after peer review?[21] Which cues do researchers use to assess preprint credibility?[22] Can unethical research practices be identified from a linguistic analysis of preprints?[23] (How) Do journalists rely on preprints?[24]

Despite such studies, preprints remain largely uncharted territory. They are generally considered "graue Literatur":[25] *Grey* literature which is neither *white* (published conventionally and catalogued by libraries) nor

18 *Sarabipour et al.*, On the value of preprints, PLoS Biol. 17 (2019), e3000151.

19 *Glymour et al.*, Counterpoint: Preprints and the Future of Scientific Publishing, Am. J. Epidem. 192 (2023), 1043.

20 *Bertin/Atanassova*, Preprint citation practice in PLOS, Scientometrics 127 (2022), 6895: "authors make use of different lexical content when citing preprints compared to the rest of citations."

21 *Nelson et al.*, Robustness of evidence reported in preprints during peer review, Lancet Glob. Health 10 (2022), e1684–87: "the correlation between estimate values before and after review was high (0.99) and there was no systematic trend."

22 *Soderberg/Errington/Nosek*, Credibility of Preprints, Royal Soc. Open Sci. 7 (2020), 201520: "cues related to information about open science content and independent verification of author claims were rated as highly important [. . .] peer views and author information were rated as less important."

23 *Teixeira da Silva*, "Tortured phrases" in preprints, Curr. Med. Res. & Opin. 39 (2023), 785: "Whereas some 'tortured phrases' might represent simple mistranslations, in other cases, an abundance of such terms in a single preprint might reveal a more serious ethical issue, such as the undeclared use of a paper mill or an unprofessional editing service."

24 *Fleerackers at al.*, Science in motion, PLoS ONE 17 (2022), e0277769: "journalists approach the decision to cover preprints as a careful calculation [. . .] journalists also reported barriers to covering preprints, as many felt they lacked the expertise or the time required to fully understand or vet the research."

25 The term's origin is unclear, but surely older than the first reference I found: *Ullner*, Afrika-Dokumentation, Africa Spectrum 4 (1969), 38, 39, 42.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

black (publically unavailable and uncatalogued, like corporate R&D);[26] there is now a whole ecosystem around it.[27]

One type of grey literature is institutional literature such as the aforementioned "Discussion Papers of the Max Planck Institute for Research on Collective Goods Bonn". I am not aware of previous research on this body of 27 years of contiguous research. So, as a contribution both to preprint bibliometrics and to charting the *Engel* legacy, I set out to explore the preprint series of his institute since 1998.

B. Methods and Data Preparation

I. Data Source

I downloaded the metadata of the preprints and discussion paper series from the Max Planck Society's institutional repository (PuRe) on 28 January 2025. Each paper of the series was recorded with its year and document number (in the format [YYYY]/[NN]), authors, title, keywords, abstracts, and URL of full text pdf. There were no metadata on the papers' discipline of origin, which to code manually would have been unreliable at best, and epistemologically unjustifiable at worst. So all papers were treated as coming from the same general field of research, i.e., *collective goods research*.

This dataset consisted of 605 repository entries. To prepare them for analysis, data were first transformed into a Stata dataset, and tested for internal consistency. I discussed any inconsistencies with the MPI staff member in charge (*Marc Martin*) whose support I gratefully acknowledge. Several errors in the repository were corrected in this process, and previously unavailable full text documents were reuploaded.

---

26 See GDR librarian *Schmidmaier*, Ask No Questions and You'll be Told No Lies, LIBRI 36 (1986), 98, 101: "'Grey literature' is that in the libraries which is not 'white' (available, catalogued, classified) or 'black' (not available, unknown, not obtainable, secret)."

27 See the Grey Literature Network Service "GreyNet" with its International Journal on Grey Literature simply called "The Grey Journal", and the "International Conference on Grey Literature" of at least 25 instalments.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Completeness and Deduplication

I ensured that data were complete by comparing a list of the paper numbers with an auto-generated list of running numbers for each year. Several gaps were identified, of which only three remained after data preparation: Preprints 2007/23, 2011/27 and 2011/29 were never published, their slot had merely been reserved for one of the institute's directors (*Martin Hellwig*). In addition, preprint 2007/11 had been replaced by two documents: 2007/11a and /11b. Conversely, three preprints were published outside the running count: 2004/1(a), 2007/4b, and 2007/6b.

In addition to ensuring overall completeness, I also tested whether metadata for each entry was complete. Whereas some data (like title and authors) were available for every preprint, others had not been recorded consistently and could not easily be interpolated or reinserted. In order to avoid biases, I did not use incomplete data in subsequent analyses. This included upload date (missing in 84 % of records), keywords (missing in 35 % of records), and abstract (missing in 16 % of records).

I then checked for metadata duplicates. There were none in the strict sense, but in a few instances, the repository entry noted a preprint being "updated by" a different one from the same series. In these cases, the earlier paper was treated as a duplicate, so as to include only the latest version in some analyses. I proceeded likewise for all papers with identical titles (all but one of which also had identical authors).[28] Some of these papers acknowledged, in their full text, the earlier version which they superseded, others did not.

Overall, I labelled 13 papers as duplicates.[29] Unfortunately there was no easy way to verify whether other papers superseded one another under altered titles, or whether even similar-sounding titles[30] indicated duplicates. This might be possible by semantically comparing abstracts with one another, or even by computing a similarity score for pairs of full text documents; both were infeasible for this initial exploration.

---

28 Paper 2019/9 had one author less than its previous version 2017/18.

29 See footnotes 34, 35 below.

30 Compare, for instance, the title of 2008/29 ("Competition Policy and Sector-Specific Regulation for Network Industries") with 2018/7 ("Competition Policy and Sector-Specific Regulation in the Financial Sector"), or 2014/9 ("Financial Stability, Monetary Policy, Banking Supervision, and Central Banking") with 2015/10 ("Financial Stability and Monetary Policy").

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. Full Text Acquisition

For most repository records, a full text had been uploaded. For others, their metadata included external URLs, albeit inconsistently: Only 82 % of records featured an external URL, which in three quarters of the cases (457) was of the form *http://www.coll.mpg.de/pdf_dat/[year]_[##]online.pdf*. This uniform scheme allowed for easy interpolation, which worked in every instance except five.[31] For these I found URLs on different domains, except in the earliest case (1999/9), whose author (*Christoph Knill*) kindly provided another document version which he assured me was identical to the one submitted to the present paper series.

After assembling the URL list, I downloaded each document and extracted their page counts, which I used for subsequent analyses. I did not use the preprints' contents yet due to the difficulty of extracting clean contents from pdf files with vastly different layouts. While they might be used for a variety of semantic comparisons and interesting measures (discipline-specific vocabulary, number of references, existence or absence of economic theory, statistical tests, etc.), this was beyond the scope of an initial exploration.

C. Data Exploration

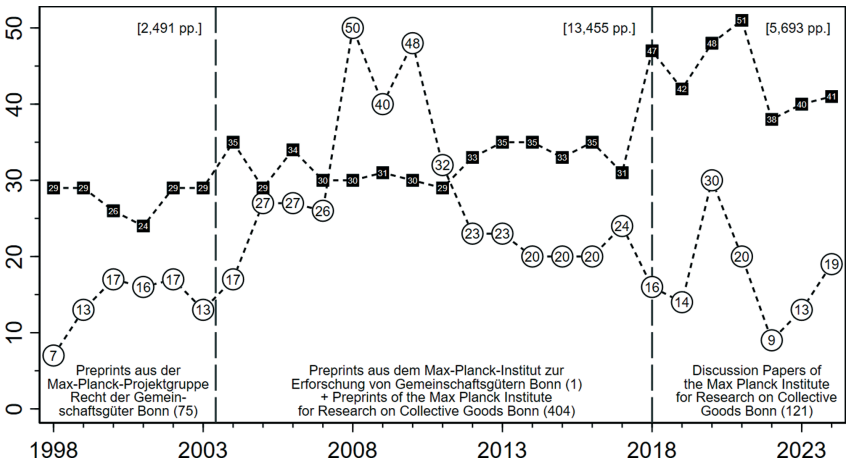
I. Distribution: 21,639 pages over 26 years

As a first exploratory analysis, I show the number and length of papers over time (including duplicates) in graph 1 below.

Vertical lines divide the graph into three phases: In the first (up to 2003/5), 75 documents of just under 2,500 pages had been called "Preprints aus der Max-Planck-Projektgruppe Recht der Gemeinschaftsgüter Bonn". Then came a single document (2003/6) under the series title "Preprints aus dem Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern Bonn" before the series was again renamed to "Preprints of the Max Planck Institute for Research on Collective Goods Bonn" for the longest part of its existence (some 13,500 pages until 2017/24), and finally to "Discussion Papers of the Max Planck Institute for Research on Collective Goods Bonn" (since 2018/1).

31 Preprints 1999/9, 2008/36, 2009/25, 2010/20 and 2012/5.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -



Graph 1: Annual number of preprints (encircled: ① etc.) and their median length (■) across all three phases of the paper series. The number before "pp." denotes total page count in each of the phases.

These phases coincided with major turning points in the institute's evolution: Around the first transition in 2003, it turned from a research group (*Projektgruppe*) into an institute, just as political scientist *Adrienne Héritier*, co-director since April 1998, left the institute in March 2003. The second transition at the start of 2018 came shortly after the accession of co-director *Matthias Sutter* in August 2017.

Consider the development of the number of preprints: Encircled numbers illustrate that peak preprint productivity was in 2008–2010, with another (albeit local) maximum in 2020. The smaller black squares show that median paper length hovered consistently at 30 ± 5 pages,[32] but increased by half (to 45 ± 5 pages) at the beginning of the third phase – likely due to the new research group.[33]

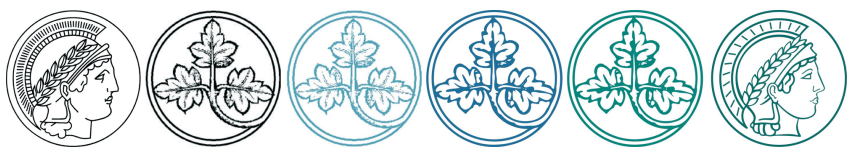
32 The extremes were rather interesting: Two of the three shortest papers were single-authored by Martin Hellwig (6 pages in 2023/5, 7 pages in 2008/45), both of the longest papers were co-authored by Sebastian Schneider (129 pages in 2023/3, 121 pages in 2023/6). As the document numbers indicate, the shortest and both longest papers were published in 2023 within a sequence of merely 4 papers.

33 Anecdotal evidence: 8 of the 12 longest papers (all 87+ pages) came from authors in the *Sutter* group.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

These bibliometric explorations neatly complement a historiography of the institute across its phases of productivity. For instance, its physical relocation marks the difference between preprint 2004/05 (which states the institute's address as Poppelsdorfer Alle 45, 53115 Bonn) and preprint 2004/06 (which mentions Kurt-Schumacher-Str. 10, D-53113 Bonn). Interestingly, even three years after the institute was incorporated (and relocated), preprints still cited its URL as http://www.mpp-rdg.mpg.de: Its subsequent (and current) URL http://www.coll.mpg.de was not mentioned until preprint 2006/11.

As a final illustration, the preprints contain qualitative data such as the emblem used on their title pages: Starting with the Max Planck Minerva, the series quickly switched to its own oak leaf emblem across four different color schemes, before returning in late 2020 to the Minerva emblem. Here the paper series comes full circle, as tides may have changed in the Max Planck Society's corporate design rules.



Graph 2: All six emblems used for the series' title pages. From left to right, the emblems were last used, respectively, in papers 1998/04, 2005/21, 2014/18, 2019/11, and 2020/24. The right-most emblem is currently still in use.

II. Topics: 588 titles in 2 languages

In order to explore the research topics covered throughout the paper series, I first removed the 13 duplicates identified earlier – seven of which were shorter[34] and six longer[35] than the successive papers replacing them.

34 By length difference: Papers 2020/26 (45 % of 2023/12), 2021/14 (83 % of 2024/2), 2023/4 (91 % of 2024/8), 2009/25 (94 % of 2011/18), 2010/2 (96 % of 2015/12), 2010/42 (96 % of 2013/23).

35 By length difference: 2007/14 (238 % of 2015/4), 2021/6 (230 % of 2023/6), 2010/20 (207 % of 2011/19), 2017/18 (143 % of 2019/9), 2020/17 (135 % of 2023/3), 2019/14 (116 % of 2020/20), 2005/23 (113 % of 2007/2).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I then coded the language of all titles of the remaining 588 papers, using a naïve classification based on nine common stop words from both German[36] and English,[37] and hand-coding the 68 titles containing none of them. In order to visualize trends in the title vocabulary, I then produced word clouds (where each keyword's font size scales with its frequency) for each of four time periods of six or seven years each.[38]

The 2×4 roster of word clouds illustrates several shifts in the institute's focus (see next page, graph 3).[39]

Firstly, the right-hand column exhibits far greater density and font size variation than the left-hand column, resulting from a greater number of English papers in each of the four periods (53, 186, 150, and 108) than German titles in any of them (30, 44, 11, and 7, respectively). Greater sample size means greater variation between frequency extremes, hence the right-hand words vary more in size than the left-hand ones, where most words only appear once.[40]

Secondly, there is an unmistakable evolution in vocabulary, away from applied and policy issues (1998–2003: *produktverantwortung*, *umweltpolitik*, *european*, *governance*; 2004–2010: *wettbewerbspolitik*, *public*, *information*, *taxation*) towards more basic research (2011–2017: *grundlagenforschung*, *punishment*, *information*, *experiment/al*, 2018–2024: *discussion*, *behavior*, *preferences*, *experiment*). Consider the prominence of *experiments*: Barely visible in the first phase, inflected in the second ("experimental"), prominent in both inflections in the third, and clearly dominant in the fourth. Just seven title words appear at least twice in each of the four eras: *public* (32 titles overall), *law* (32), *social* (26), *european* (19), *regulation* (12), *political* (11), and *court* (8). These terms represent the solid core of the institute's topics since its inception.

36 Specifically: als, das, der, des, die, eine, und, von, zur.

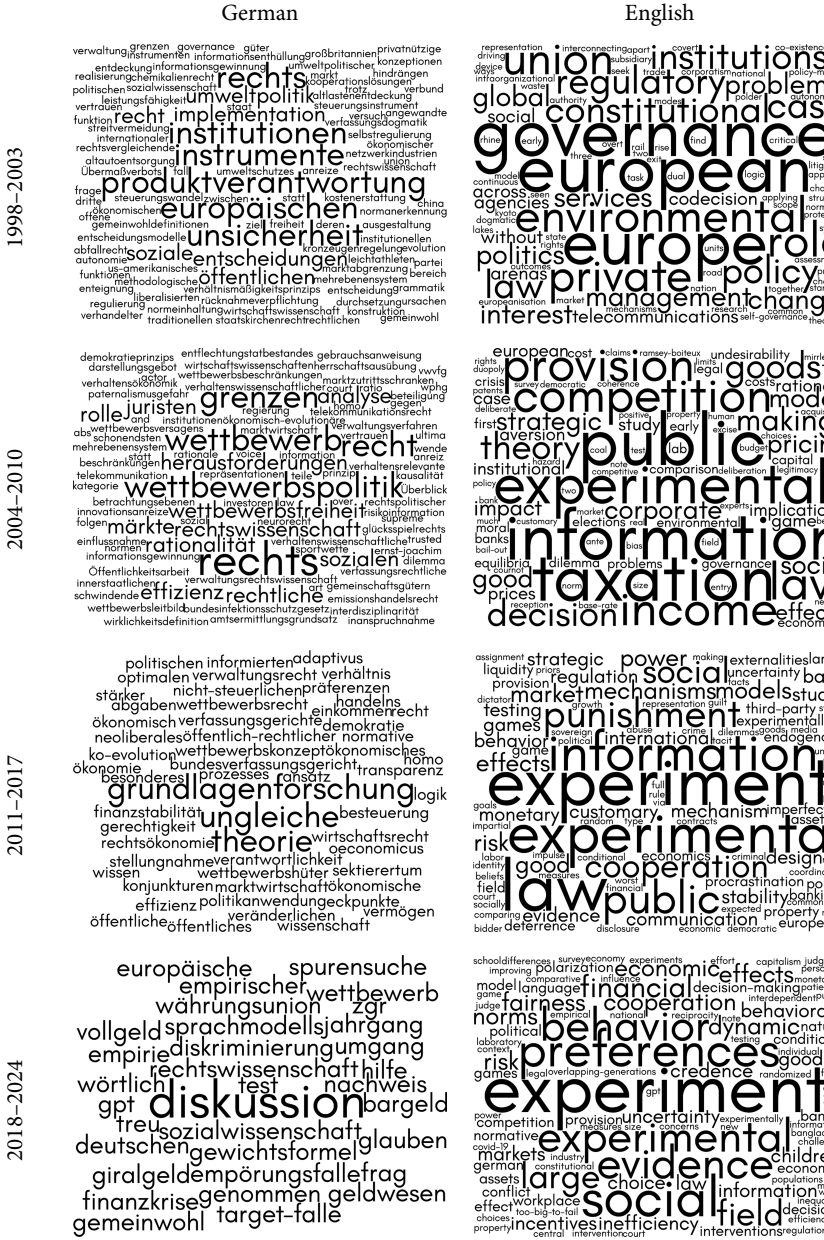
37 Specifically: a, and, for, not, of, on, the, to, with.

38 I thank my colleague *Rima-Maria Rahal* for this suggestion.

39 Graphs created using *free online wordcloud generator* at www.wordclouds.com with direction "Horizontal", drawing method "Circles", min. word size 5, word margin 2, restricted word number.

40 In fact, with just 7 and 11 German texts in the latter two periods, they even fell short of 100 words to plot: Their clouds contain 56 title words in the third, and 33 in the last period.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

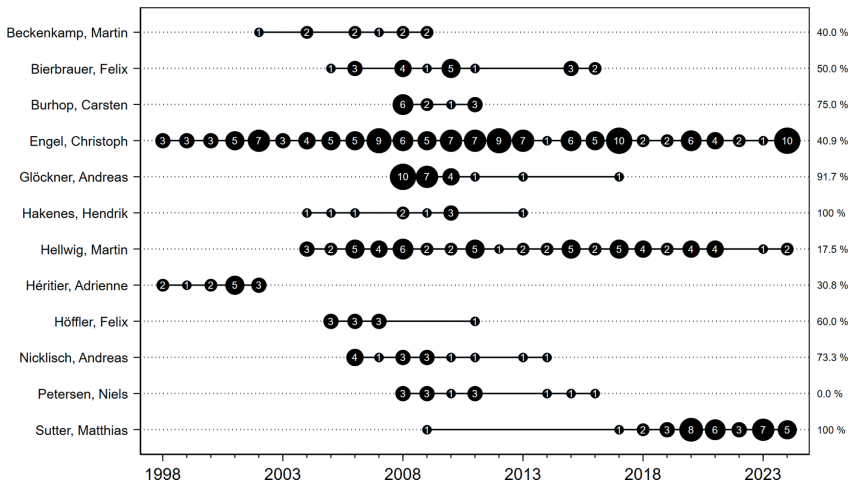


Graph 3: Word clouds of up to 100 most frequent words in paper titles in German (left) or English (right)

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. People: 372 authors in 39 clusters

Lastly, I explored which authors contributed to the series, either alone (in 313 papers, or 53.2 % of the series) or jointly (in 275 papers with 2.7 co-authors on average). Treating all 372 distinct (co-)authors equally, I found that they contributed 2.8 papers on average, and just a dozen authored at least ten papers, shown in the following string-of-beads graph.



Graph 4: Number of papers per year of the most prolific authors in the series, with their overall co-authorship ratio plotted on the right

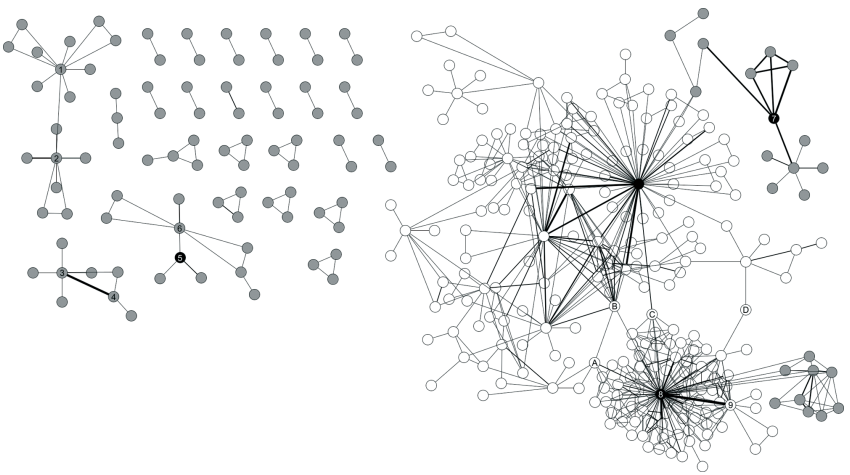
The graph reveals that *Christoph Engel* was the series' most prolific author by far, both in absolute terms (137 papers) and relative to the duration of his tenure (5.1 papers p.a.). His three co-directors (*Hellwig*, *Héritier*, and *Sutter*) also contributed papers throughout their tenure, with *Sutter* even once earlier (in 2009), and *Hellwig* still long after he retired (in 2017). Among the directors, co-authorship rates vary starkly: While *Hellwig* co-authored less than one-fifth of his papers, *Engel* co-authored more than two-fifths, and *Sutter* every single paper. Setting directors aside, a junior group leader (*Glöckner*) was the most prolific contributor (24 papers in total), with a record 10 papers in 2008 alone, the year just after *Christoph Engel* had first set this record.

Another author set a remarkable record: All of *Niels Petersen*'s 13

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

papers in the series were single-authored. Including his runner-up, *Carl Christian von Weizsäcker* (6 single-authored), there were just 30 other contributors who authored in the series exclusively by themselves; none of the others contributed more than two papers.

All other authors, who collaborated at least once, can be visualized as nodes in a network where edges represent co-authorships:



Graph 5: Network of co-authors where edge width scales with frequency of co-authorship between adjacent nodes, black nodes mark institute directors, and grey nodes highlight clusters discussed in subsequent text[41]

Consider first the right-hand side of the graph: It contains three black nodes, marking directors of the institute. Remarkably, no paper in the series had ever been jointly authored by two directors, which creates easily distinguishable sub-clusters:

At the top right, *Martin Hellwig* (#7) is at the center of a rather stable coauthorship pattern (recognizable as thicker lines) with his coauthors *Gizatulina*, *DeMarzo*, *Pfleiderer*, *Admati*, and *Bierbrauer* – with each of whom he shares three or four papers. Bottom right, we find *Matthias Sutter* (#8) centering a tight sub-cluster of 85 authors – almost a "hairball" in network analysis lingo. It is attached to the remainder of the graph

41 Graph created using yEd 3.23.1 with Layout setting "Organic".

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

through only four connectors: *Francesco Feri* (A), *Bernd Irlenbusch* (B), *Bettina Rockenbach* (C), and *Eugenio Verrina* (D). *Sutter* also anchors a smaller cluster of nine authors even further right (in gray) through his coauthors *Walzl*, *Kirchler*, *Stefan*, and *J. Huber*, and he is part of the most stable coauthorship pattern in the entire graph, comprising of six collaborations with *Sebastian Schneider* (#9).

An equally strong collaboration is located at the opposite corner, in the lower left: Another six-fold edge[42] between *Hendrik Hakenes* (#3) and *Isabel Schnabel* (#4), which forms the backbone of one of 25 self-contained mini clusters in the left half of the graph (in gray). Each of these clusters contains 2–16 authors (or 86 authors in total), who are disconnected from the remainder of the graph. At the top left, the biggest of these clusters centers around *Christian Traxler* (#1) and *Carsten Burhop* (#2), and the third somewhat larger cluster contains a black node, marking director *Adrienne Héritier* (#5), whose co-author *Christoph Knill* (#6) was most central in her cluster.

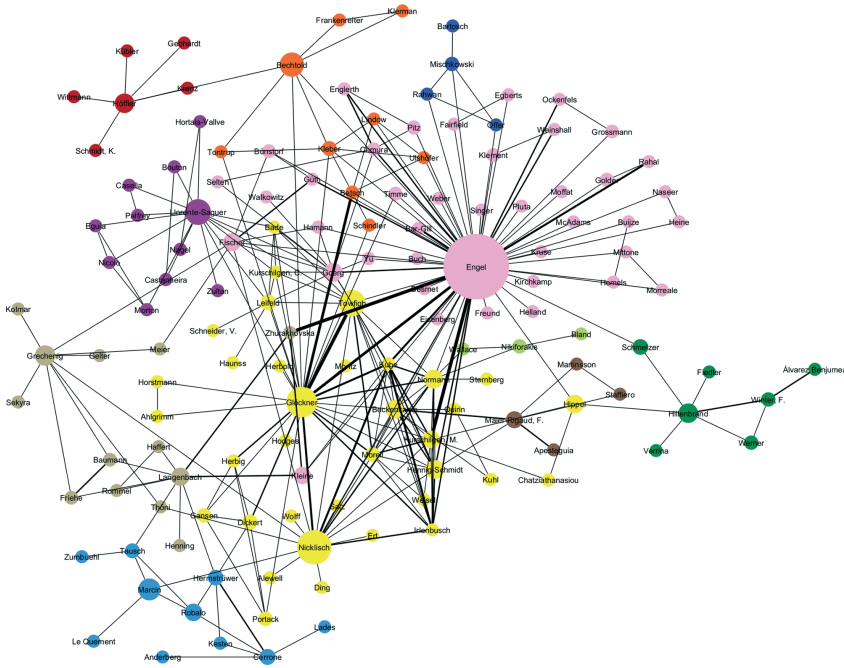
I next turn to the yet-unexplored main section of the major cluster on the right-hand side, centered around the fourth black node – unsurprisingly, *Christoph Engel*. To put this part of the graph into sharper relief, I removed all grayed clusters and cut the sparse connections at A–D. This left 143 authors, or 38 % of the full graph. On these nodes, I ran an edge-betweenness clustering algorithm,[43] which resulted in 11 clusters of between 3 and 41 authors. These clusters are color-coded in the following graph and their nodes resized to scale with their betweenness centrality. In other words: The bigger a node, the more shortest paths between all pairs of nodes run through it.

This graph lets us quickly distinguish the ten authors with the greatest centrality within the Engel sub-network. In ascending order: *Felix Höffler*, *Adrian Hillenbrand*, *Kristoffel Grechenig*, *Isabel Marcin*, *Stefan Bechtold*, *Aniol Llorente-Saguer*, *Emanuel Towfigh*, *Andreas Glöckner*, *Andreas Nicklisch*, and – most centrally – *Christoph Engel*.

42 The only other edges of at least five collaborations are between Glöckner/Towfigh, Engel/M. Kurschilgen and Engel/Zhurakhovska.

43 *Girvan/Newman*, Community structure in social and biological networks, PNAS 99 (2002), 7821–7826. Performed using yEd 3.23.1, Auto-Grouping function with Grouping Type "Natural Clusters" and Quality/Time Ratio 1.0.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -



Graph 6: Network of co-authors connected to *Christoph Engel*, where node size represents betweenness centrality.

Given *Engel*'s centrality, we may borrow a famous parameter from the bibliometry of mathematics (the "Erdős number")[44] to measure *Engel*'s collaborative distance (E) to every other author in the series.[45] We find that 24 % of the entire network (62 authors) directly collaborated with *Engel* at least once (E = 1), while 20 % of the network (50 authors) and 45 % of the network (115 authors) had *Engel numbers* of 2 and 3, respectively. Just five authors in the entire dataset maximised collaborative distance with E = 5,[46] meaning that at least one of their co-authors co-authored

44 See https://en.wikipedia.org/wiki/Erd%C5%91s_number.

45 Except for those who were disconnected from the graph altogether, i.e., 31 exclusive single authors, as well as 86 authors in the self-contained mini-clusters on the left hand side of graph 5.

46 Namely: *Rafael Aigner*, *Pierre C. Boyer*, *Lydia Mechtenberg*, *Nick Netzer*, and *Marco Sahm*.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

with someone who co-authored at least once with a co-author of *Christoph Engel*.

D. Discussion / Outlook

Given the limited space allotted to this contribution, I cannot conduct further explorations which might have been interesting. For instance, I did not attempt to manually complete (and utilize) the missing keyword or abstract data (see supra II.2), I did not use any of the full text files for which URLs had been collected (see supra II.3), I did not match preprints with published research to determine how many preprints resulted in publication, or how much they changed in between. These are just some of the areas that might be explored using the new dataset. Given that retired academics famously have too much time on their hands, I hereby present this dataset as a gift and (hopefully inspiring) pastime to the retiree.[47] Happy birthday belatedly, *Christoph Engel*!

47 Available from Zenodo: Preprints & Discussion Papers of the Max Planck Institute for Research on Collective Goods Bonn (1998–2024) (https://dx.doi.org/10.5281/zenodo.15479677).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The First

*Jeffrey J. Rachlinski**

Christoph Engel is a remarkable scholar. He is a European legal scholar who blends law and economics, psychology, behavioral economics, game theory, and empiricism. Indeed, he is a pioneer. But he is not just a pioneer; he is the first to combine all of these approaches into the study of law in Europe. He is, of course, too modest to make such a claim on his own. And there are certainly others who have brought novel methods to the more formalistic approaches on which European legal scholars customarily rely. But I will say it – he is the first. In his honor, I offer this article, which addresses the rise in assertions that a law review article is "the first" of its kind.

Claims to being the "the first" in legal scholarship have enjoyed explosive growth in law review articles in the United States. As the data I report show, for most of the history of legal scholarship, no author claimed to have produced "the first" work on any topic. Today, however, well over 200 articles per year make this extravagant claim (Although this claim may be leveling off after a period of exponential growth.) What a remarkable scholarly era we must have entered, in which so many articles are "the first" on any subject! This article (the first on the subject) documents the trend for authors to announce the primacy of their work and discusses its causes.

Of course, every law review article is the first of its kind. At an average length of 25,000 words, any published law review article is, absent severe plagiarism, the first article to use exactly those 25,000 words in exactly that order. Any article that fails to break some new ground is either an empirical replication or probably not worth reading. Authors who use this phrase surely mean to signal more than that they have adopted a unique combination of words. An article that is "the first" presumably denotes a path-breaking scholarly endeavor; something never before achieved in the

* I wrote this piece after repeatedly encountering the phrase "the first article" in drafts articles over the last six months. Research funding in support of this Article was provided by the Roger C. Hyatt Fund.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

many millions of pages of law reviews that have come before. Something like what Christoph Engel has achieved with his body of work.

Legal scholarship should be witnessing fewer firsts, not more. After over a century of legal scholarship, it should be hard to find a new thesis. Certainly scholarly focus drifts and legal scholarship might document novel trends in the development of law, technology, and society. Surely by now, however, most legal subjects worth addressing have received some attention. No one in 2026 can rightfully claim to have written "the first" article discussing the Second Amendment[1] or the first to propose strict liability for defective products.[2] Authors claiming to be "the first" are not actually marking new territory in this way. Rather, authors mean to signal their exploration of some unexplored avenue of research or to apply some existing methodological approach to a well-known topic. Even still, it is surprising that anyone can truly claim to be "the first" at this stage of legal scholarship. At the very least, these claims should be on the decline.

To document the rate at which authors claim primacy, I conducted a search in the Westlaw "Law Reviews and Journals" database.[3] Specifically, I searched for two variations on assertions to primacy: "this is the first article" or (the more common) "this article is the first." I omitted from these results those articles that used that phrase to denote the first of a series of articles by excluding those articles that used these terms within five words of the word "series."[4] This search captured a residue of articles that were not claiming primacy. For example, some authors denoted their first contribution,[5] or just happened to use the word Article too close to the

1 The first such piece I could find is S.T. Ansell, *Legal and Historical Aspects of Militia*, 26 Yale L.J. 471 (1917).

2 The well-known article by Professor James is a god candidate for that honor. Fleming James Jr., *Products Liability*, 34 Texas L. Rev. 192 (1955).

3 I updated the search on September 12, 2025.

4 For example, I excluded the 1912 piece, Jay Newton Baker, *The American Federation of Labor*, 22 Yale L.J. 73, 73 at note a1 (1912) ("This is the first article of a series of articles by the author on the subject of 'Regulation of Industrial Corporations.'"). The exact search was thus: adv: "this is the first article" or "this article is the first" BUT NOT ("this article is the first" /5 series).

5 For example, Muriel Morisey Spence, *The Sleeping Giant: Textualism as Power Struggle*, 67 S. Cal. L. Rev. 585, 585 n.* (1994) ("Because this is my first article as a law professor I want to extend special thanks to four of the people whose words and acts of support helped me decide to enter law-teaching: Regina Austin, Judge Harry T. Edwards, Lani Guinier and Robert Suggs."). I also excluded articles referring to the "first article" of a convention or treaty.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

phrase "the first", although not meaning their own article. I also excluded both articles that reported being the "first in thirty years" to address a subject.[6] I therefore had to review each article that the search captured to ensure it was a claim to primacy.[7] I cut the search off after 2023 because articles are still being added with a 2024 date.

Overall, after I excluded the articles that the search captured but did not actually claim to be the first, 2,369 articles claimed primacy. Figure 1 presents the results over time.

Surprisingly, the first use of the phrase did not occur until 1986 when two pieces claimed to be the first in their fields.[8] Over the next two decades a handful of authors included the phrase in their articles, but its use started to pick up in 2005 and took off at an exponential rate after 2007. After a plateau of about 150 such articles per year between 2015 and 2019, claims to primacy have surged further. In 2023, 220 papers declared themselves the first paper in some form.

These papers consist of several types of firsts:

– Some articles declared themselves the first "systematic" or "comprehensive discussion" of a topic. For example, Kathleen Clark laid claim

6 Victor Fleischer, *A Theory of Taxing Sovereign Wealth*, 84 N.Y.U. L. Rev. 440, 452 (2009) ("[T]his is the first article in thirty years to focus on sovereign investment and the first to focus on sovereign wealth funds."); Richard M. Hynes, *Bankruptcy and State Collections: The Case of the Missing Garnishments*, 91 Cornell L. Rev. 603, 607 (2006) ("Surprisingly, this is the first article in thirty years to carefully examine garnishment statistics." I also omitted article that quoted that language in Professor Hynes' piece. Steven L. Willborn, *Wage Garnishment: Efficiency, Fairness, and the Uniform Act*, 49 Seton Hall L. Rev. 847, 868 n. 109 (2019). Strangely, Professor Hynes did not identify the piece from thirty years earlier, which was likely the first of its kind.

7 The most difficult case concerned the use of "this article is the first step," which was fairly common. Many authors clearly meant this to open up a new area of scholarship and hence counted as a claim to primacy. E.g., Suneal Bedi & Mike Schuster, *Measuring Fair Use's Market Effect*, Wis. L. Rev. 1467, 1491–91 (2022) ("This article is a first step toward addressing whether a factual presumption is feasible and what sort of presumption would be appropriate.") Others intended the phrase as the first of a series of articles and were thus excluded.

8 Paul Burstein & Kathleen Monaghan, *Equal Employment Opportunity and The Mobilization of Law*, 20 Law & Soc'y Rev. 355 (1986) ("This is the first article to describe quantitatively the extent and outcomes of the mobilization of EEO laws at the appellate court level."); Robert J. Kaczorowski, *Revolutionary Constitutionalism in the Era of Civil War and Reconstruction*, 61 N.Y.U.L. Rev. 863, 867 (1986) ("This is the first Article to explore this question and explain the connection made by the framers between the primacy of national citizenship and the primacy of the national government's authority to enforce civil rights.").

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

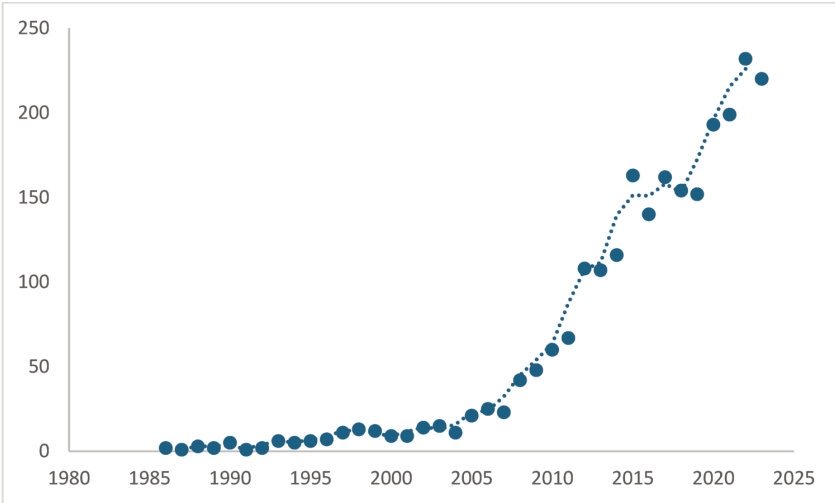


Figure 1: Number of Law Review Articles Claiming to be "The First Article" by Year

to publishing "the first article to take a comprehensive look at the [Emoluments] Clause as interpreted by the Department [of Justice]."[9] The article was not the first paper discussing the Department of Justice's view of the Emoluments Clause and that is not what Professor Clark claimed. The paper presented a review of the entire body of opinions of the Department of Justice on the subject and as such, was likely the first such assessment. I included papers in this category if they identified themselves as engaged in the first systematic or comprehensive analysis, even though some of these authors indicate that earlier discussions of the topic might exist.[10]

9 Kathleen Clark, *The Lawyer Who Mistook a President for Their Client*, 52 Ind. L. Rev. 271, 273 (2019).

10 Madison Condon, *Externalities and the Common Owner*, 95 Wash. L. Rev. 1, 7 (2020) (Declaring itself to be "the first Article to systematically examine institutional investor activism on environmental issues using the economic incentive of common ownership as a framework.") The use of the term systematic in this case might actually reflect caution by the author, much as others qualify their claim to primacy with phrase like "the first article of which we are aware." *See, e.g.*, John J. Donohue & Peter Siegelman, *Law and Macroeconomics: Employment Discrimination Litigation Over the Business Cycle*, 66 S. Cal. L. Rev. 709, 710 n. 2 (1993). I could also poten-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

- Some articles concerned highly specialized topics that previous literature presumably had neglected. For example, "Crisis-Driven Tax Law: The Case of Section 382" might well be important scholarship in an area of tax law, but it is a relatively narrow topic.[11]
- Some articles applied an existing framework of analysis to a topic that has not previously been subjected to that methodology. For example, two articles applied economic analysis to legal topics that had (somehow) previously escaped the attention of law-and-economics scholars.[12]
- Some articles made a connection between two existing lines of scholars. For example, Holning Lau and Hillary Li's "American Equal Protection and Global Governance" stated that it was "the first article to focus squarely on convergences between U.S. equal protection doctrine and global developments."[13]
- Some articles remarked on a novel trend, often one that calls for legal reform. For example, such as the first article to discuss "local approaches to regulating digital platform short-term rentals across major American municipalities," thereby addressing the rise of Airbnb and Vrbo.[14]

---

tially have categorized this paper as one of three other categories I describe below: addressing a highly specialized topic, given its particular analysis of investor activism; as the application of an existing framework to a new topic; or as making a connection between two existing lines of scholarship.

11 Albert Choi, Quinn Curtis & Andrew T. Hayashi, *Crisis-Driven Tax Law: The Case of Section 382*, 23 Fla. Tax Rev. 1 (2019).

12 A. Mitchel Polinsky & Paul Riskind, *Deterrence and the Optimal Use of Prison, Parole & Probation*, 62 J. L & Econ. 347, 349 (2019) ("this is the first article to compare the use of prison, parole, and probation to promote deterrence when offenders discount disutility and the state discounts costs."); Donohue & Siegelman, *supra* note 10, at 710 n. 2. ("This is the first article of which we are aware that shows that potential beneficiaries of an important federal statutory right are quite sensitive to the state of the macroeconomy in making decisions about whether to initiate litigation."). These two openly assert that they are using economic analysis, but other pieces claiming primacy appear also be using economic approaches. For example, Sergio J. Campos & Cheng Li, *Discovery Disclosure and Deterrence*, 71 Vand. L. Rev. 1993, 1995 (2018) (announcing itself to be the first article that considers whether orders sealing discovery might encourage misconduct.).

13 Holning Lau & Hillary Li, *American Equal Protection and Global Governance*, 86 Fordham L. Rev. 1251, 1256 n. 34 (2017).

14 Braedon Sims, Jordan Carr Peterson, *No Vacancy or Open for Business? Making Accommodations for Digital Platform Short-Term Rentals in Major American Municipalities*, 43 U. Haw. L. Rev. 123, 138 (2020).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

– Numerous first articles purported to be the first empirical or statistical analysis of their topics.[15]

One author even called the assertion that an assertion of primacy "the standard law review claim", before asserting it was actually true of their article.[16]

Declarations of primacy grace every type of journal. Many of the top flagship journals have published such articles, including the *Stanford Law Review*, the *University of Michigan Law Review*, the *University of Chicago Law Review*, and the *Cornell Law Review*. One might have expected student-edited journals to use the phrase with greater frequency, as authors feel the need to inform the less-experienced student editors of their novelty,[17] but this is not the case. Numerous claims to primacy have been published in peer-reviewed journals. In fact, as noted above, one of the first two pieces to claim primacy in 1986 was published in the peer-reviewed *Law and Society Review*.

I did not systematically assess whether the use of the phrase facilitates acceptance of the article or whether it attracts subsequent citations.[18] Although I am skeptical, the data nevertheless contain some evidence that "firsts" garner more citations. Among the articles declaring primary, 87 % received at least one citation and 52 % received more than ten citations. This compares favorably to a 2007 study showing that only 57 % of law review articles receive any citations and only 21 % are cited more than ten

15 Surprisingly, the first of these does not appear until 2006. Hon. Donald E. Shelton, Young S. Kiim & Gregg Barak, *A Study of Juror Expectations and Demands Concerning Scientific Evidence: Does the "CSI" Effect Exist?* 9 Vand. J. Ent. & Tech. L. 331, 332 (2006).

16 Brian Soucek, *The Case of the Religious Gay Blood Donor*, 60 Wm. & Mary L. Rev. 1893, 1935 (2019) ("Maybe the standard law review claim – 'This article is the first to . . . '–is actually true here.").

17 My colleague Jed Stiglitz makes this argument concerning the choice of title. Jed Stiglitz, Untitled 2 (Jan. 10, 2022) (unpublished manuscript on file with author) ("Relative to peer editors, student editors seem likely . . . to be more susceptible to strategies involving enticement or the representation of article quality through titling choices.").

18 A systematic analysis of this would likely require a full accounting of the predictors of citation rates among all law reviews, which is beyond the ambition of this modest undertaking. Ian Ayres and Frederick Vars provided a thorough analysis of the predictors of citation over twenty-five years ago but did not assess the then-rare claims of primacy. Ian Ayres, Frederick E. Vars, *Determinants of Citations to Articles in Elite Law Reviews*, 29 J. Legal Stud. 427 (2000).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

times.[19] Considering that most of these articles were publishing the last few years, they might be more valuable than most articles. That said, the 2007 study is dated; as the number of law review articles published rises (especially the short "on-line" commentary pieces), citation rates overall might be on the rise.[20]

Several patterns in these articles suggest that the increasing tendency to declare an article the first of its kind is a branding trend, not a change legal scholarship. A great many articles over the decades fit into some of the patterns of the types of types of "firsts" without declaring their primacy. For example, noting previously unobserved connections between lines of cases or lines of scholarship is a time-honored formula for a law review article, but only in recent years do some authors tout this as a breakthrough. Likewise, articles that comment on a new trend could plausibly arise from the dramatic changes in commerce and media technologies in recent decades, but technological and social trends that affect law are themselves nothing new. The Twentieth Century witnessed numerous technological and social changes without authors labeling their pieces as the first to address a topic. The first paper addressing the regulation of radio could have made similar claims, but it did not because articles published in the 1940s did not make such claims.[21] We can certainly expect a plethora of "first" articles on various aspects of the recent pandemic, however.

Also consider that only two of the "novel applications" pieces involved the application of economic analysis to an area of law. Law and economics witnessed decades of articles that were the first to apply economic analysis to law without authors announcing their novelty. The first of many articles applying an economic analysis to the suit and settlement process, for example, made no such claim.[22] By contrast, several "firsts" consisted of empirical analysis of an area. The difference lies not with the different methodologies, but with when these methodologies became popular. Empirical legal scholarship has grown as a methodology along

19 Thomas A. Smith, *The Web of Law*, 44 San Diego L. Rev. 309, 336 (2007) (reporting that among about 385,000 articles across 726 law reviews, 43% are never cited and 79% are cited ten or fewer times).

20 An earlier study reported that 57% percent of articles are never cited, which suggests a trend towards more citations over time. Deborah L. Rhode, *Legal Scholarship*, 115 Harv. L. Rev. 1327, 1331 (2002) (reporting that among 70,560 articles published between 1981 and 1993, 57% were not cited).

21 Note, *Radio Program Controls: A Study in Inadequacy*, Yale L.J. 275 (1947).

22 William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & Econ. 61 (1971).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

with the tendency to announce one's article as first, whereas the wave of law-and-economics scholarship emerged decades earlier.

Similarly, it is ironic that several articles using a framework developed in a previous piece declare their primacy, even though paradigm-shifting framework pieces did not announce themselves as the first of their kind. Consider Kristin Carpenter's article, "Real Property and Peoplehood," which claimed to be is "the first Article to consider Radin's theory in the context of American Indian real property."[23] The piece applied Professor Margaret Radin's work on market inalienability in this novel context.[24] Professor Radin herself did not claim any such primacy in her own work. Nor is any major, highly cited "framework" piece (such as Ronald Coase's "The Problem of Social Costs"[25]) among the firsts. The data reflect a trend in branding, not a change in the scholarship.

Ronald Coase's article perhaps merits special attention in a discussion of academic primacy. Guido Calabresi arguably wrote a similar article before Coase published his Nobel-Prize winning work.[26] Calabresi submitted his piece as a student note years before Coase's publication, but the student editors of the *Yale Law Journal* rejected it for being too unconventional.[27] Ultimately, Calabresi's article saw the light of day a year after Coase's more famous contribution.[28] With a Nobel Prize on the line, Calabresi surely should have submitted the piece earlier to SSRN and claimed primacy. But I suppose that is hindsight (and of course SSRN did not exist in the 1950s).

Why do law review articles increasingly declare themselves the first? As the details of the Coase and Calabresi articles illustrates, being first matters. One can hardly blame budding young authors for claiming pri-

---

23 Kristen A. Carpenter, *Real Property and Peoplehood*, 26 Stan. Env't. L. Rev. 313, 319 at n 26 ((2008)

24 Margaret Jane Radin, *Property and Personhood*, 34 Stan. L. Rev. 957 (1982).

25 Ronald Coase, *The Problem of Social Cost*, 3 J.L & Econ. 1 (1960).

26 *See* Steven G, Medema, Juris *Prudence: Calabresi's Uneasy Relationship with the Coase Theorem*, 77 Law & Contemp. Prob. 65, 65 (2014) ("[A]n argument can be made … that Calabresi 'discovered' the Coase Theorem roughly two years before Coase first stated those ideas in print.").

27 *See* Emily Sherwin, *Guido Calabresi*, *in* New Palgrave Dictionary of Economics and the Law v. 1 199, 199 (Peter Newman ed. 1998) (noting that "Calabresi wrote the initial draft of the article as a student comment … [but it] was not published until 1961").

28 Guido Calabresi, *Some Thoughts on Risk Distribution and the Law of Torts*, 70 Yale L.J. 499 (1961).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

macy. Who, after all, remembers those who come second? Norgay and Hillary were the first to summit Everest, but who comprised the second team to achieve that feat? No one I can name. The sprinter Usain Bolt won the gold medal for first crossing the finish line in the one-hundred and two-hundred-meter races in three consecutive Olympics to earn the title of "world's fastest human," but no one even marks the name of the second fastest human. The gymnast McKayla Maroney famously turned herself into an instant meme with her "not impressed" scowl while receiving the silver medal in the vault at the 2012 Olympics.[29] Who was the second human after Neil Armstrong to set foot on the moon? OK, Buzz Aldrin is also famous, but who captained the second expedition to the moon? Being second is perhaps worse in academia. We all know Watson and Crick for their first publication describing the structure of DNA; being second would not even merit publication except as a replication. As the first (and second) coach to win a Super Bowl, Vince Lombardi, put it, "winning isn't everything, it's the only thing." The only real puzzle is that the trend to declare primacy is so recent.

The rise of Scholastica might be to blame. Certainly, many woes can be laid at its doorstep. With just a few mouse clicks and a modest payment, today's authors can submit their articles to hundreds of student-run law reviews. One used to have to make copies and submit them via snail-mail. (Younger authors: really, actual physical copies – and here I feel like I do when explain to my kids how rotary phones worked.) Standing out to the student editors was always important, but it is now more important than ever as they have to sort through thousands of submissions.

Law perhaps produces more trends in scholarship than other fields.[30] The students who select articles are beginners to the profession and thus have a different perspective than the senior scholars who make editorial decisions for other journals. This perspective likely leads the students to search more for a novel trend than for articles that might support a well-accepted thesis. This tendency has some virtues, in that it avoids ossifi-

29 *See* https://knowyourmeme.com/memes/mckayla-is-not-impressed

30 I did not undertake a systematic analysis of the use of the phrase "this is the first" in fields other than law. Authors in other fields use the phrase, but it appears to be less common. A search of the SSRN database revealed 48 articles in 2020 and 2021 that use that phrase, 39 of which (81 %) were published in law reviews. The other nine consist of publications in computer science, economics, business, anthropology, sociology, and political science. I do not, however, know what percentage of articles in the SSRN database are law review articles.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

cation and capture of the discipline by an elder statesperson, but it has a downside as well. As the data show, legal scholarship now also more commonly announces its novelty.

The tendency to declare primacy perhaps sheds some light on what academics value in legal scholarship. Summarizing and synthesizing an area of law is not the ticket to academic glory at today's law schools. No matter how useful a summary of an area of law might be to practitioners and judges, it might not impress academics as much as an article that is "the first" of its kind. Chief Justice Roberts has bemoaned this tendency:

> Pick up a copy of any law review that you see and the first article is likely to be, you know, the influence of Immanuel Kant on evidentiary approaches in 18th-century Bulgaria, or something, which I'm sure was of great interest to the academic that wrote it, but isn't of much help to the bar.[31]

A piece on the influence of Immanual Kant on evidentiary approaches in 18th-century Bulgaria would surely also be labeled as the first article on the subject. (As it happens, the first such article was penned by Orin Kerr after the Chief Justice made that statement, and although it asserts that it "fills that gap" it does not explicitly assert that it is "the first."[32]) Authors surely mean declarations of primacy to be self-laudatory claims of a paper's worth and academics value pieces that change the paradigm, not restate existing doctrine.

For aspiring authors, I advise against announcing yourself as "the first," despite it being in vogue. The pronouncement can come across as needlessly and unintentionally pretentious. Well-known authors like Richard Posner and Cass Sunstein do appear as authors of papers claiming to be the first, but only once each, despite their many path-breaking contributions.[33] None of Christoph Engel's pieces in the database have made this claim.

---

31 Chief Justice of the United States John G. Roberts, Jr., Interview at Fourth Circuit Court of Appeals Annual Conference, available at www.c-span.org/video/?300203-1/conversation-chief-justice-roberts at approx. 30:40.

32 Orin S. Kerr, *The Influence of Immanuel Kant on Evidentiary Approaches in 18th–Century Bulgaria*, 18 Green Bag 2d 251 (2015).

33 Oddly, both are in the *Stanford Law Review*. Richard Posner, *An Economic Approach to the Law of Evidence*, 51 Stan. L. Rev. (1999); Cass R. Sunstein, Daniel Kahneman, David Schkade, Ilana Ritov, *Predictably Incoherent Judgments*, 54 Stan. L. Rev. 1153 (2002). Sadly, I confess one of my own pieces appears. Susan D. Franck, Anne van

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Given how specific assertions of primacy typically are, they also convey no useful information about the article. All authors are trying to write something that adds to the literature. Many of those claiming primacy seem to understand that their claim does not truly signal a path-breaking enterprise. Many add the caveat that their piece is "the first of which we are aware." Even the assertion that one's piece is the "first comprehensive" analysis is actually an admission that much has come before on the same topic. Better that you write "the last" article on a subject. Publish the piece that blankets the field, answers all questions, and leaves nothing more to be said. That would be a real triumph.

With that said and my hat off to Christoph Engel – a true first – this is the "last" article about the "the first article."

Aaken, James Freda, Chris Guthrie, Jeffrey J. Rachlinski, *Inside the Arbitrator's Mind*, 66 Emory L.J. 1115 (2017).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

On Generosity
Our Wonderful Journey with Christoph Engel

Renate Buijze & Sigrid Hemels

A. Introduction

Christoph Engel is a generous man. Especially generous with his most valuable possessions: knowledge and time. We had the privilege of experiencing this when the three of us worked together on our paper *Insuring your donation*.[1] For this paper, we conducted a lab experiment in which we tested whether individuals were willing to pay a premium to ensure their donation was used for the intended charitable purpose and/or to ensure they would receive a (tax) benefit on the gift.

Philanthropic giving was the topic of our joint research project with Christoph and a topic we – a professor of tax law and a cultural economist – are passionate about. In literature, philanthropy and generosity are often linked with one another. Generosity is used to describe the behavior and/or characteristics of a donor.[2] Wiepking[3] suggests using 'generosity behavior' as a more inclusive alternative to 'philanthropic behavior' when studying philanthropy across different cultures. Wright uses the term 'generosity' as a characterization of the philanthropic ethos in the United States.[4]

---

1 Renate Buijze, Christoph Engel, and Sigrid Hemels, "Insuring Your Donation: An Experiment" *Journal of Empirical Legal Studies* Volume 14, Issue 4, 858–885, December 2017.

2 See for example Breeze, Beth, and Wendy Scaife. "Encouraging generosity: The practice and organization of fund-raising across nations." *The Palgrave handbook of global philanthropy*. London: Palgrave Macmillan UK, 2015. 570–596. They describe how fund-raising influences philanthropic behavior, also referred to as the act of 'encouraging generosity'.

3 Pamela Wiepking, "The Global Study of Philanthropic Behavior" *Voluntas* Volume 32, 194–203, April 2021.

4 Karen Wright, "Generosity vs. altruism: Philanthropy and charity in the United States and United Kingdom." *Voluntas: International Journal of Voluntary and Nonprofit Organizations* Volume 12, Issue 4, 399–416, December 2001.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

With this Festschrift contribution we celebrate Christoph's generosity, one of his many great traits like patience, perseverance, gracefulness and open-mindedness. These characteristics are invaluable when joining forces with scholars from other disciplines to engage in interdisciplinary research. Not fearing the unexpected is another one. Christoph had no shortage of well-established academics enthusiastic to engage in a joint research project with him, as immediately becomes clear when one looks at the lengthy list of publications which he has co-authored. But still, he decided to work with us, academics with no experience in lab experiments.

Our journey with Christoph spanned four years, five conference- and seminar presentations, a dissertation chapter, one journal article, hundreds of e-mails and many conversations which were always pleasant, efficient and thought provoking. We encountered royals, overcame hurdles, benefitted from lengthy train rides, were supported by kind travel companions and saw on the horizon the impact our research could have. "It's not the destination, it's the journey", the quote famously attributed to American philosopher Ralph Waldo Emerson (1803–1882), most certainly applied to our project. Although Christoph kept reminding us that the destination was important as well. He was, thus, instrumental in getting us there.

Therefore, before bringing you onto our journey with Christoph, the next section jumps ahead and describes the destination: our joint publication.

B. The destination

The paper *Insuring your donation* investigates the concept of donation insurance, particularly as facilitated by donation intermediaries. A donation intermediary attracts funds, and channels them to the recipient chosen by the donor, which is often located outside the donor's country of residence. In exchange for a limited fee the intermediary, amongst others, ensures the donation reaches its intended goal and that the donor does not miss out on a tax benefit. This service can be interpreted as providing an insurance to donors to ensure that a) the donation reaches its intended goal (reduce efficacy risk) and; b) the donor will receive a tax benefit (reduce tax risk).

The efficacy risk concerns the risk that the donation may not serve its stated purpose, such as money meant for disaster victims ending up with local warlords. This risk is particularly pronounced for cross-border

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

donations, where it is harder to verify the spending of the donation. By leveraging their own reputation, donation intermediaries act as a 'quality stamp' and help contain the efficacy of the donation.

The tax risk is related to the fact that in many countries donors only receive a tax benefit on their domestic charitable donations. The tax risk concerns the risk that donors may not receive an expected tax incentive, e.g. when giving across borders. The donation intermediary contains the tax risk by having charitable status in the donor's country and/or certifying that foreign beneficiaries meet regulations. The paper addresses 1) the willingness of donors to pay for containing the efficacy risk, or the tax risk, or both and 2) to which degree the availability, and the actual purchase of such insurance increases the probability of donations.

The research questions were answered through a lab experiment with a 2x3x2 factorial design, including a Baseline, Effectiveness (efficacy risk), Tax (tax risk), and Effectiveness + Tax (both risks) treatments. In 'Insurance treatments', participants could buy insurance against these risks. The charity chosen for the experiment was Stichting Unite For Basic Rights (UFBR) which provided basic needs to orphans. Participants received a 5€ endowment and could donate 2.50€ to the charity, with a potential 1.25€ reimbursement. The "frivolous activity" identified for the efficacy risk was a dinner for UFBR board members. The willingness to pay for insurance was elicited using the Becker, DeGroot and Marschak mechanism.[5]

The lab experiment showed that a substantial fraction of participants donated across all treatments. The largest effect was found for the risk that the donation fails to reach its intended purpose. It significantly deterred donations (e.g. 39.3% vs. 73.3% in the Baseline).

The risk that the donation turns out more costly (tax risk) did not significantly reduce donations (60,7% v.s. 73.3% in Baseline, not significant).

The insurance option showed to have a significant effect on donation rates in the Effectiveness treatment and in the Effectiveness + Tax treatment. The insurance option was not critical for willingness to donate if the risk was confined to the tax refund. Interestingly, the willingness to pay for insurance against the risk that a donation does not reach its in-

5 Becker, Gordon M., Morris H. Degroot and Jacob Marschak (1964). "Measuring Utility by a Single-response Sequential Method." *Behavioral Science* 9(3): 226–232.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tended goal, or that it is not subsidized, was substantial and did not differ between these risks. Post-experimental tests suggest that participants paid for absolute certainty, rather than just attenuating risk.

In sum, while individuals are willing to pay for insurance against both efficacy and tax risks equally, the efficacy risk is the one that significantly deters giving if not contained. The tax risk, although people are willing to insure against it, does not deter giving if left uncontained. This suggests that from a societal perspective, insurance against the risk of donation misuse is more impactful than insurance against tax cost uncertainty.

These results did not come overnight. The project, however, did emerge out of pure circumstance.

C. Embarking on our journey

Our journey started on 8 November 2013, when Erasmus University Rotterdam (EUR) celebrated its first centennial. This was celebrated big time in the Rotterdam De Doelen Concert Hall, *mit allem Drum und Dran*, including guest of honor princess Beatrix, the former queen of the Netherlands. This royal presence (and difficult traffic) meant that the professors had to wait quite a while in the cortege before they could enter the concert hall. This had the benefit of giving lots of chatting time to professors who otherwise might not have met each other. And this is how the professor of legal empirical studies and the professor of tax law met for the first time and started to talk. Quite soon they left the usual chit-chat and started to discuss their research interests. The traffic really did its best to delay the royal highness. This allowed for a fruitful conversation on one of the tax professor's research topics – tax incentives for philanthropy – and the empirical legal studies professor's research methodology. Both became enthusiastic about possibilities for a joint research project, which could also include the tax law professor's then PhD candidate who conducted qualitative empirical research on cross-border philanthropy and taxation.

Usually, such conversations end with vague promises, warm goodbyes and then other obligations preventing anything coming from it. But not with Christoph Engel. That same evening (22:47) he sent an email titled 'Exploiting the momentum' in which he immediately showed his generosity. "It has been a true pleasure talking to you, and learning about the possibility of common ground. We should of course consider everything carefully. An empirical project is a substantial investment. But since

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I am now a member of Rotterdam Faculty, I would in principle be open to joining forces". Christoph by that time knew the professor of tax law had no clue about empirical legal research. Still, he decided to start this venture which from the start included the three of us: the empirical legal scholar, the tax lawyer and the cultural economist.

During the whole process, Christoph amazed us about how a busy man like him could always find time for the project. He often praised the train ride from the Netherlands to Bonn as giving him opportunity to come back to us,[6] usually with the speed of lightning. On 27 February 2014, a few hours after the three of us had our first meeting in which he eluded us on his research on dictator games, he wrote "This has been a highly promising kick-off meeting (terrible word, as if good science could be kicked around . . .)."

After some exchange of emails Christoph wrote to us on 3 March: "Such a jump start for an experimental project is not at all common. I have been so thrilled that I have immediately set out at work. Attached please find my first cut. [. . .]] I have the impression we have already made quite some progress. I very much look forward to the sequel, at our next encounter, and maybe before via e-mail."

It was a project we were all excited about. For us it was a fantastic opportunity to learn about this strand of empirical legal research with one of the most eminent scholars in the field.

D. The world of lab experiments

Christoph introduced us into the world of lab experiments. Renate found out that EUR at the time had two labs for such experiments, the joint lab of the faculty of Psychology and Business Administration and the lab of

6 Email of 16 April 2014 "The train ride to Rotterdam has just proven the right amount of time (and secludedness) to digest all your good work, Renate. I have the impression that our project is flourishing!" Email of 17 April 2014: "Thank you very much for an equally productive and enjoyable encounter. I have used the train ride back home to do the homework. Attached please find a version of the draft with the changes we have been talking about during the day, a second version with more changes that we have not yet discussed (. . .)" Email of 24 February 2016: "But my train is some 90 minutes delayed. This has given me a last minute chance to draft an alternative introduction for the extended abstract. (. . .)" Email of 11 May 2016: "Im Moment bin ich irgendwo im Niemandsland nicht weit von Ulm. Die Verbindung ist etwas schütter. (. . .)."

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the Erasmus School of Economics. By that time Christoph had informed us on the differences between psychology and economic experiments and Renate suggested the econ lab as being most suitable for our experiment.

Christoph not only opened the doors to a new world for us, he also generously shared all he knew about it. Renate has one particular occasion in mind. It was when Christoph had been working on the model that measured the willingness to pay for insurance. In a meeting with Renate he wanted to cross-check whether she agreed on the model. Soon he discovered that the sophistication of the mathematical equations included in this model surpassed the notion of an average cultural economist. Without hesitation he started breaking down the model. Step-by-step he wrote the composition of the model on the whiteboard, how it worked and why it worked. Even though he had to go back every now and then to ensure the cultural economist was still on board, with limitless patience he continued his explanation. By the time he got his message across, the entire whiteboard was covered with his writing. Regardless of the time and effort this took, his final message to Renate was 'you see, you understand this, you can do this!'

E. Building bridges

Christoph did not only share his time and knowledge, he also generously shared his network. A mathematician was so kind to check the model.[7] Christoph involved his colleague Lars Freund of the Max Planck Institute in Bonn to write the program for the experiment. Christoph had colleagues reviewing the paper and got the help of the language editor at his institute when our students pointed out a possible language mistake in the program.

On 5 May 2014 he invited Renate to join the Legal Institutions Seminar at the Max Planck Institute in Bonn where they presented our project and asked for (and were given!) feedback on our research question, theory and design. Like Christoph himself, his team was generous in sharing their feedback. In July 2014 he used the key features of our experiment at the

7 Email of 11 March 2014: "Happily she approves. To be on the safe side, I will occasionally also approach an economist, but might wait with that until we have advanced a bit further."

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

summer school of his graduate school in a course on experimental design. This turned out to be very fruitful, not only for those students, but also for us, as the students made some good suggestions.

Back in Rotterdam, Renate and Sigrid were able to include master students in the project as they filled in a survey on what charity was most appealing to them and what activities of such charities they would not like to fund (top number one: dinner for the board members). In addition, some of our tax law students also helped us by proofreading the instructions for the experiment and they were given the opportunity to participate in the lab experiment. Usually, tax law students do not learn about this kind of research during their studies, so it was extremely nice that we could involve our own students in the project as well.

By March 2014 the three of us were already working on the theoretical part of the paper. At the same time the research manager of the Erasmus School of Law (ESL) suggested us to apply for funding from the Frontier Research Fund of the ESL as a pilot project to work interdisciplinary in a professional lab. By the end of July 2014 we were informed that our proposal was accepted for funding. The ESL Board of Research was in particular happy with the interdisciplinary component, the experimental design, and the new contacts made through the project. This support meant that we were actually able to proceed with the lab experiment.

F. Two steps forward and one step back

Lars Freund was very quick with his programming work and during the summer of 2014 we, especially Renate, were introduced to z-Tree. A lot of emails were going back and forth and everything went like clockwork. However, in September we encountered an unexpected set back: the charities selected by our master students refused to accept money which had to be spent on a dinner for their board members. This even though they, just as well as we, knew that they were having such dinners. The charities were afraid about their reputation if they would not be able to state that 90 % of each euro they were given, was spent on their purpose. After having been turned down by several charities, in October our students provided us with a solution. Some of them had founded a charity that supported orphans living in poverty. They were willing to accept both the donations for their purpose and the donations for the board dinner. In November we also received a positive answer from the fund that would accept the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

donations to the intermediary. After this hurdle was taken as well, we could finally proceed to the lab.

After a lot of testing, in March 2015 the experiment was about to finally start. But not before another hurdle was taken. To reimburse the participants in the experiment, cash was needed in small coupons. ESL was not accustomed to, nor equipped for, cash transactions without receipts. After a thorough description of our plan, explaining that cash and anonymous payout was required, the finance controller of the law school approved. He even allowed Renate to store the cash in the safe in his office. Participants would have to sign off on receiving the cash amount. This involved a large administration, as each participant received a different amount. Also finding a bank that could provide cash in small coupons turned out to be practical obstacles the average cultural economist or tax lawyer does not consider.

In the end all went well. Participants (93 in total as there was quite a large no show) behaved properly, the computernetwork proved to be reliable, and our data were saved properly. A minor mistake in the program could be adjusted on the spot, as the programmer was on site. Christoph immediately sent an encouraging email: "In an experiment, so many things can go wrong that one should not worry too much if something happens. And in this particular case, the mistake does not seem to be all too severe. As long as the program behaves as it says on the screen, it is not very likely that participants really got it upside down. At any rate: very good you noticed and changed it." Because of the no shows, more lab sessions were held in May 2015. To get a higher participation, we invited the student pool of the psychology lab and all first-year law students who Sigrid happened to be teaching at that time. This was a success as we managed to have 244 participants in the end.

In April 2016, we had to run another experiment, in order to meet the review of one of the journals we had submitted the paper to. Renate was in the final stages of her PhD research and had to focus on that, but Christoph managed with all his contacts to organize this. Maximilian Kerk, then PhD-candidate in the Law & Economics department of ESL, joined to help execute the post-experimental tests. Thanks to his experience in running experiments in the economics lab, things went extremely smooth in this final stage of the data gathering.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

G. Getting published: the bumps in the road for a multi-disciplinary travel company

Collaboration between disciplines can be demanding. The fields of philanthropy, empirical legal studies, tax law and cultural economics all take different perspectives. Each discipline also has its own writing style, research methods and publication culture.

Christoph helped us in a very elegant way to master the art of writing for a, for us, new audience. In his email of 16 April 2014 he wrote "Experimentalists have a taste for what they tend to call a "crisp" paper. In that spirit, I have somewhat shortened your contributions. I of course meant to keep the information."

In June 2015, we started discussing the journal to which we would submit the paper.

Where submission and publication is a relatively speedy process in tax law, we learned that this takes much more time in the journals we were aiming for with this paper. Christoph suggested (as he phrased it) to first try to reach out to the sky. He warned us that odds were very much against us. However, he stressed that even if we were to be turned down, we would get valuable reviews that would help us improve the paper. Indeed, we were turned down at first, but were given useful reviews which we could use to improve the paper. This led to running an additional experiment.

The process of submitting and being turned down was at times a bit discouraging. But luckily, we had Christoph guiding us in this journey through highly valued journals. He remained positive and cheerful and spurred us to go on and not to give up or go for the lower ranks.

Nevertheless, by September 2016, Renate started exploring possibilities to publish in outlets for philanthropy research, in case the initial strategy would not succeed. She, amongst others, consulted one of the editors of the Nonprofit and Voluntary Sector Quarterly to see if the paper was a good fit. By November 2016 also Sigrid became a bit impatient, not being used to such lengthy processes of submissions and being rejected for publication and the need to make revisions again. At the Tax Research Network conference where she had presented a tax paper with the research results, she was invited to submit that paper for inclusion in the conference book. Sigrid suggested to take this opportunity to at least being able to publish the research. But Christoph made it clear that it was not the time to lose hope and faith in a publication in a high-ranking journal: "I know that this is taxing on our patience, and that these long struggles for

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a good outlet are uncommon in "black letter law" contexts. But I hope for your understanding: the peer reviewed world is just very different (but also gives very high rewards)."

H. Land on the horizon!

How right he was! Our paper had been accepted for presentation at the 2016 Conference on Empirical Legal Studies in Europe (CELSE). In addition, it was also accepted for the US Conference on Empirical Legal Studies (CELS) in the same year. Christoph graciously gave Renate the opportunity to present the paper at the conferences, but she had to focus on finishing her PhD, so Christoph presented the paper himself.

Within three weeks after the email exchange of November 2016, Christoph sent an email that at the CELS conference he met the editor of the Journal of Empirical Legal Studies, which, as he stated "is at the absolute top of the peer reviewed legal journals". We submitted the paper and in March 2017 we received encouraging reviews. We resubmitted after the revisions and in June 2017 we got the message that the paper was accepted for publication. Again, Christoph was very generous in his email in which he shared this great news: "I hope that you, Renate, can capitalize on this truly remarkable success for your future career. Your input has been essential!"

On 7 November 2017, just one day short 4 years after we started our journey and a little bit over a month before Renate defended her PhD thesis, we received the article in print. A fantastic end of an incredible journey or, as Christoph wrote "It has been an excellent experience, with a wonderful outcome." In a later email he added: "We should tell the next rector that it is critical for the academic success of Erasmus University to make scholars wait for the Queen (on Dies Natalis)."

I. Looking through the window

Indeed, for Renate's career this experience with experimental economics turned out to be of great value – although in a different manner than Christoph might have anticipated. She is now responsible for grant making at the Erasmus Trustfonds, the university foundation of EUR. The experience with setting up an experiment turned out to be truly helpful

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

when reviewing proposals for research projects by others, be it a Randomized Control Trial in medicine, a field experiment in public administration or a behavioural experiment in the economics department. Furthermore, the outcome of the research shows to be valid in her daily practice. When the Erasmus Trustfonds receives gifts from foreign donors, be it the UK, US or Switzerland, they are all willing to pay intermediaries to ensure they receive a tax benefit on their donation.

Not only did the project turn out useful for Renate's career. Already during our journey the paper posted on SSRN attracted the attention of a German insurer. A Corporate Responsibility Consultant of this company approached Christoph in November 2015. She wrote to him that she and some colleagues were working on a possible social impact insurance project. They were thinking about insurances to make sure that development aid would reach its intended goal. Christoph was invited to talk with them about our project. He wrote to us "It is fairly rare that a lab experiment attracts direct interest from industry or regulators. Ours seems to be the exception that proves the rule." It was nice and encouraging to see that we could already have an impact at that stage.

Christoph had a meeting with five employees of the company in December 2015 and immediately reported back to us. He had the impression that they were seriously interested in our work. He had had a lively discussion resulting in three potential lines of research for which the company was even prepared to give funding. Again, his creative mind had immediately set to work and he sent us a document with ideas for possible future research. He was very enthusiastic "it seems there would be room for research that is as relevant as it is scientifically interesting", but he kindly put himself in our shoes: "I can imagine that you, Sigrid, are hesitant because none of this involves tax. And you, Renate, might be hesitant because you do not want to be too heavily associated with lab or field experiments. But since we had this wonderful first joint endeavour, either of you would be most welcome to join in on a new project, to express preferences for the choice among them, and to shape a project you might want to join. Give it a moment of thought. I will certainly not be disappointed if none of this fits your plans. I would then see which of the Bonn researchers is interested. But you would, of course, have priority." Again, he showed his generosity in giving us a second possibility to work with him (where it would have been much easier of course to immediately turn to the Bonn colleagues), but also in offering the option to graciously decline. We both had to decline, because of other priorities. As always Christoph reacted

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

very kind on this and it, in our experience, had no effect whatsoever on our working relationship.

J. Conclusion

This was a truly interdisciplinary project as we all brought our knowledge to the table and were able to achieve synergy as the knowledge of each of us was necessary to come to the end result. Still, we are of the opinion that without Christoph Engel the end result would have been impossible. He was clearly the lead in the whole project, but never claimed that role. He insisted on having an alphabetical order of authors in the paper, he was always flexible in his extremely busy schedule to allow for meetings,[8] guided us through the lengthy procedure of publishing empirical studies and he generously shared his knowledge and network.

Rereading the emails we exchanged over the four years we worked on this project, makes us even more aware of how lucky we were to have had this opportunity of getting an insight of doing empirical legal research with one of the most eminent scholars in this field.

We will forever cherish the memories of our wonderful journey with this exceptional, kind and generous academic. Christoph, thank you so much for having us as one of your traveling companions!

8 Email of 21 January 2014 "My RILE colleagues have once more been very kind. They would have enjoyed my participating in the regular seminar with the students from the European Doctorate in Law and Economics, which also takes place on Thursday afternoon. But they understand that scheduling is tricky, and have told me I should seize the opportunity and meet with you. So let us go for it."

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II.
Experiments

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Using Experiments to Test Criminological Theory

Daniel S. Nagin

A. Introduction

I may be the only contributor to this Festschrift who is not a lawyer or economist with deep roots in law and economics or legal theory. I am instead a criminologist and statistician. My professional relationship with Christoph began about fifteen years ago when he invited me to Bonn to give a seminar on the state of the criminological literature on the deterrent effect of criminal sanctions, a topic of longstanding interest to me dating back to my dissertation. I was puzzled as to why the invitation was extended but I accepted because I had never been to Bonn. I thank my lucky stars that I did. Since then, our professional relationship has evolved into a valued friendship. It has also blossomed into a highly productive scholarly collaboration late in both of our careers.

The substance of that collaboration will be the focus of my contribution but first I want to share my impressions of Christoph. Other contributors are better placed than me to comment on the importances of his contributions to law and economics and legal theory. In this regard, I will observe that with over 150 publications in refereed journals and his authoring fifty books Christoph's contributions are voluminous. More importantly, however, they are cutting edge as evidenced by his countless keynote lecture invitations and the long list of his recognitions including honorary degrees from two of the world's leading universities – Hebrew University and the University of Copenhagen.

Based personal experience, I am better positioned to comment on three characteristics of Christoph that make him a valued collaborator and friend. They are his farsightedness, the clarity and rigor of his thinking and writing, and his genuine regard for others.

Concerning his farsightedness, in retrospect I now understand that the purpose of his Bonn invitation was to recruit me to his mission to bring the use of experimental methods, a staple of the law and economics research, to criminology. Criminology is a discipline that is deeply rooted in sociology and more recently developmental psychology. Experimental

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

methods are sparingly used in both these disciplines. With delay, he was successful in convincing me to join his mission and importantly there is evidence that our mission is bearing fruit! My point here is that remarkable careers such as Christoph's are rarely based on a single magisterial contribution but rather on an accumulation of linked scholarship based on a farsighted vision. Christoph clearly has the gift of vision.

Beyond a long-term vision, a successful scholarly career also requires clarity and rigor in thinking and in writing. I tell my graduate students that you don't understand it until you can write it with crystal clarity. I am consistently impressed by the clarity of his writing and the crispness of his logic; a skill I am sure he brings to all his scholarship.

I have come to a point in my career where I only collaborate with individuals whose company I enjoy and who treats others with respect and dignity. Christoph is a joy to interact with, and he treats everyone the same regardless of their station – with respect for their humanity.

B. Bringing experimental methods to criminology

Experimental methods in criminology have mostly been used to evaluate the crime preventions effectiveness of police deployment strategies such as concentrating police presence at crime hot spots or to test the efficacy of interventions to reduce recidivism of individuals under the supervision of the criminal justice system. Use of experimental methods to test criminological theories is rare.

Our first attempt to demonstrate to criminologists the usefulness of experimental methods to test criminological theory was a failure. Engel and Nagin (2015) was ultimately published in a behavioral economics journal. Recently, however, we have had a major success with the publication of Engel, Lanfear, and Nagin (2025) in the *Journal of Quantitative Criminology*, a prominent criminological outlet. The paper titled "When Do Guardians Deter Offending? An Experimental Test of Informal Social Control Mechanisms" is an outgrowth of my recent line of research on guardianship. Guardianship refers to actions taken by private citizens to disrupt or to prevent crime. It is central to collective efficacy theory (Sampson, et al., 1997), a highly influential theory in criminology and sociology which itself is a form of informal social control theory. As described by Sampson (2006a, p. 153, emphasis added), "The term collective efficacy is meant to signify an emphasis on shared beliefs in a neighbor-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

hood's capability for action to achieve an intended effect, *coupled with an active sense of engagement on the part of residents*."

My guardianship research has largely been supported by the Max Planck Society in my capacity as a Max Planck Legal Fellow. It has been heavily influenced by law and economics theoretical perspectives Christoph has shared with me.

Nagin, Barnum, and Herman (2024) distinguish two distinct forms of guardianship, reactive and proactive. *Reactive guardianship* involves an individual or a small group responding to an ongoing criminal incident or a potentially harmful situation with the aim of disrupting it. People balance prosocial motivation to help others against safety and social costs associated with intervening. The large psychological literature on bystander intervention and prosociality, along with research on personal safety, physical prowess, and social and moral attitudes, illuminate the complex decision-making processes underlying reactive guardianship. *Proactive guardianship* involves community members coming together to improve public safety, such as organizing block watches. It requires ongoing participation rather than a one-shot reactive decision to intervene. A key challenge to devising a theory of proactive guardianship which Christoph impressed upon me is to overcome the free rider problem of individuals who benefit from improved public safety without contributing to it. Collective efficacy theory and the economics literature on public goods provision provide frameworks for identifying factors that influence individual choices to contribute that might overcome the free rider problem. These include willingness to contribute, knowledge that others are contributing, and active social approval of contributors and disapproval of non-contributors.

Our recent successful foray in convincing criminologists of the value of laboratory-like experiments involved what I believe to be the first experimental demonstration of proactive guardianship. The design of the experiment, which Christoph was largely responsible for, adapts experiments from the law and economics literature on mechanism designs for discouraging free riding in the provision of public goods based on voluntary contributions.

The online laboratory experiment was designed as follows: Participants were randomly assigned to and informed of one of three roles described to them as A, B, and C but which I heretofore will referred to as the Thief, Target, and Proactive Guardian roles. In compensation for their participation, all participants received a 6£ endowment. Those as-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

signed to the Thief role were given the option of taking 3£ of the Target's endowment. Those assigned to the Proactive Guardian role were given the option of punishing the Thief for taking from the Target, by inflicting a fine. The amount contributed by the Proactive Guardian toward that fine was deducted from their 6£ endowment, thereby making it costly for them to intervene on behalf of the Target. The Target's endowment is not restored if the Thief choses to take regardless of the fining decision of the Guardian. Target participants, the would-be victims, are passive. We manipulated (a) the presence of Guardians and their number; (b) the cost of punishment for the Guardian and (c) the maximum punishment power of Guardians.

From the perspective of those in the Proactive Guardian role, the experiment was intended to mimic the dilemma confronting real world would-be agents of informal social control. Unlike police officers who have a duty to intervene if they encounter a problematic event that might escalate into a crime, would-be proactive guardians are not so obligated and must balance their sense of duty, if any, to contribute to the safety of their community against the potential cost to themselves of serving as proactive guardian (Barnum et al. 2024; Jin et al. 2024). In the real world, that cost is mainly personal safety, something that cannot be ethically manipulated in an experiment. Instead in the experiment the cost is monetary – a reduced payout for serving as a Guardian who punishes a Thief were the Thief to take. In recompense to the external validity loss resulting from this abstraction, the design makes it possible to investigate what is invisible in research on reactive guardianship: the conditions under which knowing of potential guardian intervention deters would-be perpetrators.

Key findings were:

1) Proactive guardianship works; we found that guardians are willing to intervene to punish infractions even as punishing is costly for them. Importantly, in conditions in which proactive guardians had sufficient punishment power to make taking unprofitable some would-be thieves were deterred. Specifically, in four of six punishment conditions proactive guardians role participants had sufficient punishment power to making taking unprofitable. In two other conditions the would-be thief could either take with impunity or take with no risk that the fine would exceed 3£, making taking at worst a breakeven proposition. Taking rates were significantly lower in the four higher punishment conditions than in these two conditions.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2) In addition to this deterrent effect of guardians, we observed an even more effective mechanism of prevention that is intrinsic to the individual: regard for others. Regard for others was measured by self-reports of the inappropriateness of taking and a standard incentivized test of social value orientation (SVO, Murphy et al. 2011) in which participants make six different allocations of a monetary reward between themselves and an anonymous other. Some participants with high revealed social value orientation and who self-reported that taking was inappropriate refrained from stealing, seemingly as a matter of principle. This regard for others had a far larger impact on taking decisions than the capacity of guardians to punish, which suggests individual predispositions toward offending may generally be more consequential than situational incentives such as guardianship.
3) We also found a pronounced interaction between deterrence and regard for others. Punishment exerted a deterrent effect only for those with low regard for others. On the one hand, participants high in social value orientation refrained from stealing, even in the absence of the threat of punishment. On the other hand, participants low in social value orientation were willing to steal, but less often when the threat of punishment was more severe.
4) The finding that deterrence and social value orientation strongly interact with deterrence is consistent with research that finds deterrence is relevant only for a subset of individuals for whom offending is perceived as a potential choice. Zimring & Hawkins (1973) and Pogarsky (2002) argue that some individuals will not commit a crime for ethical or moral reasons even in the absence of a sanction threat. Pogarsky calls such individuals "acute conformists".

Moral considerations, also, play a central role in Wikstrom's (2006) Situational Action Theory – morals act as a perceptual filter that governs whether crime is considered a viable option. When the moral filter rules out offending, the act will be excluded entirely from cognitive consideration. This in turn makes sanction threats irrelevant because the behavior is precluded regardless of the magnitude of the sanction threat. To our knowledge we are the first to demonstrate this prediction in an incentivized experiment that manipulates real costs of both intervention and punishment, in so doing further building our case for the use of lab experiments to test criminological theories.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

One important limitation of our study is that it is unlikely that many participants in the Thief role had histories of serious criminal involvement. At the time of this writing we are embarking on a study to address this limitation by recruiting individuals incarcerated in a US local jail to serve in the would-be thief role. Participants in the Target and Proactive Guardian Roles will continue to be drawn from the public. Participants from the public will also be recruited for the would-be thief role to serve as a comparison group for jail participants in that role.

Another extension we are embarking on is administering an adaptation of the experiment to participants from across the globe to study whether we replicate the findings of our largely US and UK sample to samples of participants from selected countries with varying average levels of SVO and of an index of the rule of law.

A final important gap in this line of research is examining the linkage of individual-level willingness to serve as a proactive guardian with the level of collective efficacy in the community in which they reside – a key prediction of collective efficacy theory is willingness to serve as a proactive guardian will be positively associated with community-level collective efficacy. Collective efficacy theory also predicts that would-be thieves in higher collective efficacy communities will more likely be deterred due to their knowledge that intervention is more likely. Testing these two predictions requires measuring collective efficacy in a sample of communities and then recruiting individual members of the community to serve in the Proactive Guardian and Thief roles.

References

Engel, Christoph and Daniel S. Nagin (2015), "Who is Afraid of the Stick? Experimentally Testing the Deterrent Effect of Sanction Certainty", *Review of Behavioral Economics*: Vol. 2: No. 4, pp 405–434.

Engel, Christoph., Charles C. Lanfear, and Daniel S. Nagin (2025). "When Do Guardians Deter Offending? An Experimental Test of Informal Social Control Mechanisms." *Journal of Quantitative Crimino*logy. https://doi.org/10.1007/s10940-025-09616-8

Jin K, Wu J, Zhang R et al (2024) Observing heroic behavior and its influencing factors in immersive virtual environments. Proc Natl Acad Sci 121:e2314590121. https://doi.org/10.1073/pnas.2314590121

Murphy RO, Ackermann KA, Handgraaf MJJ (2011) Measuring social value orientation. Judgment & Decision Making 6:771–781. https://doi.org/10.1017/S1930297500004204

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Pogarsky G (2002) Identifying "deterrable" offenders: Implications for research on deterrence. Justice Q 19:431–452. https://doi.org/10.1080/07418820200095301

Wikstrom, P-OH (2006). Individuals, settings, act of crime: Situational settings and the explanation of crime.In Wikström P-OH, Sampson RJ (eds.) The explanation of crime: context, mechanisms and development. Cambridge University Press. Pp. 61–107.

Zimring FE, Hawkins GJ (1974) Deterrence: the legal threat in crime control, 2. print. Univ. of Chicago Pr, Chicago.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Assessing How We Decide
Psychometric Development of the Decision Modes Scale*

Crystal Reeck, Hale A. Forster & Elke U. Weber

We present the development of the Decision Mode Scale (DMS), a 24-item scale that measures the use of six decision modes, the qualitatively different processes by which individuals report making decisions: calculation, affect, habit, social norms, identity, and morality. The calculation mode involves a deliberate process of assessing costs and benefits. The affective mode involves an emotion or gut-feeling driven process. The social norms mode involves consideration of what others are doing or what others condone as appropriate. The habitual mode involves following a previously determined rule or habit. The identity mode involves identifying what action is consistent with one's sense of self or own values. The moral mode involves consideration of action provides the greatest good. The final four modes (habit, social norms, identity, and morality) are all rule-based ways of making a decision that follow a previously-learned and/or socially-derived rule or convention. The DMS exhibits good reliability and convergent and discriminant validity. By providing the DMS, we integrate past research on decision modes and create an opportunity for additional research into the effects of different decision processes on their outcomes.

Personal Note by Elke Weber

This chapter is a tribute to Christoph Engel and his important work of exploring the impact of research in psychology and other social sciences, including the decisions sciences, on legal theory and legal decision making. His impacts are seen in publications that summarize important multi- and trans-disciplinary workshops organized by him (e.g., Gigerenzer & Engel, 2006; Engel & Singer, 2008) as well as in specific research contri-

* * The authors acknowledge support from National Science Foundation grants SES 1227462 and SES 1447274. The second author was funded by National Science Foundation GRFP DGE-16–44869. The authors declare no conflicts of interest.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

butions (e.g., Goldstein et al., 2006; Lubell et al., 2008). My decision to contribute this chapter on the development of a Decision Modes Scale to this volume connects current work in my lab to the enjoyable time of working with Christoph on the impact of institutions on the decision of how to decide, i.e., on the decision mode(s) used by individuals (Engel & Weber, 2007), during a sabbatical leave that he spent at Columbia University.

A. Introduction

People make decisions in qualitatively different ways, which have been categorized into three decision modes: calculation-based, affect-based, and role/rule-based decision processes (Weber et al., 2005; Weber & Lindemann, 2007). Calculation-based decision making involves weighing costs and benefits. Affective decision making involves following a gut feeling. Role/rule-based decision-making takes advantage of learned habits, if-then rules, and rules of conduct imposed by professional, ethical, and social roles to guide behavior. Decision mode usage is not mutually exclusive: decision makers rely on multiple decision modes to varying extents when making any decision (Engel & Weber, 2007; Krosch et al., 2012; Reeck et al., 2022). In this paper, we refine the classification of decision modes and develop a domain-general scale to measure the extent of their use.

B. Defining Decision Modes

Much past research has documented the existence of two decision modes, sometimes labelled rational and intuitive (Evans, 2008; Evans & Stanovich, 2013; Kahneman & Frederick, 2002). The intuitive mode is automatic and involves habitual and emotional responses, while the rational mode is conscious, deliberative, effortful, and can override these intuitive processes. This two-way categorization has been subjected to multiple critiques (see Evans & Stanovich, 2013). As a way of understanding decision processes, the two categories may be too broad to usefully differentiate between different decision strategies or omit some decision modes.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Some research has described a third class of decision modes as rule-based decision modes that activate a specific rule in response to recognizing the decision as the type to which the rule applies (for a review, see Weber & Lindemann, 2007). Rule-based modes include the previously documented identity (or role) based decision mode, case-based (or expertise-driven) decision making, as well as decision processes that include habits, social norms, and moral rules of conduct. Identity- or role-based decisions activate a social identity (e.g., as a parent) for which the decision maker has a rule and/or expectations about appropriate behavior (March, 1994). Experts use a typically unconscious recognition-driven or case-based process to make decisions based on similarities with past situations (Klein, 1998; H. A. Simon, 1990).

C. Trait Decision Styles versus State Decision Modes

Importantly, while individuals may exhibit trait-like propensities to rely more heavily on one decision mode than another, mode use is also a state-dependent construct, contingent on decision features (Weber & Lindemann, 2007). Several trait-based measures of chronic decision styles (for a review see Hamilton et al., 2016; Scott & Bruce, 1995) measure the way individuals tend to approach decisions. These existing measures of decision styles assess static, dispositional styles at the level of the individual, and do not allow researchers to quantify variance in decision making styles across contexts or types of decisions. Common sense tells us that although people may have propensities for certain decision styles, there is a great deal of variance in decision mode use across different decisions.

D. Past Research on Decision Modes

Past research has shown that understanding the degree to which decision modes are employed by a decision-maker can help us better explain and predict choices, particularly when decision makers face conflicting motivations or social roles (Krosch et al., 2012). Decision mode use influences pro-environmental decision outcomes: in particular, greater use of a role-based and/or affective decision mode is correlated with increased preference for pro-environmental choices whereas greater use of a calculation-based mode is correlated with decreased preference for pro-environmen-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tal choices (Reeck, Gamma, & Weber, 2022). Further, decision mode use can be experimentally manipulated, and interventions that increase the use of role-based decision modes increase subsequent pro-environmental decisions, including weatherization and green energy rate selection (Reeck et al., 2022).

E. The Current Research

This paper presents the results of three studies conducted to develop and validate our Decision Modes Scale (DMS).[1] We followed standard scale development procedures (Flake et al., 2017; Simms, 2008; Worthington & Whittaker, 2006). Table 1 presents an overview of the analyses conducted at each phase of the scale development process, and in which study each analysis was conducted. Study 1 was both substantive and structural, refining decision modes constructs, developing items, and then using exploratory factor analyses to evaluate scale items, finalize the factor structure, and develop a preliminary scale. Studies 2 and 3 addressed the structural phase, using confirmatory factor analysis to test the model fit for this scale. Study 3 assessed convergent and discriminant validity and test-retest reliability.

I. Overview of Participants and Decisions

We used a mix of online and student samples to validate the scale for respondents across a wide range of demographic characteristics (Table S27).

Decision mode use can be expected to differ across different decisions (Ames et al., 2005), so we created the scale using seven decisions that spanned different domains, including public policy, consumer purchase, relationship, and ethical choices (Table S1).

II. Study 1: Item generation, selection, and exploratory factor analysis

In Study 1 we expanded upon the existing three-factor (calculation, affect, role) environmental-specific decision modes scale by Reeck et al. (2022) in

1 Data and supplemental materials available at https://osf.io/uvzkx

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

two ways. First, we explored multiple types of rule-based decision strategies and second, we constructed the items in a more domain-general way across four different decisions, described below.

1. Methods

a) Item development

Building upon a pilot study that had tested 39 decision mode items (SI), we omitted 12 inadequate items from the pilot and added 21 items that represented a broader range of rule-based decision modes for a total of 50 decision strategy items (Table S5).

b) Procedures

Participants (N = 1,607) were recruited and paid via Amazon Mechanical Turk and randomly assigned to one of four decisions. After reading the decision description, participants reported their decision and wrote a few sentences explaining it. Participants then rated their likelihood to use each of 50 decision strategies on a 1 ("not at all likely") to 7 ("very likely") scale: "In making this decision in real life, how likely would you be to use each of the strategies listed below? There are no right or wrong answers; focus on what you naturally or typically do or would do." The four decisions were either environmental (whether to support a revenue neutral carbon tax), relationship (whether to break an important engagement with a romantic partner to help a friend move), moral (whether to report a stealing colleague to your supervisor), and consumer (whether to purchase an LED or a halogen lightbulb) choices. Participants then rated item clarity for a subset of nine randomly selected decision strategies by rating how confusing they thought other respondents would find the item, and finally answered demographics questions.

c) Analysis

We retained 1,161 respondents for analysis (see SI) and followed standard scale development procedures to develop the factor structure and reduce the number of items (Flake et al., 2017; Simms, 2008; Worthington & Whittaker, 2006), namely Horn's parallel analysis to implement a common factor analysis to determine the number of factors to extract,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

and exploratory factor analysis to determine the loadings of the items on each factor (Dinno, 2018; Worthington & Whittaker, 2006). Exploratory factor analyses used the "stats" package in R (R Core Team, 2019), using oblique rotation because our definition of decision modes allows use of multiple decision modes simultaneously, and thus different decision mode factors are likely to be correlated. Initial parallel analysis and factor analysis were conducted for each decision, and items that loaded on one or fewer decisions dropped. We finalized the factor structure by examining the similarity of a parallel analysis and exploratory factor analysis conducted separately on each decision and dropping factors that included two or fewer items.

After finalizing the factor structure, additional analyses reduced the number of items in the scale. To be conservative in item removal, we marked items that were marginal on the criteria below and any items where the retained item was less clearly superior to the dropped item as "reserve" to include as backup items in Study 2. First, we compared the factor loading matrix across the four decisions to look for (1) consistent loading above 0.40 and (2) cross-loading, defined as any item that loads on at least two factors where the difference in loading between two factors is less than 0.15 in magnitude. Second, we looked at clarity ratings. We dropped items that were rated as confusing by at least 10% of respondents on all decisions. Third, we looked for item pairs with unusually high (>.85) inter-item correlation and fourth, factor reliability below .70. Fifth, we looked at syntactic similarity, dropping any items that had very similar wording.

From the remaining list of qualified items, we developed a tentative scale, selecting four items for each factor that would provide as general a meaning as possible.

2. Results

The complete results of each of the phases of factor structure and item removal are presented in the SI. Parallel analysis and exploratory factor analysis identified six consistent, interpretable factors across the four decisions (Table S6, S9, S26).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

3. Discussion

In Study 1, we found consistent support for six decision modes across a range of different decision types. These decision modes were calculation, affect, social norms, habitual, identity, and moral. The last four factors each represent distinct elements of the construct of a rule-based decision mode. These factors are robust across decisions and conceptually meaningful descriptions of decision modes.

III. Study 2: Confirmatory factor analysis

The purpose of preregistered Study 2 was to conduct a confirmatory factor analysis to validate the factor structure identified in Study 1 (https://osf.io/uvzkx/). We hypothesized that the 24 items would load onto the six factors specified in Study 1.

1. Methods

This online study was conducted with 303 students at a large western public university. Students participated in exchange for course credit. The 303 respondents exceeds the 240 needed for a 10:1 item to participant ratio (Worthington & Whittaker, 2006). Participants saw a decision about whether to support a new student fee that would fund electric campus buses to replace the existing diesel buses. Participants made their choice and then rated their likelihood of using each of the 24 decision strategies identified in Study 1, plus the additional 10 reserve decision strategies, followed by demographic information.

Nineteen students gave the same rating for all decision mode items ("straightlined") and were excluded from analysis, leaving 284 respondents.

Confirmatory factor analyses used the lavaan package in R (Rosseel, 2012) to test whether the six-factor solution was a good fit to the data. We did not constrain the covariance of the factors in these models, consistent with our theoretical definition of decision modes as not mutually exclusive (Table S12).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2. Results

A confirmatory factor analysis revealed adequate model fit ($\alpha^2(237) = 498.52$, CFI = 0.905, RMSEA = 0.068, SRMR = 0.084; Table S26). Model-derived standardized covariances are shown in Table 2.

3. Discussion

The confirmatory factor analysis of Study 2 showed that the hypothesized six-factor, 24-item scale had sufficient model fit, with items loading strongly on each factor and factors exhibiting satisfactory reliability.

IV. Study 3: Validity and reliability assessment

The purpose of Study 3 was twofold. Structurally, we confirmed the model fit with additional decisions (reported in the SI). For external validity, we established the test-retest reliability and the convergent and discriminant validity of our decision modes instrument, to further demonstrate that our scale provides a robust, meaningful, and distinct contribution to the decision-making literature.

Past research on decision styles has focused on trait-based decision styles, rather than state-dependent, decision-contingent decision modes (Hamilton et al., 2016). To our knowledge, there are no other state-dependent measures of decision-making against which to compare our scale. We expect that decision mode use is somewhat influenced by these chronic, trait-based decision styles, in addition to state-based influences from the features of the decision and its context. Therefore we expect our scale to be moderately correlated with these chronic decision styles measures, identified from the Decision Making Individual Differences Inventory (Appelt et al., 2011) and a literature search for scales that measured decision making processes or cognitive styles directly related to decision making processes, identifyinf six decision styles measures (Table S13).

We assessed test-retest reliability by examining participants' ratings over time for the same decision. To examine the criterion validity of the DMS, we conducted exploratory analyses to examine the extent to which decision mode use predicts choice (Krosch et al., 2012; Reeck et al., 2022).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

1. Methods

This study, preregistered on OSF (https://osf.io/uvzkx/), involved two parts, completed two to four weeks apart, by U.S. residents through the online survey platform Prolific Academic. 252 U.S. residents completed part one of this study to provide a 10:1 ratio of respondents per item (Worthington & Whittaker, 2006), plus 10 pretest participants.

Participants completed the 24 item decision modes scale developed in Study 2 for two different decisions at Time 1. The first decision was a peacekeeper decision adopted from Krosch et al. (2012). The second decision was the carbon tax decision from Study 1 (Table S1).

The instructions were unchanged from Study 1. At Time 1, participants completed additional scales, including: the rational and intuitive subscales of the DSS (α = .86 for each subscale, 10 items, e.g., "I prefer to gather all the necessary information before committing to a decision",); the dependent (α = .77), avoidant (α = .90), and spontaneous (α = .86) subscales of the GDMS (15 items, e.g., "I often procrastinate when it comes to making important decisions," "I often make quick decisions"); the NFC (α = .91; 8 items, e.g., "I would prefer complex to simple problems"); Maximization (α = .51; 6 items, e.g., "No matter how satisfied I am with my job, it's only right for me to be on the lookout for better opportunities"); two single-item measures of rational and intuitive decision styles (2 items, "I rely on logic and reasoning when making decisions," "I trust my feelings when making decisions;"); and all five subscales of the MFQ, including harm (α = .69), fairness (α = 0.66), ingroup (α = .73), authority (α = .75), and purity (α = .85; 30 items, e.g., "Whether or not someone suffered emotionally," "When the government makes laws, the number one principle should be ensuring that everyone is treated fairly,"). At Time 1, participants provided demographic information.

Two weeks later, Time 1 participants were invited to complete the DMS a second time after making only one (the peacekeeper) decision, to assess test-retest reliability (Watson, 2004). 231 participants completed Part 2 of the study between two and four weeks after Time 1.

At Time-1 we retained 239 participants for analysis who passed the attention check, did not "straightline," and took more than 5 minutes 34 seconds to complete the study, as preregistered (Curran, 2016; Huang et al., 2012). At Time-2, we retained 214 of 231 participants based on the same preregistered criteria.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Analysis methods for the confirmatory factor analysis were unchanged from Study 2. As preregistered, we conducted analyses with both a 5- and 6-factor scale and found that the 6-factor scale showed sufficient model fit (SI).

To assess convergent and discriminant validity, we conducted Bonferroni-corrected pairwise correlations between subscales. We calculated decision modes subscale scores as the average of the four items in each subscale, on a 1 to 7 scale. We further explored alternate ways to score the decision modes scale to minimize acquiescence bias, as described in the SI.

2. Results

a) Convergent and discriminant validity

Table 3 displays the correlations between the decision mode scores and the other scales for the peacekeeper decision, with similar results for the carbon tax decision (Table S18). Observed correlations are largely consistent with our preregistered hypotheses (in black bold text). As predicted, calculation mode use was significantly positively correlated with the single item rational and the DSS rational subscales, with no significant correlations between the single item intuitive, DSS intuitive, or GDMS spontaneous subscales, respectively. We failed to find the expected significant positive correlation between the NFC and Maximization scales and calculation mode use, however. Affect mode use was significantly positively correlated with the single item intuitive scale and the DSS intuitive and GDMS spontaneous subscales, and no significant positive correlations between the single-item rational measure (these were significantly negatively correlated), NFC, Maximization, or the DSS rational subscale.

Among the other four modes, habitual mode use was directionally consistent but not significantly correlated with the MFQ authority subscale. Social norms mode use was significantly correlated with the GDMS dependent subscale and the MFQ authority subscale scores, but not with the MFQ ingroup subscale. Use of both the identity and moral modes were significantly positively correlated with the MFQ harm and fairness subscales. The use of identity mode was directionally but not statistically correlated with the MFQ purity subscale (Table 3).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

b) Between-decision variance and test-retest reliability

Participants' mode use on the peacekeeper decision correlates with participants' mode use on the carbon tax decision, but there is also considerable variation in mode use (Table S24).

We compared responses for the peacekeeper decision between Time-1 and Time-2. 93 % of respondents made the same decision at both times, and correlations between Time-1 and Time-2 mode use were equal to or greater than $r = .50$, as predicted.

c) Criterion validity

We explored whether mode use predicted choice, using a series of binomial regression models (Table S25) for the peacekeeper decision. Each mode predicts choice (Model 1). Mode ratings from the other decision, the tax decision, do not significantly predict choice for the peacekeeper decision (Model 2), and Time-1 mode use predicts Time-1 choice controlling for Time-2 mode use ratings (Model 3). Mode use also predicts choice controlling for the trait-based decision scale measures (Model 4). These results demonstrate that respondents' decisions meaningfully covary with the decision modes that they report using, suggesting that these decision mode ratings can provide a meaningful way to understand how decision processes influence choice.

3. Discussion

This study confirmed the acceptable model fit of our six-factor scale from Study 2 for two additional decisions. The six decision modes were related to, but distinct from, existing measures of trait-based decision styles, cognition, and moral foundations. While we failed to find that use of the calculation mode is correlated with cognitive styles NFC and Maximization, we do find that calculation and affect modes are correlated with decision styles measures of logical and intuitive decision styles, respectively. We found convergent validity between our social norms, identity, and moral modes and the MFQ: the foundations of harm and care are associated with both the identity and moral mode use, and the social norms mode use is associated with the authority foundation.

We confirmed our conceptual definition of decision modes as being informed by individual states and decision features, but also informed

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

by individual habits and traits: there was some stability but also considerable variance in participants' reported use of decision modes across different decisions and over time. For context, the amount of intertemporal stability we observe in decision modes is roughly equivalent to the amount of intertemporal stability observed in the Positive and Negative Affect Schedule Scale, another state measure of momentary or time-specific emotion (Watson, 1988). We found that decision mode use predicted choice for both Study 3 decisions.

F. Conclusion

The reported studies developed the Decision Modes Scale (DMS), a 24-item scale to measure six identified decision modes: calculation, affect, habitual, social norms, identity, and moral decision modes. The final four decision modes are all instances of rule-based decision processes where a decision is based on learned, socially derived heuristics and conventions.

This research makes two main theoretical contributions. First, we expand on the idea of context-dependent decision modes and provide a validated measure of decision mode use. Second, we expand the number of decision modes from three (calculation, affective, and rule-or role-based) to include additional subcategories of rule-based decision modes: social norms, habitual, identity, and moral.

Future research can address limitations of this study. While the current scale was designed to be domain-general, it may predict some types of decisions better than others. Future research can also refine the nature of the relationship between modes, e.g., between calculation and affective modes and the four rule-based modes. Finally, future research is needed to validate this scale with more diverse populations.

The fundamental limitation of this research is the use of self-report to measure decision modes: do people have accurate introspection into their decision processes to answer these questions? We know that self-reports of decision processes and their influences have limitations (Ericsson & Simon, 1980; Pronin & Kugler, 2007). Nevertheless, a wide range of decision research suggests that self-report data is valuable (Chan, 2009).We believe that an imperfect measurement of these constructs can advance research in this area and pave the way for different measurement approaches, e.g., neuroimaging methods that capture the relative influence of different considerations during decision making (e.g., Hutcherson et al., 2015).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

References

Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, *6*(3), 11.

Chan, D. (2009). So why ask me? Are self-report data really that bad. In *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–336). Routledge.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006.

Dinno, A. (2018). *Paran*. https://cran.r-project.org/web/packages/paran/paran.pdf.

Engel, C., & Weber, E. U. (2007). The impact of institutions on the decision of how to decide. Journal of Institutional Economics, 3, 323–349. doi: https://doi.org/10.1017/S1744137407000744.

Engel, C. & Singer, W. (2008). Better than Conscious: Implications for Performance and Institutional Analysis. Boston MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063.

Gigerenzer, G. & Engel, C. (2006). Heuristics and the Law. Oxford University Press.

Goldstein, D.G., Arkes, H.R., Beckenkam, M., Cooter, R., Ellickson, R. C., Engel, C., Guthrie, C., Hertwig, R., Kurzenhaeuser, S., & Weber, E. U. (2006). How do heuristics mediate the impact of the law on behavior? (pp. 439–465). In Gigerenzer, G. & Engel, C., Heuristics and the Law. Oxford University Press.

Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment*, *98*(5), 523–535. https://doi.org/10.1080/00223891.2015.1132426.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18(3–4), 117–144. https://doi.org/10.1080/03610739208253916.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8.

Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Areas and Integrated in Ventromedial Prefrontal Cortex. *Journal of Neuroscience*, *35*(36), 12593–12605. https://doi.org/10.1523/JNEUROSCI.3402-14.2015.

Kahneman, D., & Frederick, S. (2002). Heuristics of intuitive judgment: Extensions and applications. *Heuristics of Intuitive Judgment: Extensions and Applications*, 1–30. https://doi.org/10.1038/2251090a0.

Klein, G. A. (1998). *Sources of Power: How People Make Decisions*. MIT Press.

Krosch, A., Figner, B., & Weber, E. U. (2012). Choice processes and their post-decisional consequences in morally conflicting decisions. *Judgment and Decision Making*, *7*(3), 224–234.

Lubell, M., Engel, C., Glimcher, P., Hastie, R., Rachlinski, J., Rockenbach, B., Selten, R., Singer, T., & Weber, E. U. (2008). Institutional Design Capitalizing on the Intuitive Nature of Decision Making. In: Engel, C. & Singer, W. (Eds.), Better than Conscious: Implications for Performance and Institutional Analysis (pp. 413–432). Boston MA: MIT Press.

March, J. G. (1994). *Primer on Decision Making: How Decisions Happen*. Simon and Schuster.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, *43*(4), 565–578. https://doi.org/10.1016/j.jesp.2006.05.011.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org.

Reeck, C., Gamma, K., & Weber, E. U. (2022). How we decide shapes what we choose: Decision modes track consumer decisions that help decarbonize electricity generation. *Theory and Decision*, *92*(3), 731–758. https://doi.org/10.1007/s11238-022-09874-z.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818–831. https://doi.org/10.1177/0013164495055005017.

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, *2*(1), 414–433. https://doi.org/10.1111/j.1751-9004.2007.00044.x.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*(1), 1–20. https://doi.org/10.1146/annurev.ps.41.020190.000245.

Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, *38*(4), 319–350. https://doi.org/10.1016/j.jrp.2004.03.001.

Weber, E. U., & Lindemann, P. G. (2007). From intuition to analysis: Making decisions with our head, our heart, or by the book. *Intuition in Judgment and Decision Making*, 191–208.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Tables

Table 1: Overview of Scale Development Phases

Phase	Analysis	Study
Substantive	Item development	1
Substantive	Exploratory factor analysis	1
Structural	Item reduction	1
Structural	Confirmatory factor analysis	2 & 3
External	Convergent and discriminant validity	3
External	Test-retest reliability	3
External	Measurement invariance	1 & 2 & 3

Table 2: Standardized Covariances of Study 2 Factors

Factor	Affect	Calculation	Identity	Habitual	Social Norms
Calculation	–.03				
Identity	.61	.44			
Habitual	.69	.09	.51		
Social norms	.51	–.09	.11	.65	
Moral	.15	.77	.60	.37	.14

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 3: Decision Mode Correlation with Existing Trait-based Scales

Scale	Calculation		Affect		Habitual		Social Norms		Identity		Moral	
	H	*r*	H	*r*	H	*r*	H	*r*	H	*r*	H	*r*
Intuitive 1-Item	**0**	**–.02**	+	**.45*****		.28 **		.07		.31 ***		.29 ***
Rational 1-Item	+	**.35*****	**0**	**–.26***		.06		.01		–.01		.10
NFC	+	**.11**	**0**	**–.09**		–.12		–.25 *		.12		.19
Maximization	+	**–.07**	**0**	**.09**		.05		.10		–.03		–.07
DSS: Rational	+	**.44*****	**0**	**–.15**		.04		–.13		.14		.30 ***
DSS: Intuitive	**0**	**–.13**	+	**.53*****		.27 **		.20		.21		.12
GDMS: Spontaneous	**0**	**–.07**	+	**.32*****		.24 *		.18		.09		.01
GDMS: Dependent		.11		.00		.05	+	**.23**		–.06		.05
GDMS: Avoidant		–.17		.03		.01		.13		–.15		–.23
MFQ: Harm		–.04		.30 ***		.08		.00	+	**.26****	+	**.37*****
MFQ: Fairness		.07		.29 ***		.15		.00	+	**.36*****	+	**.42*****
MFQ: Ingroup		.03		.14		.21	+	**.12**		.07		.09
MFQ: Authority		.01		.04	+	**.21**	+	**.32** ***		–.06		–.04
MFQ: Purity		.06		.22		.20		.08	+	**.14**		.15

Note. Correlations are provided for the peacekeeper decision. Bold, black font indicates preregistered hypotheses (the direction of the hypotheses is indicated by the + or 0 in H labeled columns). Correlations use Bonferroni correction. * $p < .05$, ** $p < .01$, *** $p < .001$.

https://doi.org/10.5771/9783748847257 https://www.inlibra.com/de/agb - Open Access - (cc) BY

Appendix: Decision Modes Scale Administration and Analysis

In making this [optional description] decision in real life, how likely would you be to use each of the strategies listed below? **There are no right or wrong answers; focus on what you naturally or typically do or would do.**

1 Not at all likely / 2 / 3 / 4 / 5 / 6 / 7 Very likely

Item	**Wording** [randomize presentation order]	**Decision mode** [For scoring, not displayed]
A1	Follow your gut feelings	Affect
A2	Trust your immediate emotional reactions	Affect
A3	Act based on how you feel now	Affect
A4	Follow your heart	Affect
C1	Consider the overall costs and benefits	Calculation
C2	Weigh potential benefits against risks	Calculation
C3	Analyze all the options and pick the overall best	Calculation
C4	Work out the different pros and cons of each option	Calculation
I1	Act in a way that is consistent with your identity	Identity
I2	Be true to yourself	Identity
I3	Be true to your values	Identity
I4	Do what is consistent with who you are	Identity
M1	Do what is fair	Moral
M2	Do the "right" thing	Moral
M3	Consider what will do the most good	Moral
M4	Think of the greater good	Moral
H1	Do what you always do when in this situation	Habit
H2	Follow what you've done in the past	Habit
H3	Assess similarity to prior situations and make the same decision	Habit
H4	Decide based on what you've done in similar situations	Habit
S1	Do what others are doing	Social norms
S2	Do what other people think is the right thing to do	Social norms
S3	Follow social norms	Social norms
S4	Do what people expect you to do	Social norms

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Scoring. Compute six separate mode use scores by averaging respondent scores on the four mode items.

Analysis. As decision mode use can covary, we recommend analyzing mode use while controlling for use of other modes. See Supplementary Materials for preliminary instructions for alternate administration and scoring if conducting analyses with only one mode.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The Roles of Anger in Moral Courage

Anna Baumert & Julia Sasse

Word of Gratitude

We were fortunate to work at the Max Planck Institute for Research on Collective Goods in Bonn, with Prof. Christoph Engel as director, from 2017 to 2022. During this time, as part of the Moral Courage Research Group at the institute, we devoted much of our research to understanding the roles of anger in moral courage. Our work was shaped by the intellectual environment that Prof. Engel created. His commitment to methodological rigor inspired us to employ economic games as a central tool for investigating moral courage, and his insistence on stringent theoretical formulation guided us in striving for precision and clarity in our work. These influences have left a lasting imprint on our approach and continue to shape the way we develop and conduct our research.

A. Introduction

Standing up against others' violations of social norms and moral principles can avert harm from individuals and contribute to the well-functioning of society. Yet, such displays of *moral courage* are rare (Baumert et al., 2013, 2024; Goodwin et al., 2020; Kemper et al., 2022). Accordingly, there is great scientific and applied interest in understanding the psychological barriers to moral courage and under which conditions individuals overcome them. Empirical work across disciplines has highlighted key roles that anger plays in moral courage (Eriksson et al., 2017; Fehr & Gächter, 2002; Halmburger et al., 2015; Niesta Kayser et al., 2010; Sasse et al., 2022): First, anger experienced in reaction to an observed wrongdoing is a causal factor that "fuels" moral courage. Experienced anger is thought to shape information processing and decision making in ways that facilitate overcoming barriers to moral courage. Second, anger expression has been proposed as a low-cost form of intervention which might be employed strategically by witnesses of moral violations. Third, anger expression can accompany morally coura-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

geous action, and these expressions are thought to shape its interpersonal effects, particularly how the action is judged by others.

In the present chapter, we (selectively) review existing evidence, discuss potential mechanisms for how anger exerts its influence on moral courage, and highlight open research questions. In sum, the research on the roles of anger in moral courage fosters our understanding of moral courage, but also refines how we think about anger as a functional emotion.

B. What is moral courage?

Moral courage (also termed "civil courage", "Zivilcourage" or "courage civique") manifests when initially uninvolved observers act to stop or redress the violation of social or moral norms committed by others despite risking incurring costs from intervening (Baumert et al., 2020; Osswald et al., 2010; Skitka, 2012). Examples of situations that afford moral courage are the unfair treatment of individuals, racism, sexual harassment, bullying, or corruption (Dungan et al., 2019; Goodwin et al., 2020; Kemper et al., 2022; Fischer et al., 2006; Pouwels et al., 2019). Morally courageous actions can take various forms, including verbal or physical confrontation, sanctioning the perpetrator, moving a victim to a safe space, or calling on authorities. Intervening in such ways can be costly for the individual who might experience physical harm, social isolation, or financial loss due to the perpetrator or others who react negatively to the intervention. Yet at the level of democratic societies, acts of moral courage by initially uninvolved witnesses are considered key for maintaining societal norms (Fehr & Fischbacher, 2004; Yamagishi, 1986).

There are several related and partly overlapping concepts, such as high-risk helping (Fischer et al., 2006), social control (Brauer & Chekroun, 2005), and second-party punishment (Molho et al., 2020). Critically, in partial distinction from these concepts, moral courage pertains to situations that involve a norm violation perceived as unfair or immoral (rather than violations of descriptive norms) and reactions from initially uninvolved third-parties (rather than victims). As a core element, moral courage refers to acts of *defending* moral principles and social norms (Skitka, 2012), and is thereby distinct from helping, caring or other ways of *adhering* to moral principles (e.g., donating or volunteering; Greitemeyer et al., 2006).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The empirical investigation of moral courage often proves challenging. Assessing moral courage through self-reported intentions seems ill–suited since intentions do not necessarily translate into action (e.g., Kemper et al., 2022). Directly observing morally courageous behavior, instead, is difficult due to ethical constraints. Notably, moral courage is directly compatible with the concept of third-party punishment, and the so-called Third-Party Punishment Game (3PPG) provides a behavioral paradigm for studying (a variant of) the phenomenon. The 3PPG consists of a classic dictator game in which one person – under full anonymity – allocates resources between themselves and a recipient, with the addition of a third person who can intervene against the dictator's decision through costly punishment. The 3PPG setting incorporates some of the defining features of a situation affording moral courage: If a dictator allocates resources very unequally, this is typically perceived as a violation of a fairness norm (Krupka & Weber, 2013), and, per design, the third person is not themselves affected by the dictator's (unfair) allocation, but intervening is associated with costs. While the 3PPG lacks several features of "real-life" norm violations, such as the dynamic unfolding of social situations, ambiguity of the norm violation and uncertainty of possible costs (Toribio-Flórez et al., 2023), or a wide range of behavioral options (Pedersen et al., 2018), it offers a high degree of experimental control. This makes it an ideal starting point for investigating behavioral manifestations of moral courage and much of the work reviewed here on the roles of anger has relied on the 3PPG.

C. A causal role of anger in "fueling" moral courage

As an early indication of a potential key role of anger in the punishment of social norm violations, Fehr and Gächter (2002) demonstrated that participants in a public goods game tended to punish free-riders, even though punishment was costly and without any possible individual benefit to the punisher. Importantly, they found that participants reported more anger when free riding was more (vs. less) extreme, and anger was also generally expected as a reaction to free-riding. Based on these findings, Fehr and Gächter proposed that anger could be a proximate mechanism for the altruistic punishment of norm violators.

Addressing reactions from the perspective of an observer more specifically, several studies found a positive association between angry

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

reactions to an unfair allocation in a 3PPG and costly third-party punishment (e.g., Jordan et al., 2016; Molho et al., 2017). For example, Lotz and colleagues (2011) asked participants in the role of the third person to write a message to the (unfair) dictator, and the intensity of anger in these messages, as judged by independent raters, correlated with the level of costly punishment applied by the writer. Further studies in more complex "real-world" settings yielded similar findings. Using behavioral observation paradigms, for instance, we found unique positive associations between experienced anger and intervention against (staged) norm violations in the lab (Halmburger et al., 2015; Sasse et al., 2022). In a similar vein, dispositional anger proneness was found to distinguish between people who engaged in morally courageous action and those who did not (Baumert et al., 2025; Niesta Kayser et al., 2010).

Critically, these correlational studies cannot establish causality: Anger and punishment could be independent products of other psychological processes, such as perceptions of unfairness or harm. Experimental manipulation of anger is necessary to rigorously test for causality, but such manipulation poses exciting challenges. Emotion induction procedures can be grouped into two broad clusters: incidental and integral (Ferrer & Ellis, 2021). *Incidental emotion induction* rests on the assumption that emotional states "spill over" from one situation to a subsequent, unrelated situation. Some studies used incidental anger induction and found effects on third-party punishment (Gummerum et al., 2016; Rodrigues et al., 2020), yet only under certain (unpredicted) conditions. In our own research (Sasse et al., 2025), across five highly powered experimental studies, we were unable to replicate a causal effect of incidentally induced anger on punishment in the 3PPG.

In *integral emotion induction,* in contrast to incidental, a naturally occurring emotion is experimentally intensified or attenuated through modifications of situational features. It thus allows a close approximation of naturally occurring emotional processes in the situation of interest. At the same time, there is a risk of confounding other psychological processes with the target emotion. For instance, Nelissen and Zeelenberg (2009) manipulated anger in the 3PPG by varying the intentionality of an unequal allocation. Anger and punishment were higher when the act was intentional (vs. unintentional). Critically, an unequal allocation without intentional decisions might not be perceived as a norm violation, thereby confounding appraisals with the anger reaction. In three experimental studies, we aimed to overcome these challenges by manipulating anger

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

while holding the norm violation identical. For this purpose, we employed emotion regulation instructions (Sasse et al., 2025). Before presenting an unfair dictator decision to participants in the role of the third person in the 3PPG, we instructed them to regulate their emotions in specific ways: In one condition, they were told to focus on their emotional reactions and intensify them. In the other condition, they were told to focus on the facts and to retain a neutral mood. Results showed that in the former (compared to the latter) condition, anger was more intense, and third-party punishment was heightened. Notably, this induction procedure keeps the characteristics of the situation, most critically the unfairness, equal between conditions. Our analyses further confirmed that the effects of anger remained robust when perceived unfairness was statistically controlled for.

In sum, these experimental studies provide evidence for a causal role of anger in enhancing third-party punishment of unfair allocations. Interestingly, the pattern of results suggests that being angry does not lead to punishment in situations unrelated to the object that triggered the emotion. In other words, the motivational impact of anger does not seem to "spill over" to other situations. Rather, anger seems to motivate punishment in a goal-oriented fashion: rectifying the norm violation that has triggered the anger reaction.

D. How does anger facilitate moral courage?

With convincing evidence for a causal effect of anger on intervention against norm violation, the question follows how exactly anger exerts its impact. What are the psychological mechanisms by which anger "fuels" moral courage?

Existing process models of moral courage have not incorporated emotions explicitly. Several authors have built on the seminal model of helping in emergencies by Latané and Darley (1970) and proposed that in reaction to an observed norm violation, once it is detected and interpreted as such, processes of taking responsibility, choosing effective means, and appraisals of costs and risks should determine whether a witness intervenes or not (e.g., Ashburn-Nardo et al., 2014). Accordingly, *high* ambiguity, *low* perceived responsibility, *low* perceived efficacy, and *high* perceived costs and risks of intervening should be critical barriers to moral courage.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Integrating anger in process models of moral courage can be achieved based on the general assumption that emotion, information processing, and motivation are inherently interconnected. The characteristic appraisal pattern of anger involves the perception that standards of "what ought to be" are violated (Frijda, 1986; Ortony et al., 1988). A characteristic motivation consists of wanting to rectify such violation (Fischer & Roseman, 2007), even when the violation affects others rather than oneself (Hechler & Kessler, 2018). The theoretical and empirical background on anger suggests that anger functions to overcome barriers in the pursuit of rectifying a situation. For moral courage, this means that anger should facilitate intervention against norm violations, especially in situations that involve situational barriers.

At present, rigorous tests of the exact mechanisms of anger in "fueling" moral courage are still missing. We can only speculate that there are three plausible ways in which anger could exert its behavioral impact: First, anger could directly activate intervention behavior. Second, anger could modulate subjective appraisals such that situational barriers to intervention are perceived to be low. Third, anger could modulate the subjective relevance assigned to situational barriers in decision making. In other words, when deciding whether to intervene, angry (vs. less angry) individuals might be just as aware of barriers but weigh them less. Future experimental research could scrutinize how anger shapes or interacts with perceived barriers to intervention against norm violation. Such studies will work toward establishing a comprehensive understanding of the causal mechanisms of anger in moral courage.

E. Anger experience and anger expression in moral courage

To this point, we have discussed that *anger experience* can arise from the perception of a norm violation and provide the motivational "fuel" to act against it. Next, we want to turn to the roles of *anger expressed* in situations affording moral courage. While emotion expression is linked to and can follow spontaneously from emotion experience, this link is not perfect. For example, in an observational study in which participants were confronted with a norm violation in the lab (Sasse et al., 2022), we found that their anger experience (as assessed through in-situ self-report) and their anger expression (as assessed through video ratings) were only weakly associated. Possibly, anger expression sometimes oc-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

curs in deliberate and strategic ways, rather than being a mere side-product of anger experience. Furthermore, in our study, anger expression and behavioral intervention against the norm violation were also weakly associated (Sasse et al., 2022). These findings suggest that anger expression can at times accompany explicit intervention, such as verbal or physical confrontation or sanctioning of a transgressor, but at other times, anger expression can occur as an only response to a transgression, potentially intended as a subtle form of intervention. In other words, anger expression might be employed in a social-communicative function intended to signal disapproval to other witnesses and the transgressor.

Here, we address two questions that follow from this line of thinking: First, do individuals sometimes choose anger expression deliberately as response to others' norm violations, and if so, with what goals? Second, what are the interpersonal effects of anger expression?

F. Anger expression as a subtle form of moral courage

Moral courage is usually operationalized through explicit behavioral reactions to norm transgressions, such as verbally or physically confronting a transgressor or sanctioning them financially. This approach disregards more subtle forms of communication that are characteristic of human interaction, where emotion expression plays an important role. The notion that emotion expression is regularly used for communicative purposes in human interaction is rooted in the social function of emotions. Expressed emotions carry information that others are able to decode. Others attribute emotion expressions to dispositional or situational origins, and these attributions shape how they think about and react to such expressions (Hareli & Hess, 2010; van Kleef et al., 2011). As such, expressed emotions fulfill an important role in social interactions that goes beyond the content of spoken words or behavioral acts. It stands to reason that emotions are – at times – expressed strategically, intended to communicate with and influence others (Sasse et al., 2018).

In the context of moral transgressions, uninvolved witnesses might express anger, not (just) because they feel angry, but because they want to signal to others (i.e., the perpetrator, the victim, or other bystanders) that they disapprove of the observed behavior. Moreover, if we consider anger expression in response to others' wrongdoings as a subtle form of moral

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

courage, witnesses might express anger with the intention to redress the norm violation or change a perpetrator's future behavior.

To explore potential intentions with which anger is expressed by witnesses of others' norm violations, we modified the 3PPG such that participants in the role of the third party could choose to react to an unfair allocation by imposing financial sanctions and/or by expressing anger by sending an emoji to the dictator (Sasse et al., 2024). Our results indicated that anger expression was more than a mere reflection of anger experience but was predicted by the perceived unfairness of the dictator's decision. Moreover, anger expression and financial punishment were associated with distinct goals that participants reported: Financial punishment was best explained by the goal of restoring fairness, whereas anger expression was tied to the goal of motivating behavioral change in the dictator.

These findings lend some support to the idea that anger expression may indeed be a distinct and subtle form of moral courage, but further empirical work, involving experimental designs and a broader range of potential goals is warranted.

G. Interpersonal effects of anger expression in moral courage

Since anger plays an important role in fueling moral courage, and could even be expressed strategically as a subtle form of moral courage, we need to ask: What are the interpersonal effects of anger expression in reaction to others' norm violations? So far, research has started to address this question with regards to the judgements of uninvolved people who learn about the situation.

In a study employing hypothetical descriptions of norm violations in which the perpetrator was confronted by a witness, Eriksson and colleagues (2017) found that participants judged the witness' behavior as more aggressive and less appropriate when the witness was described as angry (vs not described as angry). By contrast, a recent line of studies (Shen et al., 2025) revealed reputational gains for witnesses of moral violations who displayed anger. When a witness was described as displaying anger in these studies participants judged the witness to be more moral and trustworthy.

These diverging findings might be reconciled when considering the severity of the norm violation as a key boundary condition. Indeed, Eriksson et al. (2017) found that negative judgements of angry interveners

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

only occurred when transgressions were mild, but not when transgressions were perceived as more severe. In a similar vein, Shen et al. (2025) reported that the witness of a norm violation was judged more favorably when feeling angry in response to a moral norm violation, but more negatively when angry feelings occurred in a situation without any violation. So, potentially, morally courageous interveners might not incur reputational costs, or even have reputational gains when expressing anger towards severe, morally-relevant transgressions.

In our own work, we asked how angry intervention behavior would be judged if information on the initial transgression was incomplete. Think of the dynamic unfolding of social interactions in everyday life: You might join into a conversation among colleagues, when one colleague responds to the other, possibly in an angry tone "Hey, that's not ok to say". Obviously, you missed parts of the conversation in which a norm violation might have occurred, but you don't know what exactly had been said. How do people make sense of and judge such behavior – and what role does anger expression play? Could anger expression serve as additional information to disambiguate that a norm violation has occurred?

To address these questions, we used an auditorily presented scenario and experimentally varied the ambiguity of the norm violation as well as anger expression during intervention. Participants overheard a conversation between two colleagues, similar to the example above. In the low ambiguity condition, participants heard how one colleague makes a homophobic statement and how the second colleague reacts to it. In the high ambiguity condition, participants only heard the second colleagues' reaction; the actual norm violation was thus ambiguous for them. Depending on the anger expression condition, the second co-worker's reaction consisted in a verbal reprimand that was emotionally neutral or in a very angry tone, or in a third variant there was no verbal reprimand at all. As expected, the colleague's reaction shaped how participants judged the situation and determined in turn their support intentions. However, the perception of the situation as constituting a norm violation and support for the second colleague depended primarily on whether or not this colleague reprimanded, not how (i.e., whether in an angry or neutral tone), and this pattern of results was largely unaffected by ambiguity. At the same time, participants evaluated the colleague who reprimanded angrily as more aggressive and less positive than the colleague who did so in a neutral fashion or who did not reprimand at all. This suggests that the mode of intervening against a transgression (expressing anger or not)

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

might primarily affect how the intervener is judged, rather than how the situation is interpreted.

Further research is needed to gain a clearer understanding of the role of anger expression in morally courageous acts. Existing evidence suggests that there may be a fine line between reputational gain and reputational loss for an intervener who expresses anger, with outcomes depending on situational characteristics such as the severity of the transgression. Other potentially relevant factors include the social status of the transgressor relative to the intervener, the interpersonal relationship between perpetrator and victim, or shared group memberships among the people involved.

Notably, the studies reviewed here focused on the judgements of uninvolved others who learn about or overhear a situation. Yet, it seems plausible that individuals involved in the situation in other roles, such as the victim or the perpetrator of a norm violation may respond quite differently to angry intervention. For instance, for victims of norm violations anger expressed by third-parties might be a source of comfort and validation, and for perpetrators, it might lead them to reconsider their behavior or, to the contrary, react defensively and further escalate the situation. Previous research has addressed reactions to morally courageous intervention considering the perspectives of victims, perpetrators and uninvolved others (e.g., Jordan et al., 2016; Mahon, 2022). However, so far, research has not comprehensively addressed the specific role that anger expression could play for the effectiveness and reputational consequences of moral courage.

H. Conclusion

Our selective review in this chapter illustrates that anger plays multiple key roles in moral courage. First, there is ample correlational evidence that anger is a common emotional response in situations affording moral courage. Importantly, experimental studies demonstrate that anger experienced in response to norm violations can indeed causally contribute to promoting third-party punishment. Second, beyond anger as a motivational fuel, we also reviewed evidence suggesting that anger expression may constitute a subtle form of moral courage. In these cases, anger may serve goals distinct from those of behavioral intervention, such as communicating disapproval and prompting behavioral change in perpetrators. Third, anger expression in response to others' norm violations

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

shapes the interpersonal consequences of morally courageous intervention. Research on how such expressions of anger are judged and reacted to by others has mainly focused on further uninvolved bystanders and suggests that anger first and foremost shapes the perception of the intervener rather than of the situation, but also that effects may be highly context-dependent.

Thus, while our review shows the marked advances we have made in our understanding of the roles of anger in moral courage, it also provides a roadmap for future research: After having established a causal role of anger in "fueling" moral courage, more research is needed to understand the exact mechanisms through which anger exerts this effect. Furthermore, experimental research is needed to understand when and why anger expression is chosen as a means of intervention. Lastly, the contextual factors that shape judgements of and reactions to anger expression in situations affording moral courage need to be investigated to better understand when it helps and when it hurts.

The 3PPG has provided an ideal starting point to study costly third-party punishment of unfairness as a variant of morally courageous action, under high experimental control. To mimic aspects of "real-life" situations affording moral courage, the 3PPG lends itself to modifications, such as introducing ambiguity of the norm violation (Toribio-Flórez et al., 2023) or anger expression as an alternative behavioral option (Sasse et al., 2024). Nevertheless, on the road towards a comprehensive understanding of the roles of anger in moral courage, it will be necessary to scrutinize the generalizability of effects of anger experience and expression to the broad array of situations that afford moral courage in everyday life.

References

Ashburn-Nardo, L., Blanchar, J. C., Petersson, J., Morris, K. A., & Goodwin, S. A. (2014). Do You Say Something When It's Your Boss? The Role of Perpetrator Power in Prejudice Confrontation. *Journal of Social Issues*, *70*(4), Article 4. https://doi.org/10.1111/josi.12082

Baumert, A., Halmburger, A., & Schmitt, M. (2013). Interventions Against Norm Violations: Dispositional Determinants of Self-Reported and Real Moral Courage. *Personality and Social Psychology Bulletin*, *39*(8), Article 8. https://doi.org/10.1177/0146167213490032

Baumert, A., Li, M., Sasse, J., & Skitka, L. (2020). Standing up against moral violations: Psychological processes of moral courage. *Journal of Experimental Social Psychology*, *88*, 103951. https://doi.org/10.1016/j.jesp.2020.103951

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Baumert, A., Mentrup, F. E., Klümper, L., & Sasse, J. (2024). Personality processes of everyday moral courage. *Journal of Personality*, *92*(3), 764–783. https://doi.org/10.1111/jopy.12850

Baumert, A., Sckopke, A., Küchler, G., Sasse, J., & Wagner, J. (2025). Personality characteristics of morally courageous individuals: An extreme group comparison. Manuscript under review.

Brauer, M., & Chekroun, P. (2005). The Relationship Between Perceived Violation of Social Norms and Social Control: Situational Factors Influencing the Reaction to Deviance. *Journal of Applied Social Psychology*, *35*(7), Article 7. https://doi.org/10.1111/j.1559-1816.2005.tb02182.x

Dungan, J. A., Young, L., & Waytz, A. (2019). The power of moral concerns in predicting whistleblowing decisions. *Journal of Experimental Social Psychology*, *85*, 103848. https://doi.org/10.1016/j.jesp.2019.103848

Eriksson, K., Andersson, P. A., & Strimling, P. (2017). When is it appropriate to reprimand a norm violation? The roles of anger, behavioral consequences, violation severity, and social distance. *Judgment and Decision Making*, *12*(4), Article 4.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), Article 6868. https://doi.org/10.1038/415137a

Ferrer, R. A., & Ellis, E. M. (2021). Preliminary evidence for differential effects of integral and incidental emotions on risk perception and behavioral intentions: A meta-analysis of eight experiments. *Journal of Behavioral Decision Making*, *34*(2), Article 2. https://doi.org/10.1002/bdm.2209

Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology*, *36*(2), Article 2. https://doi.org/10.1002/ejsp.297

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology*, *93*(1), 103–115. https://doi.org/10.1037/0022-3514.93.1.103

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Goodwin, R., Graham, J., & Diekmann, K. A. (2020). Good intentions aren't good enough: Moral courage in opposing sexual harassment. *Journal of Experimental Social Psychology*, *86*, 103894. https://doi.org/10.1016/j.jesp.2019.103894

Greitemeyer, T., Fischer, P., Kastenmüller, A., & Frey, D. (2006). Civil courage and helping behavior: Differences and similarities. *European Psychologist*, *11*(2), Article 2. https://doi.org/10.1027/1016-9040.11.2.90

Gummerum, M., Van Dillen, L. F., Van Dijk, E., & López-Pérez, B. (2016). Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *Journal of Experimental Social Psychology*, *65*, 94–104. https://doi.org/10.1016/j.jesp.2016.04.004

Halmburger, A., Baumert, A., & Schmitt, M. (2015). Anger as driving factor of moral courage in comparison with guilt and global mood: A multimethod approach. *European Journal of Social Psychology*, *45*(1), Article 1. https://doi.org/10.1002/ejsp.2071

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition & Emotion*, *24*(1), Article 1. https://doi.org/10.1080/02699930802613828

Hechler, S., & Kessler, T. (2018). On the difference between moral outrage and empathic anger: Anger about wrongful deeds or harmful consequences. *Journal of Experimental Social Psychology*, *76*, 270–282. https://doi.org/10.1016/j.jesp.2018.03.005

Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, *19*(4), Article 4. https://doi.org/10.1007/s10683-015-9466-8

Kemper, N. S., Campbell, D. S., & Reiman, A.-K. (2022). See something, say something? Exploring the gap between real and imagined moral courage. *Ethics & Behavior*, *0*(0), 1–22. https://doi.org/10.1080/10508422.2022.2104282

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. *Journal of the European Economic Association*, *11*(3), 495–524.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Appleton-Century Crofts.

Lotz, S., Baumert, A., Schlösser, T., Gresser, F., & Fetchenhauer, D. (2011). Individual differences in third-party interventions: How justice sensitivity shapes altruistic punishment. *Negotiation and Conflict Management Research*, *4*(4), Article 4. https://doi.org/10.1111/j.1750-4716.2011.00084.x

McMahon, S. (2022). How Helpful Is Bystander Intervention? Perspectives of Dating and Sexual Violence Survivors. *Violence Against Women*, 10778012221117596. https://doi.org/10.1177/10778012221117596

Molho, C., Tybur, J. M., Güler, E., Balliet, D., & Hofmann, W. (2017). Disgust and anger relate to different aggressive responses to moral violations. *Psychological Science*, *28*(5), Article 5. https://doi.org/10.1177/0956797617692000

Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-17286-2

Nelissen, R. M. A., & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision Making*, *4*(7), Article 7.

Niesta Kayser, D., Greitemeyer, T., Fischer, P., & Frey, D. (2010). Why mood affects help giving, but not moral courage: Comparing two types of prosocial behaviour. *European Journal of Social Psychology*, *40*(7), Article 7. https://doi.org/10.1002/ejsp.717

Ortony, A., Clore, G. L., & Collins, A. (2022). *The Cognitive Structure of Emotions*. Cambridge University Press.

Osswald, S., Greitemeyer, T., Fischer, P., & Frey, D. (2010). What is moral courage? Definition, explication, and classification of a complex construct. In *The psychology of courage: Modern research on an ancient virtue*.

Pedersen, E. J., McAuliffe, W. H. B., Shah, Y., Tanaka, H., Ohtsubo, Y., & McCullough, M. E. (2020). When and Why Do Third Parties Punish Outside of the Lab? A

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Cross-Cultural Recall Study. *Social Psychological and Personality Science*, *11*(6), 846–853. https://doi.org/10.1177/1948550619884565

Pouwels, J. L., van Noorden, T. H. J., & Caravita, S. C. S. (2019). Defending victims of bullying in the classroom: The role of moral responsibility and social costs. *Journal of Experimental Social Psychology*, *84*, 103831. https://doi.org/10.1016/j.jesp.2019.103831

Rodrigues, J., Liesner, M., Reutter, M., Mussel, P., & Hewig, J. (2020). It's costly punishment, not altruistic: Low midfrontal theta and state anger predict punishment. *Psychophysiology*, *57*(8), Article 8. https://doi.org/10.1111/psyp.13557

Sasse, J., & Baumert, A. (2024, September). Ärgerausdruck als Form der Zivilcourage [Research talk]. Presentation at the 53rd Conference of the German Psychological Society & 15th Conference of the Austrian Psychological Society, University of Vienna, Austria.

Sasse, J., Richiusa, C., & Baumert, A. (2025). Investigating the Causal Effects of Anger on Costly Third-Party Punishment. https://doi.org/10.31234/osf.io/c369u_v2

Sasse, J., Halmburger, A., & Baumert, A. (2022). The functions of anger in moral courage – Insights from a behavioral study. *Emotion*, *22*(6), 1321–1335. https://doi.org/10.1037/emo0000906

Sasse, J., Spears, R., & Gordijn, E. H. (2018). When to reveal what you feel: How emotions towards antagonistic out-group and third party audiences are expressed strategically. *PLOS ONE*, *13*(9), Article 9. https://doi.org/10.1371/journal.pone.0202163

Shen, X., Anderson, R. A., & Pizarro, D. A. (2025). People judge third-party anger as a signal of moral character. *Journal of Experimental Social Psychology*, *120*, 104765.

Skitka, L. J. (2012). Moral convictions and moral courage: Common denominators of good and evil. In M. Mikulincer & P. R. Shaver (Hrsg.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (S. 349–365). American Psychological Association. https://doi.org/10.1037/13091-019

Toribio-Flórez, D., Saße, J., & Baumert, A. (2023). "Proof Under Reasonable Doubt": Ambiguity of the Norm Violation as Boundary Condition of Third-Party Punishment. *Personality and Social Psychology Bulletin*, *49*(3), 429–446. https://doi.org/10.1177/01461672211067675

Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2010). An Interpersonal Approach to Emotion in Social Decision Making: The Emotions as Social Information Model. In *Advances in Experimental Social Psychology* (Bd. 42, S. 45–96). Academic Press. https://doi.org/10.1016/S0065-2601(10)42002-X<

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*(1), Article 1. https://doi.org/10.1037/0022-3514.51.1.110

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

On the Effects of Transparency on Labour Supply in Redistribution Systems

Marina Chugunova, Andreas Nicklisch & Kai-Uwe Schnapp

This paper analyzes whether there is a strategic incentive for redistribution systems to reveal information on the productivity of transfer recipients. The transparency of productivity is potentially able to stimulate both the work effort of taxpayers and transfer recipients. Experimental results do not support this claim. Rather than stimulating effort provision, transparency creates a negative reference point for many taxpayers. Thus, monitoring is negative for the overall productivity suggesting that there is an institutional advantage of deliberate ignorance for redistribution systems. Choosing not to reveal information yields superior outcomes and serves in the best interest of all members of the society.

A. Introduction*

In 2016, Christoph Engel and Ralph Hertwig wrote their seminal paper on deliberate ignorance, the decision to shrink from knowing information as this creates a strategic advantage. Their paper "Homo ignorans: Deliberately choosing not to know" represents a milestone in our understanding of human behavior as it departs from the traditional assumption that curiosity is a dominant human characteristic. They present examples of deliberate ignorance from a wide range of domains, demonstrating that not knowing sometimes yields superior outcomes.

A large body of literature refers to the approach by Hertwig and Engel (2016). Investors, for example, avoid looking at their financial portfolios when the stock market is down (Andries and Haddad, 2020; Sicherman et al., 2016). Consumers do not check whether the consumption of cheap products causes a negative externality to some third parties (Bartling et al., 2015). Kozyreva et al. (2020) argue that mental habits that help to

* Financial supports by the German Research Foundation (DFG) for Nicklisch (Ni 1610/1-1) and Schnapp (Schn 1282/4-1) are gratefully acknowledged.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

avoid being tempted by attention-grabbing content are important components of modern digital information literacy. Finally, De Bruin et al. (2021) investigate the degree of news avoidance during the first months of the Covid-19 pandemic.

In this study, we discuss another facet of strategic ignorance, the advantage of a non-transparent institutional setting that support poor members of the society by means of redistribution mechanisms. That is, government institutions organize a transfer of income between rich or advantaged members of society and poor or disadvantaged members of society. The rich members provide the poor sufficient financial resources so that the transfer guarantees the sufficient income to participate in the normal economic, social, and political life of a society.

However, the willingness to contribute to the redistribution system for those who give depends on everybody's effort contributing fairly and according to his or her ability. In their study on redistribution in Denmark and the USA, Aarøe and Petersen (2014) show that the support of those who give depends on how much information on the recipients is available: if recipients are judged as being lazy, less support is granted in comparison to the situation in which the recipient is going through a streak of bad luck.

Increasing the transparency on the recipients' performance prior to the redistribution (i.e., monitoring the recipients' effort) would allow those who give to reciprocate: they would provide little support to those who are considered to be lazy and would provide a lot of support to those who are considered to be unlucky. Yet, a redistribution system seeks to guarantee a constant flow of transfers to all in need. Therefore, increasing the transparency could potentially lead to decreasing efficiency of the redistribution system.

To test this claim, we run a series of real-effort laboratory experiments with the provision of information on the performance of transfer recipients by means of unilateral monitoring. Each experiment consists of three phases with the game being played by pairs of players. In each phase participants solve math tasks for 30 minutes. Participants are paid for each correctly solved task and are free to choose when to work and when to be idle.

In the first phase all participants add or subtract pairs of three-digit numbers. In the second phase one player of each pair is randomly assigned high productivity while the other is assigned low productivity. Both players solve math exercises again for 30 minutes. The high productivity player adds or subtracts three-digit numbers as before, while the low productivity player adds or subtracts five-digit numbers. The third phase is

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

equivalent to the first for both players in a group but players earn twice the amount of money for every correctly solved task. Only players who earned an income net of transfers beyond a threshold level in the second phase are allowed to take part in the high-income third phase.

Our baseline treatment (BA) follows the description above and has no interaction between players. The transfer treatment (TR) introduces a thirty percent income tax on the highly productive players. At the end of the second phase the tax revenue is transferred to the low productivity player within a group. The transfer increases the likelihood that disadvantaged players take part in the high-income third phase of the experiment despite their randomly assigned low productivity in the second phase.

The feedback treatment (FB) is identical to TR but introduces additionally unilateral monitoring. The highly productive players (taxpayers) can constantly observe the performance of the low productivity players (transfer recipients). However, there is no feedback in the opposite direction (from taxpayers to recipients), since we want to isolate the effect of this unilateral transparency on the overall productivity and avoid an interaction between taxpayers and transfer recipients. As such, in TR and FB transfers systematically target disadvantaged peers. What varies endogenously is recipients' own contribution to meeting the subsistence income threshold and to qualifying for the third phase.

We analyze the effects of transparency (i.e., provision of information) on the productivity for both taxpayers and transfer recipients. In our experimental societies players have heterogeneous productivity, but a homogeneous minimum income constraint (a subsistence income). In BA there is no redistribution, in TR redistribution is introduced, and, finally, in FB redistribution is coupled with the unilateral monitoring. Our results indicate that ignorance is a blessing. In TR both players in a pair increase their productivity significantly as compared to BA. The introduction of monitoring in FB does not, however, affect the productivity on the side of the taxpayers, and significantly reduces the productivity of transfer recipients. Hence, it seems that recipients try to stimulate large transfers by showing little productivity. Consequently, for our experimental economy as a whole, monitoring harms rather than promotes the total amount of labour supplied and thus efficiency.

These results relate to a small number of studies on circumstances in which transparency backfires and – against the traditional view in public economics – is not positive and desirable, but leads to lower production in a society. For instance, public monitoring of decision procedures in

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

developing countries is generally considered to be usefull against corruption, but provides important to identify those who are clues recipients of bribes (Kolstad and Wiig, 2009; Lindstedt and Naurin, 2010). Likewise, if people have the opportunity to embezzle part of a common resource while others cannot sanction this behaviour, full transparency increases rather than prevents embezzlement (Khadjavi et al., 2017).

The rest of the paper proceeds as follows: In Section 2 we present our experimental design and procedure, in Section 3 we develop both our theoretical analysis and hypotheses, in Section 4 we present our results, and finally in Section 5 we draw conclusions based on our findings.

B. The Experimental Design

I. The Game

In our real-effort experiment, we asked participants to conduct simple computations and did so in three phases of 30 minutes each. In Phase I we measure the players' motivation and ability to perform the computing task, awarded them an income depending on how well they did, and advanced all participants to Phase II. In Phase II we applied our treatment but only those that passed an income threshold were advanced to Phase III. Phase III is a high income phase. Players need to reach Phase III in order to earn a decent amount of money in the experiment.

In all three phases all players are tasked with computing sums and differences without technical support, however, the concrete assignments, differ during the phases. In Phases I and III players compute the sums or the differences of two three-digit numbers. If a task is solved correctly, a new task appears on the screen, if not, the same task is presented again. Per correctly solved task players earn a piece rate ϕ. In all phases, the number of tasks is unlimited and thus players are free to decide how hard to work. By opting for a tedious and repetitive task across an extended period of time, we minimize the joy and intrinsic motivation of work and hence increase the importance of monetary incentives. To avoid an experimenter demand effect, players are not offered an alternative leisure activity.[1]

1 Research on productivity shows that not performing a dull task that requires a lot of concentration can also be considered as leisure as it saves cognitive bandwidth (e.g, Kool et al., 2010).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The design of Phase II, the treatment phase, differs in two respects: while players are still paired in groups of two each player is assigned a different computing task. And while one player solves the same task as in Phase I, that is, computing the sum or the difference of two three-digit numbers, the other player has to solve a more difficult and thus more time-consuming task, that is, computing the sum or the difference of two five-digit numbers. Irrespective of the task's difficulty, both types of players receive the same piece rate ϕ for a completed task.

As explained to the players between Phase I and II, advancing from Phase II to Phase III requires for all players to meet an an exogenously imposed income constraint. If a player earns less than S Euros in Phase II, she will not be advanced to the high income Phase III but is still required to wait in the laboratory until the other players have completed Phase III. We choose $\phi = 0.05$ Euros, $S = 3.20$ Euros. Both parameters are calibrated such that on average the threshold is slightly too high to be reached by players assigned the five-digit task, but easy to reach for players assigned with computing three-digit numbers.

In treatment conditions TR and FB, players who solve three-digit tasks in Phase II have to pay a tax rate $\tau = 0.3$ of their income. The tax revenue is transferred to the player of the other type in their group. We therefore refer to these players as "taxpayer" (T) and to the other players as "recipients" (R).[2] The types in each pair are randomly assigned to the players at the beginning of the experiment.

Our three treatment conditions Base (BA), Transfer (TR) and Feedback (FB)[3] introduce both different "production" conditions between taxpayers and different degrees of awareness about those production conditions by recipients: the treatment condition BA follows this outline without any change. In TR and FB, the transfer is meant to enable players R to reach the substance threshold (or make reaching the threshold substantially more likely). TR and FB differ on one dimension only: in FB throughout the second phase player T receives a real-time feedback on the performance of the player R, this is not the case in TR. Thus, on one hand, player T can exactly see how much effort the paired player R exercises to reach the threshold

2 Explanations given to the subjects in the lab use the labels "Type A" and "Type B" to avoid any unintended connotations. Instructions and example screens are available online at https://osf.io/bft6w

3 We reported the results from the former two treatment conditions already in Chugunova et al. (2022).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

on her own. On the other hand, player T knows approximately how much player R still needs to pass the threshold. Therefore, in TR the willingness to assist the disadvantaged will rather follow a logic of unconditioned (or even unintended) solidarity, while FB relies on some sort of reciprocity (R extends effort, T supports via obligatory transfer).

All parameters of the game (i.e., difference in the complexity of tasks between player types, piece rates for both types, subsistence income, unilateral monitoring) are common knowledge. The implemented monitoring scheme is somewhat imperfect. While the taxpayer observes how many tasks the assigned partner solves, she has no information on player R's performance in the first phase. Hence, a taxpayer is not able to tell whether her counterpart performs to her true ability.

We would like to stress two important features of our design: (1) The accountability principle has been proven to play a role in redistribution decisions. That is, people are less willing to pay for redistribution if recipients are accountable for their situation (Cappelen et al., 2013). The type of disadvantage implemented in our experiment, however, cannot be ascribed to the accountability of players of type R. This interpretation is further supported by the random assignment of players to their types. Hence, we deliberately eliminate this rationale on the side of the taxpayers. (2) Previous studies on donations show that if donor incomes are earned (i.e., donors work for their income), donors are less likely to redistribute (e.g., Cherry et al., 2002; Ogawa et al., 2012). That is, if they feel entitled to their earnings, they are less likely to give. By employing a real effort design we create unfavorable conditions for redistribution by reinforcing the perception of being entitled to one's earnings. So in contrast to earlier experiments (e.g., Agranov and Palfrey, 2015), our design diminishes the easiness (acceptance) of giving.

II. Procedure

The experiment was conducted at the Wiso research Lab at the University of Hamburg using z-tree (Fischbacher, 2007) in spring and autumn 2016. Subjects were mostly students of various majors of the University of Hamburg, recruited online via hroot (Bock et al., 2014) and randomly assigned to treatments. No subject participated more than once.

Participants were seated in cubicles. Before the experiment started, they placed their cellphones into envelopes provided and sealed by the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

experimenters. The envelopes were returned to them only after they had been paid. They were also asked to keep their bags and personal belongings outside of their cubicles, ensuring that they could not use any electronic devices to help with the calculations. They were warned that the experimenter would monitor the use of forbidden devices throughout the experiment and that participants breaching the rules would be immediately removed from the experiment without being paid. The use of pen and paper, which were provided, was allowed.

After the participants were settled in, they received the instructions for Phase I in writing and by the experimenter reading them out loud. Participants were then randomly paired in sets of twos; those pairs stayed unchanged throughout the experiment, and they played simultaneously but independently of each other. The participants learned if they were player A or B right before the calculations in Phase I began, without, however, knowing the meaning attached to A and B. After Phase I was over the instructions for Phases II and III were distributed, and read out aloud, and the meaning attached to being player A or B was revealed. After each phase there was a three-minute break for each participant, without them having the opportunity to communicate with each other.

We conducted 10 sessions with 180 participants in total. Sessions run with 16–20 subjects per session. Overall, our data set consists of 32 pairs (i.e., independent observations of 31 players T and 32 players R) for BA, 30 (30 players T and 30 players R) for TR, and 28 (27 players T and 28 players R) for FB.[4] At the end of the experiment subjects were paid privately and in cash. Payments ranged from 3.75 to 52.40 Euros with an average of 24.30 Euros for approximately 140 minutes.

C. Hypothesis

I. Effort provision without redistribution

In the following, we assume that players hold a twice differentiable, additive utility function with concavely increasing benefits and convexly

4 We removed two participants (one participant in BA and one in FB) from the sample and did not use their data in the analysis, since individuals behaved nonsystematically different (their choices were more than 2.5 standard deviations away from the mean of relative performance measured as described below).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

decreasing working costs. Both, benefits and costs depend on effort, while we expect players to differ with respect to their costs (i.e., their skills). That is, for two players i and j we label player i as more talented than player j if for every given work speed $\underline{v}$ it holds $c_j(\underline{v}) > c_i(\underline{v})$. In turn, players gain benefits from working which depends on the piece rate payment ϕ they receive and the work speed (per minute) v_i with which they choose to work (i.e., their endeavor). In sum, this yields

$$u(v_i) = w(\phi v_i) - c_i(v_i) \tag{1}$$

such that utility u_i of player i depends on the benefit w with $u'(w) > 0$ and $u''(w) < 0$, and the individual costs c_i with $u'(c_i) < 0$ and $u''(c_i) < 0$. It holds that $w'(v_i) > 0$ and $w''(v_i) < 0$, and $c'_i(v_i) > 0$ and $c''_i(v_i) > 0$.

In the first and – if possible – the third phase of the experiment, players maximize their utility by choosing v_i such that $\phi w'(\phi v_i) = c'_i(v_i)$ (first order condition). In the second phase, we have to consider that v_i may or may not be high enough to qualify i for the third phase. Thus, i's utility function changes to

$$u(v_i) = \begin{cases} w(\phi v_i) - c_i(v_i) & \text{if } 30 v_i \phi < S \\ w(\phi v_i) + U_i - c_i(v_i) & \text{otherwise.} \end{cases} \tag{2}$$

for U_i denoting i's expected utility in the third phase. This creates the incentives for players in the "medium effort cost range" to choose exactly a work speed v' that allows them to just reach the subsistence income. In turn, both players with very low or very high effort cost choose the same work speed as in phase I: their "initial" work in phase I is so low or high, so that the subsistence income requirement does not affect their decisions.

Hence, in phase II of BA, one can predict a partition of both advantaged and disadvantaged players into three groups. The first group chooses a work speed which yields an income below the subsistence income requirement. Within this group, people increase their work speed if they are better at solving the tasks (i.e., if they have lower effort costs). The second group chooses a work speed v' which yields an income at the subsistence income threshold. All players within this group choose this speed regardless how good they are at solving the tasks. Finally, players in the third group choose a work speed which yields an income beyond the subsistence income requirement. Within this group, people increase their work speed if they are better in solving tasks.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Effort provision with redistribution and monitoring

When analyzing the choice of a work speed in the treatments with redistribution, one has to consider the two types of players separately. For players of type T, the piece rate per solved task ϕ decreases to $(1 - \tau)\phi := \phi'$. For players of type R, the subsistence income requirement S decreases to $S - U' := S'$ with U' being the expected transfer they receive from their matched taxpayer.

Replacing ϕ with $\phi' < \phi$, it thus becomes "harder" for a taxpayer to meet the subsistence income requirement, as they earn only $(1 - \tau)\phi$ for each task. This implies that the range for the three groups of players T "moves to the right": a broader range of potential optimal work speeds now belong to the first group earning less than the subsistence income, while a smaller range of players choose a work speed leading to more than the subsistence income. Finally, taxpayers who were to choose a work speed leading to an income beyond the subsistence income now choose a work speed corresponding with the subsistence income.

In turn, replacing S with $S' < S$, it becomes "easier" for transfer recipients to meet the subsistence income requirement implying that the range for the three groups of players "moves to the left".

In other words, from a theoretical point of view, redistribution is nothing more than a wage cut for taxpayers, similar to Kessler and Norton (2016), and an expected decrease in the subsistence income threshold for transfer recipients. Since we calibrated the threshold, piece rate and tax rate such that many taxpayers belong to those who were to choose a work speed leading to an income beyond the subsistence income even with taxation, whereas recipients belong to those who choose a work speed close to the subsistence income even without taxation, we predict a positive effect of redistribution on productivity in phase II:

H_1: Redistribution imposes benefits since players produce in total more in TR than in BA.

With transparency as designed in our experiment a taxpayer can observe the effort of a recipient, but not vice versa. Earlier research has shown that monitoring among peers has enormous effects on behavior (e.g., Bandiera et al., 2005; Mas and Moretti, 2009). The reason for this is a human inclination to reciprocate and match others' behavior (e.g., Fischbacher et al., 2001): if two people work on a joint project or on similar tasks, they tend

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

to harmonize their effort. In our context, this could mean that transfer recipients and taxpayers approximate their work speed with monitoring:

H_2: Transparency amplifies the connection between work speeds of players R and T in FB, while the connection is weaker in TR.

Yet, knowing that both work speeds by R and T become more similar while R being monitored offers R a strategic leverage: if R accelerates her work speed (at least in the beginning of phase II), she will stimulate the work speed of T as well and it increases her chances to meet the subsistence income. In other words, the unilateral monitoring of R by T may create an additional incentive to increase R's work speed. In this case, there would be a strategic benefit of transparency. Those considerations lead us to formulate a hypothesis that follows the former argumentation:

H_3: Transparency increases the benefits of redistribution, since particularly players R produce more in FB than in TR.

However, we are aware that also the opposite argumentation appears plausible: if R do not realize the strategic leverage, they may even create a negative reference point for T, slowing T down. In this case, transparency backfires: the unilateral monitoring yields a decrease in efficiency and a redistribution system that keeps taxpayers ignorant with respect to recipients work effort would provide superior efficiency.

D. Results

I. Does redistribution lead to more production on the side of the taxpayers?

To measure the performance of the participants controlling for their initial ability we divide the number of correctly solved tasks in the second phase by the number of correctly solved tasks in the first phase. We further on call this relation between stages "relative performance". The relative performance of 1 means that a player solves in the second phase exactly the same number of tasks as in the first one; a number less than 1 means that she reduces her effort and solves less, and the number more than 1 means respectively that she improves her performance. Thus, one can also interpret relative performance as a percentage of change as compared to the first phase.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

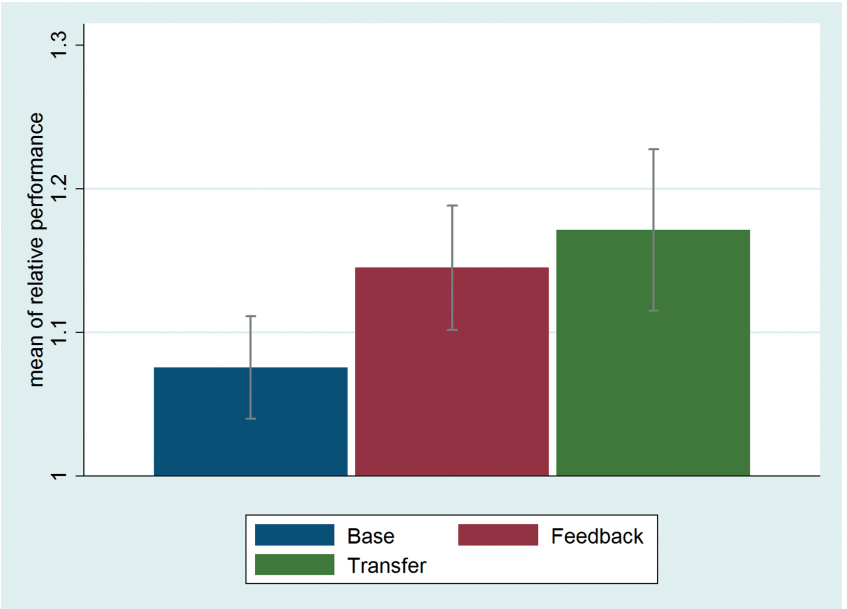


Figure 1: Average relative performance of players T across treatments. 95% confidence interval.

Figure 1 displays the response to taxation on the aggregate level. Overall, there are no implicit costs of redistribution for players T, but rather benefits in terms of productivity. The relative performance of players T increased on average by 7,6% in BA, by 17,1% in TR, and by 14,5% in FB. As shown in the Fig. 1, redistribution in TR and FB improves productivity as compared to BA. Mann-Whitney U-tests[5] confirm H_1 by showing that the relative performance of players T in FB (1.15) and TR (1.17) is significantly higher than in BA (1.08; p=0.03 with FB and p=0.005 with TR). However, there is no significant increase or decrease due to the introduction of the monitoring. In other words, the relative performance in TR and FB are not significantly different from each other (p=0.45). Thus, we cannot confirm our hypothesis H_3. The presence of monitoring leads to a slight, but insignificant decrease in relative performance.

5 All nonparametric statistical tests reported are two-tailed and take individuals as units of observations.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Does monitoring lead to more production on the side of the taxpayers?

We calibrate the subsistence income requirement such that it provides an extremely mild restriction for players T. Fig. 2 displays how restrictive the imposed subsistence income threshold is for every individual as well as individual response to the introduction of taxation. Blue bars show how players T need to adjust their performance in the second phase to meet the subsistence income requirement; the vast majority of players could lower their performance considerably and still meet this requirement. Red bars show the actually observed relative performance; only 3 (2) players T in FB (TR) are not able to meet the requirement and the majority of subjects boosted their performance.

To analyze how differences in initial skills influence the relative performance in TR and FB in greater detail, we divide players T into high and low "talent" groups according to the median of performance in the first phase of each treatment. We observe that it is the higher relative perfor-

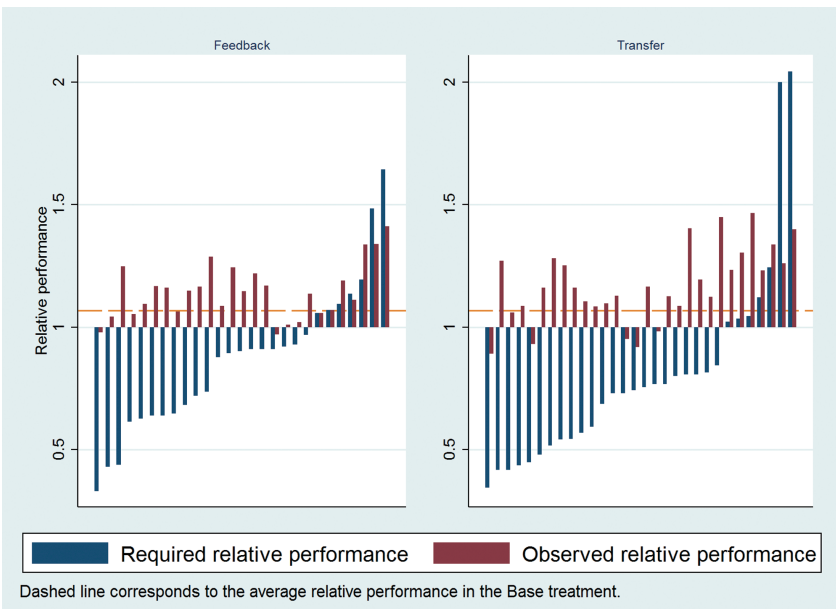


Figure 2: Relative performance required to reach the subsistence threshold and actual relative performance for each taxpayer. Dashed line depicts the median relative performance in the Base treatment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

mance among low talent taxpayers that predominantly drives the overall higher relative performance level in TR treatment (see Fig. 3). Among low talent taxpayers the highest performance can be observed in TR (1.25), which is significantly different from performance in BA (1.09; p=0.02) and in FB (1.16; p=0.07). The performance of low talent players T in FB is non-different than that in BA (p=0.24).

On the other hand, among high talent taxpayers, there is no significant difference between relative performance in TR (1.09) and FB (1.13; p=0.45) and TR and BA (1.06; p=0.35). The relative performance in FB is, however, significantly different from BA (p=0.04). Thus the introduction of monitoring causes some increase in effort among the high talent taxpayers, but at the same time decreases the effort of the less talented ones resulting on average in the slight but insignificant decrease in productivity due to the introduction of monitoring.

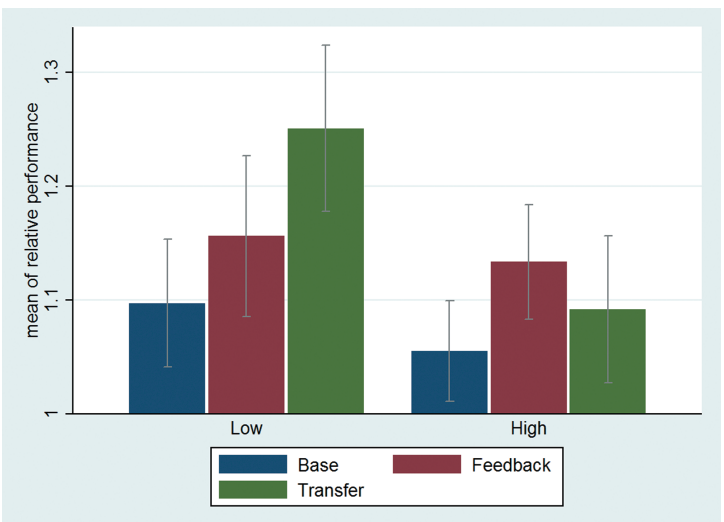


Figure 3: Relative performance of taxpayers: high versus low talent.

Summarizing all of the above:

Result 1: Redistribution without monitoring leads to a substantial increase in productivity for low talented players T; there is no such effect for high talent players T. Yet, monitoring of transfer recipients reverses the effect bringing about higher relative performance for high talent taxpayers and eliminating the enhancing effect for low talent players T. The

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

magnitude of the effect is bigger for the former. On average productivity is enhanced by the introduction of the redistribution, but there is no additional boost due to transparency.

III. Does monitoring lead to more production on the side of the recipient?

Since transfer recipients faced a more difficult task in the second phase, they slowed down by approximately half, resulting in relative performance of 0.48 in BA, 0.55 in TR and 0.49 in FB (see Fig. 4). Mann-Whitney U-tests confirm that the relative performance of players R in the TR is significantly different than in BA (p=0.07), however, results reveal no significant differences in relative performance between FB and TR (p=0.40) or FB and BA (p=0.59). That is, against our H_3, recipients in the system of unilateral monitoring do not use their performances strategically to affect the behavior of the taxpayer by triggering reciprocity. Rather, it seems that players R are not aware of the strategic implication of their own work speed. Furthermore, it seems that monitoring demotivates players R as we observe that without monitoring more players attempt to reach the threshold: the correlation of performance in the first and second phase among players R is 0.40 (p=0.02) in BA, but 0.71 (p=0.000) in TR, while the correlation between performances during the two phases drops again in FB (0.39, p=0.04) (see Fig. 6).

To gain additional insights we divide players R into low and high talented subgroups according to the median of performance in each treatment. In Fig. 5 it can be seen that the higher relative performance in TR is predominantly driven by less talented players. Although the differences between high and low talent recipients are not significantly different across all treatment conditions, we find a substantially enhanced performance of the low talent transfer recipients in TR. In contrast, it seems that monitoring neither spurs the performance of high talent nor of low talent players R.

Fig. 7 sheds the light on the dynamics of the work speed of recipients over time. It shows the average time in seconds needed by players R to solve a task over the course of the second phase[6]. We observe a decrease

6 To adopt to the random fluctuation in the difficulty of the task, we take a moving average of 5 tasks: the current task and two tasks before and after all weighted equally.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

in working speed around task 15, which is the most pronounced in the FB treatment. It can be connected to the understanding of how hard meeting the threshold is for disadvantaged players. The speed improves (i.e., decreases) over time as bad performers "drop out". The average work speed is the highest in TR, against opposing H_3. As a consequence of the higher speed the total number of tasks solved by the poorest performing quarter of players (i.e., the dashed lines) is significantly higher in TR (0.48) than in FB (0.24 p=0.005) and BA (0.28 p=0.02). Summarizing all of the above:

Result 2: Monitoring leads to a substantial decrease in productivity, particularly for low talent players R; there is little evidence that monitoring affects the performance of high talent recipients.

IV. Do taxpayers take the performance of the transfer recipients into account?

Answering the question of whether taxpayers indeed reciprocate or rather react to the current need of the recipients requires a deeper look into the performance of individual pairs. To do so we run an Arellano-Bond dynamic panel data estimation. Including a lagged dependent variable as an explanatory variable allows the model to partially adjust. We split the overall time of the second phase into one-minute sub-periods and run the panel regression with the number of correctly solved tasks of the taxpayer in the one-minute sub-period as a dependent variable. As the implemented monitoring system is unilateral, the direction of the effect is certain.

Table 1 displays two specifications (1–2)[7] of the estimation which differ only with respect of how quickly the taxpayer updates her behavior in response to the performance of the transfer recipient: in the column (1) we assume that the update is immediate (i.e., takes place at the same minute), and in the column (2) we allow for a one minute lag (i.e., taxpayers observe the performance and adopt in the next minute). Apart from the measure of performance of the recipient (i.e., number of tasks solved by the recipient in the respective one minute subperiod), we include in the estimation the number of the tasks the recipient solved up till the current period ("total solved by recipient")[8] and three dummy variables:"threshold with

7 Specifications 3 and 4 will be discussed in Section IV.

8 It is the actual number taxpayers observe on their screen in any respective minute.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

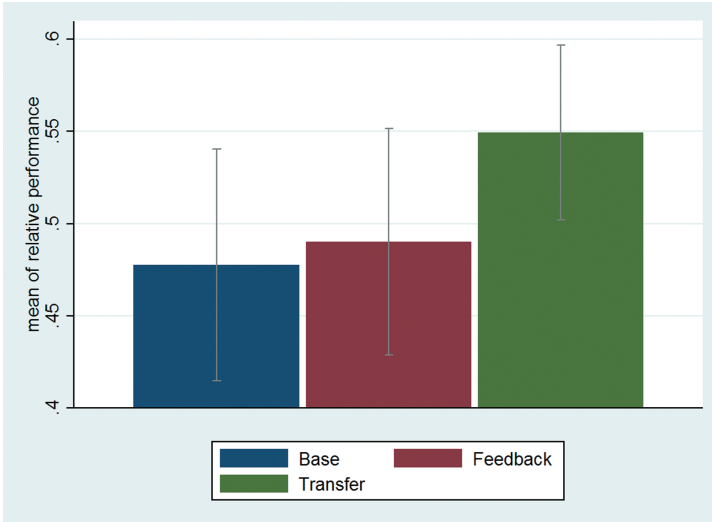


Figure 4: Average relative performance of recipients across treatments. 95 % confidence interval

transfer (recipient)" takes the value of one if the recipient is already above the threshold taking into account the transfer which is due to happen, respectively "threshold (taxpayer)" takes the value of one if the taxpayer is already above the threshold taking into account the transfer which is due to happen and finally, the control "faster" takes the value of one if the recipient happens to solve more tasks than taxpayer in the respective one minute period. Individual abilities are eliminated through fixed effects, standard errors clustered at the individual level.

Our results show that neither current nor lagged performance of the transfer recipients affects the performance of the taxpayer: recipient outperforming the taxpayer and the taxpayer reaching the threshold are the only significant coefficients. If the recipient outperforms the taxpayer, the taxpayer solves roughly 1.2 tasks per minute less. Considering the size of the constant it implies a slowdown of almost 30 %. The coefficient is negative indicating that if the recipient is so fast that she does not require any help. The positive effect of passing the threshold is somewhat more surprising. It means that after reaching the threshold, the taxpayer speeds up and gets more productive. It can be explained by the motivational spike of taxpayers from being able to reach the threshold and qualifying for further production. Fig. 8 looks deeper into the performance of the players three

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

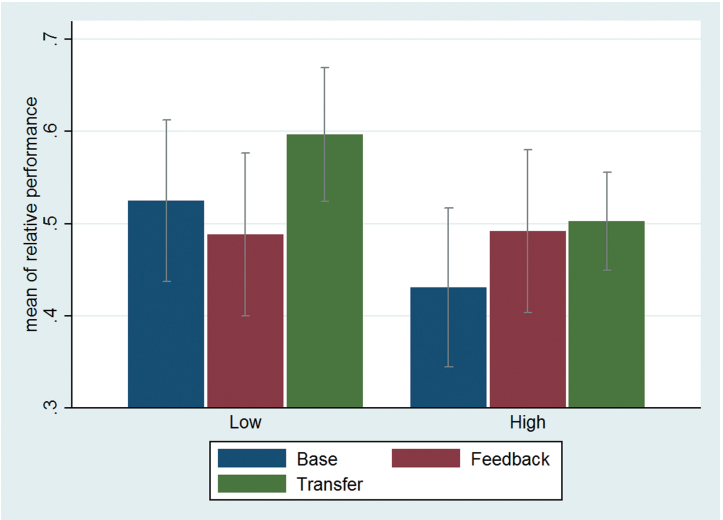


Figure 5: Relative performance of recipients: high versus low talent.

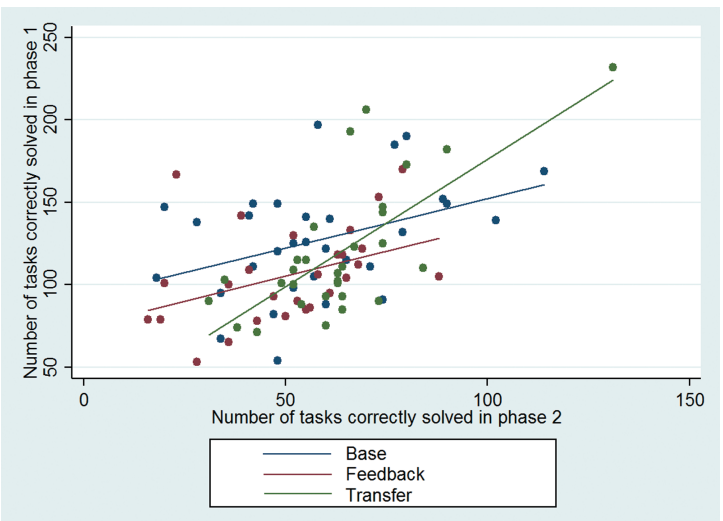


Figure 6: Correlation between correctly solved tasks of players R in the first and second phase.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

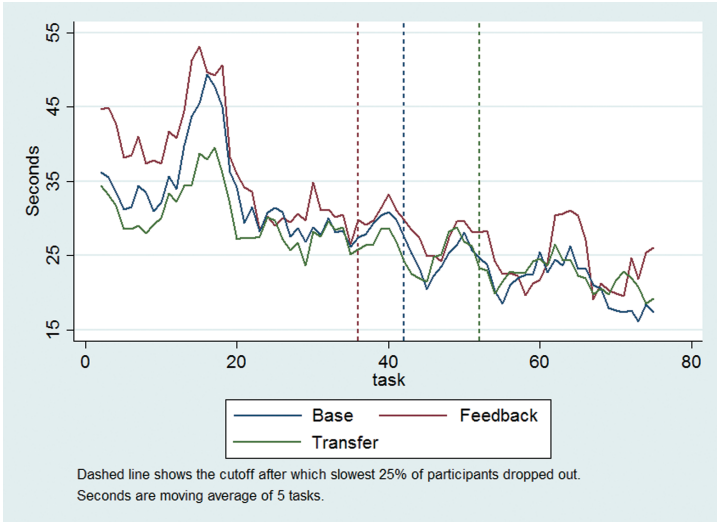


Figure 7: Work speed dynamics of recipients.

minutes (i.e., 10 % of the phase time) before and after the threshold. It is possible that some players did not reach the threshold at least 3 minutes before the end of the phase; consequently, the average number of tasks completed after the threshold may be affected by the attrition of slower players. In other words, fast players produced for 3 minutes before and 3 minutes after the threshold, while slowest players produced for 3 minutes before and – for example – 2 minutes after and thus the average may be influenced by the lower number of slower players after the threshold as such. Although we can not eliminate this bias completely, Fig. 8 shows how many minutes were taken into account for calculating the aggregate (white numbers at the bottom of each bar)[9]. That is, in the FB treatment some slower players had less than three minutes after the threshold: two players had one minute after the threshold only and three players had two minutes. Wilcoxon signed-ranks test confirms that the speed of production is not significantly different three minutes before and after the threshold (BA p=0.26, TR p=0.46, FB p=0.69). In BA and FB taxpayers slightly slow down in solving the exercises, while in TR they slightly sped up.

9 Time data are missing for some sessions in BA and TR, therefore the graph is based on the data for 27 taxpayers in BA, 20 taxpayers in TR and 27 taxpayers in FB.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 1: DPD Estimation: Determinants of Performance.

Variables	(1)	(2)	(3)	(4)
L.Correctly Solved Tasks by Taxpayer	–0.0192	–0.0192	–0.0193	–0.0195
	(0.0333)	(0.0331)	(0.0331)	(0.0333)
Correctly Solved Tasks by Recipient		–0.00958		0.0399
		(0.0524)		(0.0575)
L. Correctly Solved Tasks by Recipient	0.0139		0.00810	
	(0.0647)		(0.0637)	
Total Solved by Recipient	7.86e-05	0.000259	–1.54e-05	–0.000168
	(0.00753)	(0.00758)	(0.00772)	(0.00777)
Threshold with Transfer (Recipient)	0.0999	0.0997	0.125	0.127
	(0.245)	(0.244)	(0.251)	(0.254)
Threshold (Taxpayer)	0.613***	0.611***	0.596***	0.592***
	(0.183)	(0.183)	(0.183)	(0.179)
Faster	–1.273*	–1.262*	–1.331*	–1.366*
	(0.708)	(0.713)	(0.752)	(0.761)
Desperate			0.624*	0.708**
			(0.319)	(0.352)
Constant	4.467***	4.501***	4.198***	4.112***
	(0.355)	(0.362)	(0.336)	(0.343)
Observations	754	754	754	754
Number of id	27	27	27	27
Wald chi2	33.44	33.59	51.98	50.99

Robust standard errors in parentheses
** p<0.01, ** p<0.05, * p<0.1

V. When do taxpayers take the performance of the transfer recipients into account?

So far, there is little evidence supporting our H_2. In our argumentation in favour of this hypothesis, we stressed the reciprocal nature of redistribution: those recipients who work hard receive much help. However, the finding that taxpayers pay little attention to the performance of the recipients may indicate that their solidarity is simply an altruistic act of helping those incapable of helping themselves. If solidarity is motivated by unconditional help to those in need, it could explain the absence of changes in the labor supply of taxpayers.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

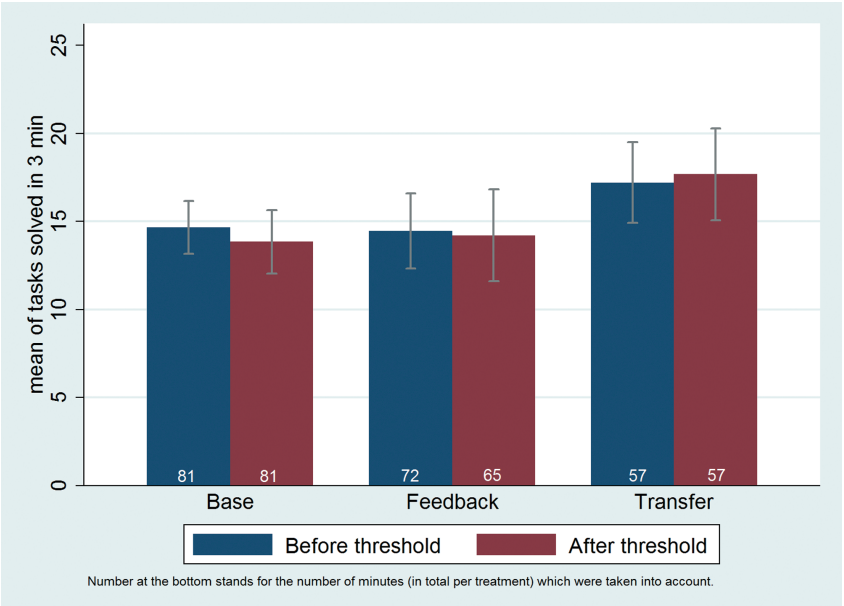


Figure 8: Mean of the number of correctly solved tasks three minutes before and after the threshold, 95 % CI.

Although our experiment does not allow to clearly separate these two logics of social exchange, we will attempt to gather some suggestive evidence in this regard. Fig. 9 shows non-linear relation between relative performance of the taxpayer and the overall performance of recipients in the second phase: if the recipient is far from the threshold, the taxpayer's effort is high, as well as in the case where the recipient is just below it or above. This U-shape can explained by the fact that the logics of reciprocity and solidarity may be complementary.

Very needy transfer recipients trigger taxpayers' solidarity which leads to higher relative performance. In turn, transfer recipients who make it almost on their own deserve support: taxpayers reciprocate and boost their relative performance as well. Players in between do not trigger solidarity, but do not perform well enough to trigger positive reciprocity. Thus, it seems that as soon as the recipient is good enough to be seen as "could have done better", taxpayers' relative performance is the lowest.

The shape of the curve suggests that due to solidarity desperate recipients are supported by taxpayers. It is possible that some players did not

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

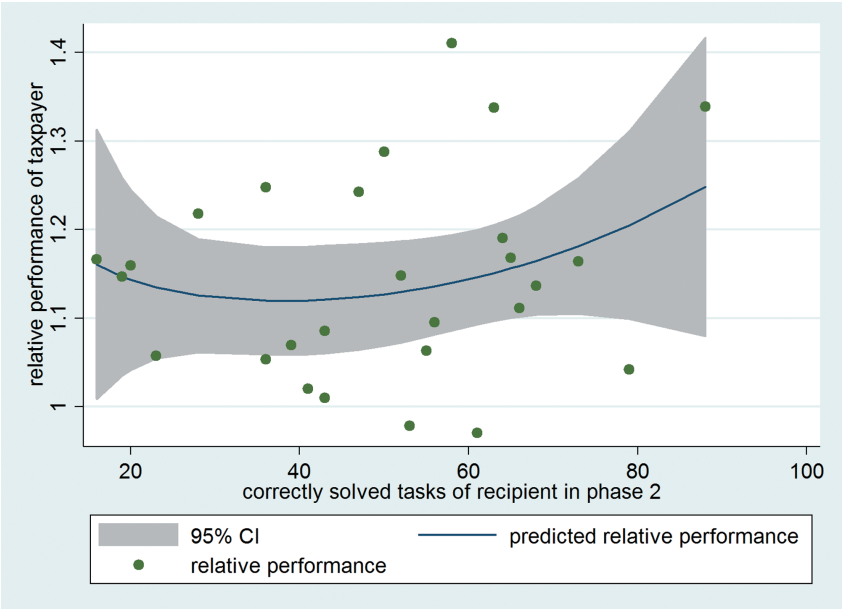


Figure 9: Relative performance of taxpayers depending on the performance of the recipient in Feedback.

reach the threshold at least 3 minutes before the end of the phase; consequently, the average number of tasks completed after the threshold may be affected by the attrition of slower players. We classify players as desperate if the total number of solved tasks and the number of tasks solved in the respective one minute sub-period are below median. Although such rule is rather arbitrary, it allows to single out those who perform poorly in general as well as in the respective sub-period. The estimation confirms that if a transfer recipient is desperate, a taxpayer speeds up by approx. 0.6 tasks per minute. This evidence is, however, suggestive, since we do not elicit the beliefs of taxpayers regarding the effort and ability of the transfer recipients.

Summarizing all of the above:

Result 3: There is some evidence suggesting that both recipients far off the threshold and recipients just below or at the threshold trigger high effort by the taxpayers. For those recipients, transparency increases efficiency in the redistribution system, while recipients between those two

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

extreme cases would benefit from taxpayers' ignorance about their work speed.

E. Conclusion

Large-scale redistribution is often criticized or even opposed because of the common fear of taxpayers being exploited by transfer recipients. In this paper we deal with the widely expressed public opinion that transparency of recipients' efforts may be one mechanism to promote the acceptance of redistribution and its ability to create incentives for transfer recipients to provide extra effort. To study the effects of transparency, we conduct a series of real-effort experiments in which taxpayers can unilaterally observe transfer recipients' level of needs and level of efforts and thus adjust their performance according to the received information.

Transparency is expected to increase productivity by stimulating both the work speed of taxpayers and transfer recipients. Our results do not support this expectation. We find that monitoring does not enhance the performance of the taxpayers but negatively affects the productivity of the transfer recipients. It seems that recipients do not use their performance strategically to trigger positive reciprocity among taxpayers. In turn, only very poor and very good performing recipients stimulate higher effort by taxpayers. Thus, the effect of monitoring on the overall productivity is negative.

Therefore, our results indicate an institutional advantage of deliberate ignorance for redistribution. Rather than stimulating effort provision by taxpayers and transfer recipients, transparency creates a negative reference point for many taxpayers. Overall, transparency backfires and sealing off information increases efficiency. This result demonstrates the importance of Hertwig and Engel's approach: strategic ignorance, that is, deliberately choosing not to know or institutionally choosing not to reveal information, yields sometimes superior outcomes. Hence, in some circumstances, cautiously using strategic ignorance is in the best interest of all members of the society.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

References

Aarøe, Lene and Michael Bang Petersen (2014). "Crowding out culture: Scan- dinavians and Americans agree on social welfare in the face of deservingness cues". In: *The Journal of Politics* 76.3, pp. 684–697.

Agranov, Marina and Thomas R Palfrey (2015). "Equilibrium tax rates and income redistribution: A laboratory study". In: *Journal of Public Economics* 130, pp. 45–58.

Andries, Marianne and Valentin Haddad (2020). "Information aversion". In: *Journal of Political Economy* 128.5, pp. 1901–1939.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2005). "Social preferences and the response to incentives: Evidence from personnel data". In: *The Quarterly Journal of Economics* 120.3, pp. 917–962.

Bartling, Björn, Roberto A Weber, and Lan Yao (2015). "Do markets erode social responsibility?" In: *The Quarterly Journal of Economics* 130.1, pp. 219–266.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg registration and organization online tool". In: *European Economic Review* 71, pp. 117–120.

Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden (2013). "Just luck: An experimental study of risk-taking and fairness". In: *The American Economic Review* 103.4, pp. 1398–1413.

Cherry, Todd L, Peter Frykblom, and Jason F Shogren (2002). "Hardnose the dictator". In: *The American Economic Review* 92.4, pp. 1218–1221.

Chugunova, M., Nicklisch, A., & Schnapp, K.- U. (2022). Redistribution and production with a subsistence income constraint: a real-effort experiment. FinanzArchiv, 78(1/2), 208–238.

De Bruin, Kiki, Yael De Haan, Rens Vliegenthart, Sanne Kruikemeier, and Mark Boukes (2021). "News avoidance during the COVID-19 crisis: Understanding information overload". In: *Digital Journalism* 9.9, pp. 1286– 1302.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr (2001). "Are people conditionally cooperative? Evidence from a public goods experiment". In: *Economics Letters* 71.3, pp. 397–404.

Hertwig, Ralph and Christoph Engel (2016). "Homo ignorans: Deliberately choosing not to know". In: *Perspectives on Psychological Science* 11.3, pp. 359–372.

Kessler, Judd B and Michael I Norton (2016). "Tax aversion in labor supply". In: *Journal of Economic Behavior & Organization* 124, pp. 15–28.

Khadjavi, Menusch, Andreas Lange, and Andreas Nicklisch (2017). "How transparency may corrupt – experimental evidence from asymmetric public goods games". In: *Journal of Economic Behavior and Organization*.

Kolstad, Ivar and Arne Wiig (2009). "Is transparency the key to reducing corruption in resource-rich countries?" In: *World Development* 37.3, pp. 521– 532.

Kool, Wouter, Joseph T McGuire, Zev B Rosen, and Matthew M Botvinick (2010). "Decision making and the avoidance of cognitive demand." In: *Journal of Experimental Psychology: General* 139.4, p. 665.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Kozyreva, Anastasia, Stephan Lewandowsky, and Ralph Hertwig (2020). "Citizens versus the internet: Confronting digital challenges with cognitive tools". In: *Psychological Science in the Public Interest* 21.3, pp. 103– 156.

Lindstedt, Catharina and Daniel Naurin (2010). "Transparency is not enough: Making transparency effective in reducing corruption". In: *International Political Science Review* 31.3, pp. 301–322.

Mas, Alexandre and Enrico Moretti (2009). "Peers at work". In: *The American Economic Review* 99.1, pp. 112–145.

Ogawa, Kazuhito, Toru Takemoto, Hiromasa Takahashi, and Akihiro Suzuki (2012). "Income earning opportunity and work performance affect donating behavior: Evidence from dictator game experiments". In: *The Journal of Socio-Economics* 41.6, pp. 816–826.

Sicherman, Nachum, George Loewenstein, Duane J Seppi, and Stephen P Utkus (2016). "Financial attention". In: *The Review of Financial Studies* 29.4, pp. 863–897.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The Engel Effect?
Experimental Evidence on Rule Compliance

Sebastian J. Goerg, Sebastian Kube & Michael Kurschilgen

A. Introduction

In Prof. Engel's research group at the Max Planck Institute for Research on Collective Goods, each Monday used to feature the "Ratio Runde." During its sessions, members and guests would present their work and receive valuable feedback in a very constructive environment. Given the interdisciplinary nature of his group, which consisted of researchers from many disciplines – including Law, Economics, Psychology, Sociology, Political Science, and Computer Science – the speakers always faced a great number of questions. At some point, Professor Engel introduced a rule that only clarifying questions were allowed during the main part of the presentations. While he always took great care to ensure the audience was aware of this rule, there were no sanctions attached to breaking it, and yet most participants adhered to it. In this paper, we explore whether the high rates of rule adherence can be attributed to Professor Engel's authority as the rule-setter, or if the adherence was merely due to the guideline being explicitly called a "rule".

Already Prof. Engel himself has conducted research on rule-following behavior. His work on rule following shows that compliance is not only a function of intrinsic morality or deterrence: i) compliance is socially mediated: rules succeed or fail through the expectations they induce, ii) information design is crucial, since what authorities publicize about typical behavior can either crowd in or crowd out compliance, iii) a rule taking the form of a law works best in concert with credible sanctions. In the following, we briefly summarize his corresponding work on these topics – some of which is jointly with us – and for which we are very grateful to have had the chance to collaborate with him.

In Desmet & Engel (2021), he demonstrates that individuals are conditional rule followers: their willingness to obey an arbitrary rule increases when they expect higher compliance by peers. This pattern establishes

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a social-contagion channel for compliance that operates even when the rule itself has no instrumental payoff, highlighting why public messaging about typical behavior can shift adherence. In Engel (2023) he pushes this insight toward deviance. In a laboratory experiment, complemented by vignette evidence, he shows that very lean cues about others' noncompliance are sufficient to increase the propensity and intensity of rule violations. The study isolates a peer-effect mechanism for "conditional rule violation," underscoring the risk that transparency about low compliance or visible nonenforcement may propagate offending. Interpreting the rule violation as a crime, as Prof. Engel does in that paper, a straightforward implication for legal design arises: information environments that give rise to a norm of noncompliance can undo the gains from sanctions or moral suasion. Furthermore, observing rule violations reduces compliance not only in the same task but also in others, especially when rules are perceived as similar between tasks (Goerg et al., 2024a). They also show that spillovers – even though they are bounded – are stronger when individuals first break the same rule themselves, suggesting a "slippery slope" effect. Taken together, (not only) these findings from lab and field settings point to conditional norm compliance as a key mechanism: individuals weigh norms less in their decision process after seeing others violate them.

In a related strand of research, Prof. Engel has zoomed in on the importance of norms and the mechanisms by which they guide behavior. Engel, Kube & Kurschilgen (2021) show – by experimentally inducing differences in what people see as "normal" behavior – that selected information about past cooperation shifts beliefs and, through them, cooperation and punishment in social dilemmas. Notably, "bad news" about low cooperation depresses both contributions and the effectiveness of peer punishment, making recovery from poor expectations difficult. This identifies expectations management as a powerful yet delicate policy tool: highlighting undesirable behavior can entrench it. Norms and sanctions are also explored in Engel & Kurschilgen (2013), where the relationship between institutions, expectations, and behavior is made explicit. In public-goods experiments with sanction regimes and legal framing, behavior and normative (i.e. injunctive) expectations coevolve. A "legal frame" by itself does little; paired with sanctions, however, it boosts their effectiveness. Apparently, the law's expressive function complements deterrence rather than substituting for it. A light-touch alternative to sanctions is proposed in Engel & Kurschilgen (2020), namely self-set contribution standards

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

that operate like private injunctive norms. These "self-nudges" initially raise cooperation, but small wording changes or already-high baselines can erase the effect. Their lesson for policymakers is cautionary: norm-based nudges are fragile; what works in one environment may not work in another.

Apart from Prof. Engel's work on rule-following, there has been a growing body of experimental research lately that investigates individuals' willingness to follow rules – particularly in situations where compliance is costly, the rule is seemingly arbitrary, and explicit enforcement is absent (Kimbrough and Wilson, 2022). A first approach was introduced by Kimbrough and Vostroknutov (2016), who developed a "costly rule-following" task. In their design, participants decide whether to wait at a series of red traffic lights before proceeding, with the rule being to wait until each light turns green. While waiting reduces monetary payoffs, the presence of the rule substantially increases compliance: 62.5 % of participants always waited in the rule condition, compared to 12.5 % in the control where no such rule was given to the subjects. This task has initiated some further methods to elicit intrinsic motivation to follow arbitrary rules. For example, to enhance portability and scalability, Kimbrough and Vostroknutov (2018) proposed a second task in which participants allocate a series of balls into one of two buckets. The rule instructs them to place the balls in the lower-paying bucket, while placing them in the higher-paying bucket yields greater financial rewards. This design allows researchers to measure individual variation in rule adherence by observing willingness to incur financial costs to follow an unenforced rule. Again, the rule is presented in a neutral manner and no punishment is associated with not adhering to the rule – and again, rule-following is frequently observed across several subject pools from different countries. Moreover, as shown in Kimbrough et al. (2024), rule-following rates in this task are remarkably stable at the individual level, at least over a one-month period.

Building on the results of these abstract rule-following paradigms, Gächter et al. (2025) introduced and experimentally tested the CRISP framework, which conceptualizes rule conformity (C) as a function of four motivational forces: intrinsic respect for rules (R), extrinsic incentives (I), social expectations (S) and social preferences (P). Their experimental evidence, which uses one-shot variants of the traffic light rule-following tasks, highlights how social context shapes compliance: observing others break rules reduces one's own compliance, while introducing strong probabilistic financial penalties increases it significantly. Furthermore, tying

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

rule-following to altruistic preferences by not donating 1 $ to the red cross if the rule is violated significantly increases rule-following rates. Their design also incorporates the elicitation of empirical (descriptive), normative and personal beliefs, which are shown to be important predictors of rule-following behavior.

While these prior studies primarily investigate responses to arbitrary rules set by the experimenter, recent work explores how the perceived legitimacy and identity of the rule-setter influence compliance. In particular, in politically polarized contexts, such as during the 2024 U.S. election cycle, Suri et al. (2025a) test if individuals respond differently to the same rule depending on whether it is attributed to a political ingroup or outgroup. While rule-following is found to be remarkably robust with respect to ideological polarization, affective polarisation (perceived social distance to the rule setter), seems to pose the stronger behavioural threat for rule-following.

In this paper, we used the task from Suri et al. (2025a) to explore if rule-following changes if the rule is attributed to Prof. Engel (in the following referred to as treatment Engel), compared to treatment Baseline where the rule is given by an anonymous user from the online platform on which we ran our study (Prolific Academic). Subjects in both treatments (n=100 per treatment) are given exactly the same experimental instructions and face the same financial incentives to follow or break the rule. They are shown 20 coins on screen, each being worth 10 Pence to them, and each second one coin disappears and reduces the payoff to the subject correspondingly. Subjects can stop the coins from disappearing whenever they like, but a rule instructs them to do so only after a visual signal on screen has occurred. The only difference between treatments is that this rule is either presented as a rule given by another prolific user or as a rule given by Prof. Engel (see the instructions, which are in the appendix, for the exact wording). After they made their decision, they are additionally asked about their personal norm to follow the rule in this situation, their normative expectations about how others view following or breaking the rule in this situation, and their descriptive belief about how many of the other participants they expect to have followed the rule. Since Prof. Engel has started working on Large Language Models (LLMs) in recent years (Engel and McAdams 2024, Engel 2025, Engel et al. 2025), in addition to our online experiment we also provided the instructions and treatment descriptions to four popular LLMs and asked them to guess the impact of Prof. Engel being the rule setter in this experiment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Our comparison between the LLMs' forecasts and the behavior of human participants reveals a striking divergence. All four LLMs (ChatGPT 5.0, Perplexity, Google Gemini, Claude Sonnet 4) predicted that attributing the rule to a well-known legal scholar would substantially increase rule-following, with expected rule-following rates in the Engel treatment exceeding the baseline by 20 to 30 percentage points. These predictions were mirrored in the models' forecasts of personal, normative, and descriptive beliefs: the LLMs consistently anticipated that participants would view compliance as the modal response, that rule-following would be seen as very socially appropriate, and that others would be expected to follow the rule at higher rates in the Engel treatment than in the baseline. In other words, LLMs projected a strong "Engel Effect", both at the behavioral and at the belief level.

In contrast, the actual experimental data provide little evidence for an Engel effect on people's behavior. Rule-following rates did not significantly differ between treatments. They were fully in line with rates that were previously reported in rule-following studies using the same or related rule-following tasks (see the corresponding literature above). Similarly, participants' own normative judgments about whether one ought to follow the rule were unaffected by authority framing. We do, however, find some shifts in expectations. Participants in the Engel treatment were more likely to believe that others considered compliance very socially appropriate. Regression analysis shows that actual compliance was best explained by personal and descriptive beliefs, not by the authority source of the rule. Taken together, the results suggest that while LLMs readily anticipate authority effects on rule-following, human subjects appear less responsive, with behavior shaped instead by individual norms and expectations of peer behavior.

The remainder of this paper is structured as follows. Section 2 presents the experimental design and introduces the main hypotheses. Section 3 reports the empirical results. Section 4 discusses the findings and concludes with a summary.

B. Design

Our experiment is designed to obtain a clean behavioral measure of rule compliance to explore if the added authority of Prof. Engel introducing a rule affects rule-following behavior. To achieve this, we use the rule-fol-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lowing task (coins-task) used in Suri et al. (2025a) and Suri et al. (2025b). The task is inspired by the work of Kimbrough and Vostroknutov (2016) and Gächter et al. (2025). It is slightly more abstract than the traffic light tasks used in their research, with the key advantage that the cost of complying with the rule is very clear to participants.

The task works as follows: Participants see 20 coins on their screen (see figure 1). Every second, one coin disappears. Participants can press a button at any time to stop the coins from disappearing (in Suri et al., 2025a,b, there was no button, but subjects had to press the space bar). The number of coins they have left when they press the button determines their payout, with each coin worth £0.10. Participants are explicitly told they can press the button at any time, which means they can receive a payout from 0 to 20 coins, i.e. between £0 to £2 (in addition to the £1.00 show-up fee that subjects received for participation).

In addition to the coins, a sign is displayed on the screen. It starts as a "minus" symbol and changes to a "plus" after 12 seconds. The exact time of the change is not disclosed to the participants. The following non-enforced rule is implemented: "Press the stop button only after the sign has changed from 'minus' to 'plus'." Importantly, following or breaking the rule has no monetary consequences beyond the number of coins a participant has left.

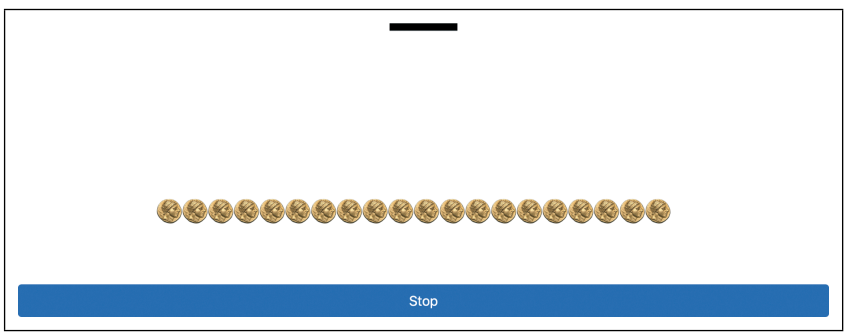


Figure 1: The coins-task

Note: This is a screenshot of the main decision screen. Here, the coins haven't started to vanish and, correspondingly, the sign hasn't changed yet from minus to plus. See the appendix for more screenshots and the full set of instructions.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Before beginning the task, participants are shown a visual representation of the process and must correctly answer two comprehension questions. These questions ask how many coins they would receive if they pressed the space bar at two different time points shown in a screenshot (similar to the image above). This is to implicitly emphasize that there are no financial penalties for breaking the rule, as well as to ensure that each subject has understood the instructions. Each participant plays the coins-task exactly once.

We have two treatments, Baseline and Engel. In treatment Engel, the rule is introduced as follows: *"There is a rule for this. The rule is based on a study by the professor of law, Prof. Dr. Dres. h. c. Christoph Engel. The rule is: Press the stop button only after the sign has changed from minus to plus."* In addition, a picture taken from Prof. Engel's website is displayed on screen (see figure 2). In the baseline treatment, subjects are instead told that the rule was chosen by another user on Prolific, and a generic person icon is shown instead of Prof. Engel's picture (the decisions of other Prolific users were pre-sampled and taken from Suri et al. 2025a).

Procedure: The experiments were programmed in oTree (Chen et al., 2016) by Maren Bermúdez, whom we thank for excellent research assistance. The study was run in August 2025. We sampled German-speaking participants living in Germany from Prolific Academic until we had collected data of 200 subjects that completed the entire experiment, randomly assigning them to either treatment Engel or baseline. In total, we ended up with 99 independent observations in treatment Engel and 101 independent observations in the baseline condition. It took participants about 5 minutes (median) to finish the study. They were paid a base payment of 1.00 GBP for their participation and could earn an additional bonus payment of up to 2.00 GBP, depending on their decision in the coins-task. On average, subjects earned 1.24 GBP in the bonus task. The order of tasks and questions was the same for all subjects: after subjects provided their consent, they entered their Prolific ID and were then provided with the instructions of the coins-task. After they successfully answered a set of control questions about the task at hand, they were once again reminded about the rule setter (other Prolific user or Prof. Engel) and then made their decision in the coins-task. Afterwards, we asked them about their personal norm (personal belief), their normative expectations and their descriptive belief. The study ended with a short socio-demographic questionnaire, asking them for their gender, age, highest school and vocational degrees and whether they had any connection to the le-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

gal domain (e.g., studying or having studied law, or working in a related field).



Figure 2: Treatment manipulation Note: The left-hand side shows the stylized person used in treatment Baseline to introduce the rule given by a Prolific user, while the right-hand side shows the photograph of Prof. Engel used in treatment Engel to introduce the rule. The speech bubble with instructions was present in both treatments.

Note: The left-hand side shows the stylized person used in treatment Baseline to introduce the rule given by a Prolific user, while the right-hand side shows the photograph of Prof. Engel used in treatment Engel to introduce the rule. The speech bubble with instructions was present in both treatments.

To complement our lab experiment with human participants, we also explored how state-of-the-art language models would forecast behavior in the same setting. We gave four LLMs (ChatGPT 5.0, Perplexity, Google Gemini, and Claude Sonnet 4= the identical instructions and treatment manipulations that participants saw. Then, in German since also the experimental instructions were in German, we asked each model to reason about what a typical participant might do and to produce quantitative hypotheses. Specifically, we asked them to estimate: i) how many participants in each treatment would wait for the signal to switch from minus to plus before acting, ii) how responses to each of the three normative belief questions would be distributed across the four answer categories, and iii) what proportion of other subjects the participants would expect to follow

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the rule. This allowed us to directly compare LLM forecasts with actual human behavior.

C. Hypotheses

We build on authority theory and social influence (Milgram, 1974; Cialdini, 2001) to conjecture that the identity of the rule-setter shapes willingness to comply with the rule. Early work emphasized problematic obedience – most famously Milgram's experiments – but subsequent research documents potentially beneficial consequences of legitimate authority for coordination and cooperation (e.g., Posten et al., 2025; Petty & Cacioppo, 1986). In applied domains, messenger credibility and reputational signals amplify compliance: for instance, Attari et al. (2016, 2019) show that the smaller the personal carbon footprint of a climate scientist – and thus arguably the larger their perceived authority and legitimacy – the more effective they are at encouraging energy conservation and mobilizing support for climate-friendly policies. In addition, our authority cue (law professor) can render the situation more "law-like," strengthening the rule's expressive meaning and perceived obligatoriness.

In this paper, we therefore test whether attributing the rule to a renowned law professor – a figure plausibly seen as legitimate in matters of rules and law – elevates compliance relative to a neutral peer source. Formally, we hypothesize that participants in the Engel treatment, compared to baseline, will exhibit higher compliance with the given rule in our experiment.

H1 (Authority-compliance): Rule compliance is higher in treatment Engel than in the baseline condition.

Beyond behavior, we expect the rule-setter's identity to shift both injunctive (normative) and empirical (descriptive) expectations, i.e., beliefs about what is socially appropriate or "ought" to be done and beliefs about what others will do, via two mechanisms.

Normative-cue channel: First, a salient, legitimate rule setter (e.g., a respected authority, an elected leader) strengthens injunctive/normative cues about appropriate conduct. Classic work on social norms shows that making injunctive norms salient does guide behavior even holding incentives fixed (Bicchieri, 2005). In law and policy contexts, perceived legitimacy robustly predicts compliance beyond deterrence, and democratic or otherwise "owned" rules tend to elicit higher cooperation than exoge-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

nously imposed ones – consistent with stronger normative expectations under a rule setter seen as rightful (see, for instance, Cialdini, 1991, Tyler, 2006a, 2006b; Sunshine and Tyler, 2003). The messenger's identity can heighten that salience. Perceived authority and legitimacy increase participants' sense of obligation and their anticipated disapproval for deviance. Legal rules also carry an expressive signal that can complement deterrence and coordinate expectations (Silverman et al., 2014; McAdams, 2015; Posner, 2000). Relatedly, "dynamic norm" messages about what people approve of (or are moving toward) can increase compliance (Sparkman & Walton, 2017; Tankard & Paluck, 2016). As a result, experimental participants might view rule violations as more socially inappropriate – and evaluate rule-following as more socially appropriate – when the rule is attributed to Prof. Engel. Authoritative attribution can also reduce strategic uncertainty and make the compliant action ("wait until sign changes to plus") focal, easing coordination on the normatively appropriate choice (Engel et al. 2021). Such attribution may activate deontic self-concepts – "I am the kind of person who respects (official) rules or professional expertise" – thereby strengthening personal norms (Tyler, 2006a, 2006b; Hoeft et al., 2025). Even with no explicit sanctions, naming a legal authority can raise perceived scrutiny or reputational costs, modestly increasing expected enforceability (Silverman et al., 2014; French & Raven, 1959).

H2 (Authority-injunctive beliefs): Personal and normative beliefs are more favorable to rule-following in treatment Engel than in the baseline condition.

Coordination-beliefs channel: The rule setter's identity also acts as an informational cue about what others will do, following a coordination logic central to social-norms theory. If a rule appears backed by a highly legitimate source, individuals infer higher compliance by peers and condition their own behavior accordingly (Bicchieri & Xiao, 2009; Krupka & Weber, 2013; Desmet & Engel, 2021). In other words, authority can shift descriptive expectations that are known to be pivotal for rule adherence (Fehr & Schurtenberger, 2018). As individual subjects respond to authority, comply more with the rule, and evaluate rule compliance as more socially appropriate, this translates into their beliefs about other subjects' behavior and therefore they might hold higher expectations of compliance by others.

H3 (Authority-descriptive beliefs): Descriptive beliefs about others' rule-following behavior are higher in treatment Engel than in the baseline condition.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

D. Results LLMs

In this section, we examine the responses of commonly used LLMs (ChatGPT 5.0, Perplexity, Google Gemini, and Claude Sonnet 4). Each model was provided with the instructions of the experiment described above and given an identical prompt to predict participants' behavior. Specifically, all LLMs were asked to forecast both the decision outcomes and the belief responses of participants. Table 1 summarizes the predicted responses on decisions and beliefs.

Table 1: LLM Responses with expected results for rule following and beliefs.

		Chat GPT 5.0	**Perplexity**	**Google Gemini**	**Claude Sonnet 4**
Rule Following	**Engel**	60–70 %	65–75 %	70–80 %	75–80 %
	Baseline	30–40 %	40–50 %	50–60 %	45–55 %
Descriptive Belief	**Engel**	60 %	70 %	60–70 %	65–70 %
	Baseline	40 %	45 %	40–50 %	40–45 %
Personal Belief	**Engel**	5/10/35/50	5/10/35/50	1/9/40/50	5/15/30/30
	Baseline	10/20/40/30	15/25/35/25	5/25/55/15	15/25/40/20
Normative Belief Follow	**Engel**	0/5/30/65	2/8/30/60	0/1/24/75	5/10/45/40
	Baseline	0/10/40/50	5/20/35/40	1/4/80/15	10/20/50/20
Normative Belief Break	**Engel**	50/35/10/5	55/30/12/3	25/60/14/1	35/40/20/5
	Baseline	30/35/25/10	30/30/25/15	5/45/45/5	20/35/35/10

Notes: Personal Belief: percentages for Never / Rarely / Often / Always the rule

Normative Beliefs: percentages for Very inappropriate / Somewhat inappropriate / Somewhat appropriate / Very appropriate to either follow the rule (Normative Belief Follow) or break the rule (Normative Belief Break)

All LLMs predict higher levels of rule-following behavior in the Engel treatment compared to the baseline condition. These behavioral predictions are consistent with their respective forecasts of descriptive beliefs. As such, the predictions of the LLMs are in line with the formulated hypotheses for the experiment.

Although there are differences in the distributions of predicted beliefs across models, the overall pattern is consistent. For the Engel treat-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ment, all LLMs anticipate that the modal response to the question of personal rule adherence is to always follow the rule (with Claude showing a tie between always and mostly). In the baseline condition, all models predict that the modal response is to mostly follow the rule.

Regarding the perceived social appropriateness of rule-following behavior, all models predict that, in the Engel treatment, following the rule is considered very socially appropriate, while not following the rule is viewed as very socially inappropriate.

In the baseline treatment, predictions vary more substantially across models. ChatGPT 5.0 and Perplexity predict very socially appropriate as the modal response to rule-following, while Google Gemini and Claude Sonnet predict socially appropriate. For the appropriateness of not following the rule in the Engel treatment, ChatGPT 5.0 and Perplexity predict a modal response of very socially inappropriate, whereas Google Gemini and Claude Sonnet anticipate socially inappropriate. In the baseline condition, all models predict socially inappropriate as the modal response; however, Perplexity shows a tie between socially inappropriate and very socially inappropriate, while Google Gemini and Claude Sonnet predict a tie between socially inappropriate and socially appropriate.

E. Results Experiment

We first examine the effect of treatment on actual rule-following behavior. In the baseline condition, participants followed the rule in 59.41 % of trials, whereas in the Engel treatment – the condition in which the rule was stated by a law professor – the rate was slightly lower at 54.55 % (see Figure 3). This difference is statistically insignificant (χ^2, $p = 0.488$), indicating that the authority framing of the rule did not have a direct effect on participants' compliance in this task.

Next, we consider participants' personal beliefs regarding rule compliance (see figure 4). Participants were asked: *"What do you think people should do in decision-making situations like this?"*, with response options ranging from always to never following the rule. In baseline, 54.46 % of participants indicated that one should always or often follow the rule, compared to 52.53 % in the Engel treatment. 23.23 % indicated one should rarely follow the rule in baseline, compared to 15.84 % in the Engel treatment, and 24.24 %, resp. 29.7 %, said one should never follow the rule in decision making situations like this. These differences are not statistically

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

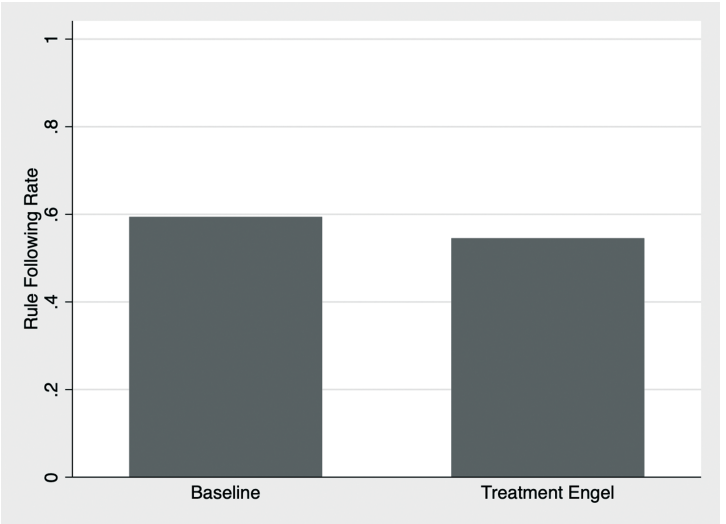


Figure 3: Relative frequency rule following behavior in treatments Baseline and Engel

significant (χ^2, p = 0.784), suggesting that framing the rule as coming from a law professor did not meaningfully influence participants' own normative judgments about rule-following.

We then turn to participants' normative expectations about how others would evaluate rule-following behavior. Specifically, participants were asked: *"Suppose a Prolific user had followed or violated the rule. What do you think, how would most other Prolific users evaluate this behavior?"* For expectations regarding compliance, we observe a significant effect of the Engel treatment. In this condition, 58.59% (40.4%) of participants expected that others would consider following the rule very (somewhat) socially appropriate, compared to 41.58% (51.49%) in baseline. Only very few subjects considered following the rule as inappropriate (somewhat/very inappropriate in treatment Engel: 1.01%/0%, in baseline 4.95%/1.98%). These differences are statistically significant (χ^2, p = 0.032), suggesting that there is an association between treatment Engel and the normative expectations. In contrast, expectations regarding rule violation were similar across treatments: in both conditions, roughly 80% of participants indicated that breaking the rule would be viewed by others as somewhat or very inappropriate, 15.15%/18.81% expected that

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

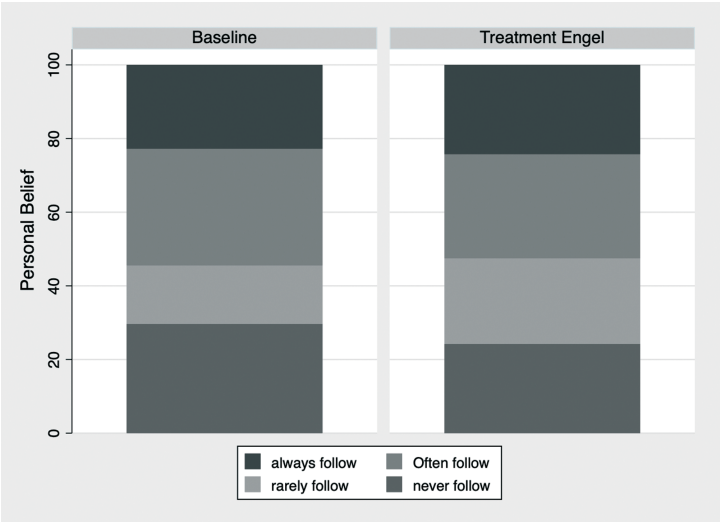


Figure 4: Personal Belief on rule following

others would view violating the rule as somewhat appropriate in treatment Engel/baseline, and only 4.04%/0.99% answered very appropriate (χ^2, p = 0.480). Taken together, these results indicate that authority framing increases the perceived social approval of rule compliance but does not affect perceptions of disapproval for rule-breaking.

Participants were also asked to report their descriptive beliefs about others' behavior, specifically the proportion of other participants in the same situation who followed the rule. The mean descriptive belief was 55.4% in baseline and 60.9% in the Engel treatment. While this suggests a positive tendency associated with authority framing, the difference is not statistically significant (Wilcoxon Rank-Sum test, p = 0.1785).

To examine the determinants of individual rule-following behavior more formally, we estimate a series of linear probability models, reported in Table 2. Model 1 includes only a dummy variable for the Engel treatment and shows a small, negative, and insignificant effect (β = -0.049, SE = 0.070). In Model 2, we add participants' personal beliefs about whether one should follow the rule. Personal belief is strongly and significantly associated with compliance (β = 0.263, SE = 0.025, p < 0.01), while the Engel treatment remains insignificant (β = -0.062, SE = 0.057).

Model 3 further incorporates normative beliefs regarding whether

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

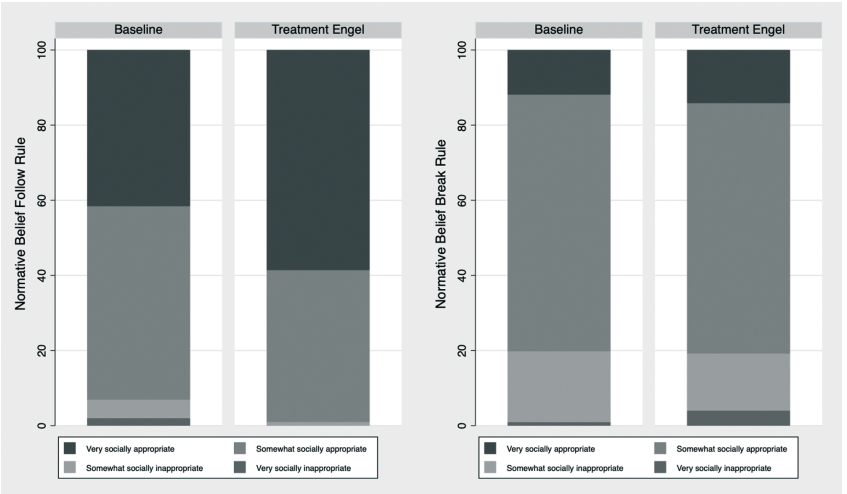


Figure 5: Belief on normative ratings on rule compliance and rule violation among other subjects.

others will follow or violate the rule. Here, the coefficient on expectations that others will follow the rule is positive and significant ($\beta = 0.120$, SE = 0.051, $p < 0.05$), whereas beliefs about others violating the rule are not significant ($\beta = 0.012$, SE = 0.049, $p = 0.80$). The Engel treatment coefficient is slightly larger in absolute value but still not significant ($\beta = -0.091$, SE = 0.057).

Finally, Model 4 adds descriptive beliefs, that is, participants' expectations about the proportion of others who actually followed the rule. Descriptive belief is a strong positive predictor of individual compliance ($\beta = 0.004$, SE = 0.001, $p < 0.01$), and personal and normative beliefs remain significant. In this full specification, the Engel treatment coefficient is marginally negative and weakly significant ($\beta = -0.107$, SE = 0.056, $p < 0.10$), suggesting that, once beliefs are controlled for, framing the rule as coming from a law professor does not increase compliance.

Overall, our results indicate that rule-following behavior is primarily driven by participants' personal beliefs and their expectations of others' behavior and social approval. Authority framing has limited direct impact on compliance, although it does appear to increase participants' perceptions that others would socially approve of following the rule.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 2: Regressions explaining rule-following behavior

	(1)	(2)	(3)	(4)
Treatment Engel	–0.049	-0.062	-0.091	–0.107*
	(0.070)	(0.057)	(0.057)	(0.056)
Personal Belief		0.263***	0.248***	0.205***
		(0.025)	(0.026)	(0.029)
Normative Belief others will follow			0.120**	0.106**
			(0.051)	(0.050)
Normative Belief others will violate			0.012	–0.014
			(0.049)	(0.048)
Descriptive Belief				0.004***
				(0.001)
Constant	0.594***	-0.056	-0.454**	–0.442**
	(0.049)	(0.074)	(0.181)	(0.177)
Observations	200	200	200	200

Note: The table presents the results of linear probability models (OLS regressions). The dependent variable is a dummy variable indicating whether an individual subject has followed the rule (=1) or violated the rule (=0). The independent variables are Treatment Engel, a dummy variable which equals 1 if the rule was given by Prof. Engel, the categorical variables "Personal Belief" (1=never, 2=rarely, 3=often, 4=always follow the rule), "Normative beliefs others will follow" (1=very inappropriate, 2=somewhat inappropriate, 3=somewhat appropriate, 4=very appropriate to follow the rule) and "Normative beliefs others will violate" (1=very appropriate, 2=somewhat appropriate, 3=somewhat inappropriate, 4=very inappropriate to violate the rule), and the continuous variable "Descriptive Belief" representing the relative frequency that the subject expects others to follow the rule, measured as a percentage ranging from 0 to 100. Standard errors in parentheses. Levels of significance are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

F. Summary

In this study, we combined experimental economics with forecasts from generative AI to examine whether attributing a rule to a prominent legal scholar influences compliance. Across treatments, we observed robust levels of rule-following, broadly consistent with prior work using similar paradigms. However, we found no evidence for an Engel effect in the narrow sense of increased compliance when the rule was attributed to Prof. Engel. While participants in the Engel treatment were more likely to believe that others considered compliance socially appropriate, this shift in expectations did not translate into higher rates of actual rule-following behavior. Instead, our results confirm that compliance is primarily driven by individual norms and expectations of peer behavior.

The absence of a measurable Engel effect in this study does not imply that authority framing or scholarly attribution is irrelevant. Rather, it suggests that the effect may depend on contextual factors that were not present in our design. Framing and norm-based nudges are fragile; what works in one environment may not work in another. This mirrors findings on information-provision nudges in less formal norm contexts, such as gender norms or support for specific policies. In these settings, individuals often underestimate the extent of support for certain norms and adjust their behavior when provided with information (e.g., Bursztyn et al., 2020; Goerg et al., 2024b) – though in some cases they do not (Goerg et al., 2024b), depending on the environment and the source of the information, again pointing to the fragility of norm-based nudges.

One possibility for the non-result in this study is that the domain matters: our rule was arbitrary and payoff-irrelevant, whereas rules that are more meaningful or explicitly connected to legal norms might elicit stronger responses. Another factor may be social distance. As Suri et al. (2025a) show, perceived distance to the rule-setter is predictive of compliance, and in our sample only very few participants reported studying or working in fields related to law. For many, a law professor may thus appear socially distant, weakening any authority effect. Relatedly, our framing was deliberately cautious in order to avoid deception: participants were told that the rule was "based on a study by the professor of law, Prof. Dr. Dres. h. c. Christoph Engel," rather than that it was "given by" him. A stronger attribution may have produced different results. Future studies might therefore target subject pools with a closer connection to the legal domain, vary the wording of authority attributions, or directly frame the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

rule as a "law" rather than a neutral "rule." Introducing sanctions could also provide a way of making the authority of the rule more salient. These extensions would help clarify the scope conditions under which an Engel effect emerges, and whether it depends on the interplay of domain relevance, wording, and social distance.

Finally, while our data did not reveal an Engel effect in rule-following, our study itself reflects Engel's legacy. The combination of behavioral law and economics, social norms, and experimental methodology – enriched here with the use of generative AI – builds directly on themes that he advanced and shaped. For us, the Engel effect has certainly existed: in his role as a researcher, mentor, and innovator who encouraged us to ask new questions and to embrace new tools. We are deeply grateful for his inspiration and support, and we wish him many more years full of creativity and enthusiasm for continuing to advance our academic knowledge frontier.

References

Attari, S. Z., Krantz, D. H., & Weber, E. U. (2016). "Statements about climate researchers' carbon footprints affect their credibility and the impact of their advice". Climatic Change, 138(1), 325–338.

Attari, S. Z., Krantz, D. H., & Weber, E. U. (2019). "Climate change communicators' carbon footprints affect their audience's policy support". Climatic Change, 154(3), 529–545.

Bicchieri, C. (2005). The grammar of society: The nature and dynamics of social norms. Cambridge University Press.

Bicchieri, C., & Xiao,E. (2009). "Do the right thing: but only if others do so". Journal of Behavioral Decision Making, 22(2), 191–208.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020) "Misperceived social norms: Women working outside the home in Saudi Arabia". American Economic Review 110, 2997–3029.

Chen, D. L., Schonger, M., & Wickens, C. (2016). "oTree – An open-source platform for laboratory, online, and field experiments". Journal of Behavioral and Experimental Finance, 9, 88–97.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). "A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior". In Advances in experimental social psychology (Vol. 24, pp. 201–234). Academic Press.

Cialdini, R. B. (2001). Influence: science and practice (4th ed.). Boston, MA: Allyn and Bacon. ISBN 0-321-01147-3. OCLC 43607370

Desmet, P. T., & Engel, C. (2021). "People are conditional rule followers". Journal of Economic Psychology, 85, 102384.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Engel, C. (2023). "How little does it take to trigger a peer effect? An experiment on crime as conditional rule violation". Journal of Research in Crime and Delinquency, 60(4), 455–492.

Engel, C. (2025). "The negotiation trap: An experiment on a large language model". Max Planck Institute for Research on Collective Goods, Discussion Paper 2025/8.

Engel, C., Hermstrüwer, Y., & Kim, A. (2025). "Human realignment: An empirical study of LLMs as legal decision-aids in moral dilemmas". Max Planck Institute for Research on Collective Goods, Discussion Paper 2025/3.

Engel, C., Kube, S., & Kurschilgen, M. (2021). "Managing expectations: How selective information affects cooperation and punishment in social dilemma games". Journal of Economic Behavior & Organization, 187, 111–136.

Engel, C., & Kurschilgen, M. (2013). "The coevolution of behavior and normative expectations: An experiment". American law and economics review, 15(2), 578–609.

Engel, C., & Kurschilgen, M. (2020). "The fragility of a nudge: the power of self-set norms to contain a social dilemma". Journal of Economic Psychology, 81, 102293.

Engel, C., & McAdams, R. H. (2024). "Asking GPT for the ordinary meaning of statutory terms". University of Illinois Journal of Law, Technology and Policy, 235.

Fehr, E., & Schurtenberger, I. (2018). "Normative foundations of human cooperation". Nature Human Behaviour, 2, 458–468.

French, J. R. P., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), Studies in Social Power (pp. 150–167). Ann Arbor, MI: University of Michigan Press.

Gächter, S., Molleman, L. & Nosenzo, D. (2025) "Why people follow rules". Nature Human Behaviour 9, 1342–1354.

Goerg, S. J., Himmler, O., & König, T. (2024a). "Norm violations and behavioral spillovers – Evidence from the lab and the field". European Economic Review, 166, 104776.

Goerg, S.J., Pondorfer, A., & Stöhr, V. (2024b). "Regional variation in social norm nudges". Scientific Reports, 14(1), p. 16773.

Hoeft, L., Kurschilgen, M., & Mill, W. (2025). "Norms as obligations". International Review of Law and Economics, 81, 106235.

Kimbrough, E. O., Krupka, E. L., Kumar, R., Murray, J. M., Ramalingam, A., Sánchez-Franco, S., Sarmiento, O. L., Kee, F., & Hunter, R. F. (2024). "On the stability of norms and norm-following propensity: a cross-cultural panel study with adolescents". Experimental Economics, 27(2), 351–378.

Kimbrough, E. O., & Vostroknutov, A. (2016). "Norms Make Preferences Social". Journal of the European Economic Association, 14(3), 608–638.

Kimbrough, E. O., & Vostroknutov, A. (2018). "A portable method of eliciting respect for social norms". Economics Letters, 168, 147–150.

Kimbrough, E. O., & Wilson, B. J. (2022). Rule-following. In C. M. Melenovsky (Ed.), The Routledge Handbook of Philosophy, Politics, and Economics (pp. 108–120). Routledge.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. Journal of the European Economic Association, 11(3), 495–524.

McAdams, R. H. (2015). The Expressive Powers of Law: Theories and Limits. Cambridge, MA: Harvard University Press.

Milgram, S. (1974), Obedience to Authority: An Experimental View, London: Tavistock Publications.

Petty, R. E., & Cacioppo, J. T. (1986). Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer.

Posner, E. A. (2000). Law and Social Norms. Cambridge, MA: Harvard University Press.

Posten, A.-C., Uğurlar, P., Kube, S., & Lammers, J. (2025). "Maintaining Cooperation through Vertical Communication of Trust When Removing Sanctions". Proceedings of the National Academy of Sciences 122 (12): e2415010122.

Silverman, D., Slemrod, J., & Uler, N. (2014). "Distinguishing the role of authority "in" and authority "to"". Journal of Public Economics, 113, 32–42.

Sparkman, G., & Walton, G. M. (2017). "Dynamic norms promote sustainable behavior, even if it is counter-normative". Psychological Science, 28(11), 1663–1674.

Sunshine, J., & Tyler, T. R. (2003). "The role of procedural justice and legitimacy in shaping public support for the police". Law & Society Review, 37(3), 513–548.

Suri, D., Kube, S., & Schultz, J. (2025a). "Evaluating Online Data Collection Platforms Using A Simple Rule-Following Task". Economics Letters 255 (September): 112509.

Suri, D., Gächter, S., Kube, S., & Schultz, J. (2025b). "The resilience of rule compliance in a polarised society". mimeo.

Tankard, M. E., & Paluck, E. L. (2016). "Norm perception as a vehicle for social change". Social Issues and Policy Review, 10(1), 181–213.

Tyler, T. R. (2006a). "Psychological perspectives on legitimacy and legitimation". Annual Review of Psychology, 57(1), 375–400.

Tyler, T. R. (2006b). Why people obey the law. Princeton university press.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The (Im)Moral Man?
An Experimental Study on Hypocrisy and Gender in Dictator-Game Giving

Heike Hennig-Schmidt & Gari Walkowitz

A. Introduction

The dictator game (DG)–due to its very simple design–is a powerful tool to systematically vary the conditions that affect altruistic behavior and explore behavioral heterogeneity (Engel, 2011). In the standard version of the game, the dictator is provided with an amount of money (the pie) she can distribute between herself and the receiver, who does not make any decision and cannot react to the dictator's proposal (Forsythe et al., 1994). Roth (1995) notes that the dictator's allocation "… can be interpreted as a pure expression of her preferences" (p. 270). Totally selfish dictators are expected to keep the whole pie. Yet, Forsythe et al. (1994) and later behavioral studies report the existence of other-regarding dictators allocating shares of about 20 % to 30 % to receivers in incentivized standard versions of the DG (e.g., Camerer, 2003; Engel, 2011). The literature following Forsythe et al. (1994) altered experimental conditions to understand when and why altruistic behavior is observed by analyzing, e.g., repetition, stakes, anonymity, proximity, or gender effects (e.g., Camerer 2003; Engel 2011). Having conducted–to the best of our knowledge–the first meta-analysis of dictator game experiments, Engel (2011) concludes: "Dictator games are helpful for exploring human sociality, because the situation can be tightly controlled, because the simple game can be played with all classes of participants and, of course, because decisions are incentivized." (p. 608).

In recent years, Roth's (1995) notion that dictators' seemingly other-regarding behaviors express true altruistic preferences was questioned, for instance, when a wiggle room exists (e.g., Dana et al., 2007; Fischbacher & Föllmi-Heusi; 2013; Abeler et al., 2014; Cerrone & Engel, 2019), when people have image concerns (Andreoni & Bernheim 2009; Lönnqvist et al., 2014), or when they have normative expectations believing that others

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

expect them to conform to a given norm, although they would like to behave differently (e.g., Bicchieri, 2005; Bicchieri & Xiao, 2009; Bicchieri, 2016).

This study focuses on a situation in which there is a wiggle room for the dictator. In our experiment, the dictator has the option of flipping a fair coin to determine how the pie is divided, with the result of the coin flip being private information known only to the dictator and unverifiable by the receiver and the experimenter. These types of (seemingly) fair procedures, in which, on the one hand, the dictator has control over the outcome of the coin flip and, on the other hand, asymmetric information exists about how the ostensibly fair procedure was actually carried out, are often used in practice. For example, consider social auditing companies whose activities cannot be monitored from the outside. These companies typically act on competitive markets and depend on the auditee they are supposed to monitor to give the assignment to do so. The auditor therefore has an incentive to achieve an audit result that does not conflict with the audited company's objectives. Other examples include influencing and selecting expert commissions, or arranging seemingly fair elections.

Lönnqvist et al. (2014) and Apffelstaedt et al. (2025) analyze this topic in a specific variant of the DG where the dictator can directly choose either an equal split of the pie, a payoff in favor of herself, or use the coin to determine the allocation (Batson et al., 1997; Bucciol & Piovesan, 2011). If dictators use the coin honestly, the outcomes should reflect the 50%-probability of payoffs favorable to themselves. Lönnqvist et al. (2014) and Apffelstaedt et al. (2025) report that most dictators take advantage of the wiggle-room option as much more than half of the reported coin-flip outcomes favor the dictator. This result suggests that many dictators are not primarily motivated to *be* altruistic or fair (also report an equal-split coin-flip outcome or choose the fifty / fifty split directly) but to *appear* altruistic (choose the coin but report only favorable outcomes).

To raise non-selfish and honest behavior, Lönnqvist & Walkowitz (2019) in a similarly structured DG employed perspective taking (Batson, 1991; 2011) to increase the dictator's empathy for the receiver. Before their decision, dictators were asked to write three sentences about how they imagined the matched receiver's feelings and how their decision would affect the receiver. In the control condition, dictators had to write three sentences about the weather of the previous day. Finally, empathy-concern was measured by having dictators in both treatments rate on a scale from 0 to 6 their feelings towards the receiver (Davis, 1980; 1983). Al-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

though the average empathy-concern measure was significantly higher in the empathy treatment compared to the control treatment, Lönnqvist & Walkowitz (2019) did not find a significant effect of the empathy priming on dictators' other-regarding behavior.

In the papers mentioned above, gender has not been analyzed. Yet, some of the reported effects regarding altruistic behavior and/or honesty in DGs may have been driven by gender as women are expected to be more other-regarding than men. The evidence, however, is inconclusive. For instance, Bolton & Katok (1995) and Frey & Bohnet (1995) find no gender differences, whereas Eckel & Grossman (1998) and Andreoni & Vesterlund (2001) do, the latter in either direction depending on the cost of altruism. Based on his meta-analyses, Engel (2011) concludes that women are significantly more generous than men. Another gender difference may relate to honesty. The evidence, again, is inconclusive, see the meta-analysis of Gerlach et al. (2019) and the literature cited therein. The review analysis suggests, however, that men behave slightly more dishonestly in the coin-flip task than women do, and this difference is statistically significant. Given these findings, we take the conclusion of Andreoni & Vesterlund (2001) seriously that gender differences in economic experiments may be a factor that researchers have to examine carefully.

Our study contributes to the literature on gender effects in the dictator game by combining the above strands. We analyze how gender affects behavior when besides a direct choice the option exists that the private flip of a fair coin decides on an even (fair) distribution of the pie or a selfish allocation favoring the dictator. We further analyze whether, and if so, how empathy priming affects behavior.

B. Experiment

I. Experimental design

The dictator is provided with a pie of 10 € he has to distribute between himself and the receiver. All dictators also receive a coin labeled "8/2" on one side and "5/5" on the other side. The dictator can choose from three options. First, he can directly choose the selfish option "8/2" (8/2d) yielding 8 € for himself and 2 € for the receiver. Second, he can directly choose the equal split "5/5" (5/5d) yielding 5 € for both. Third, he can toss the coin and determine the distribution according to the outcome

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of the coin flip ("8/2c" or "5/5c") (Lönnqvist *et al.* 2014). The dictator is then asked to indicate his choice and whether he flipped the coin ("Yes" or "No"). Neither the experimenter nor the receiver could observe whether the dictator actually did flip the coin or whether he only reported having tossed it. The receiver does not make any decisions, and the game is played only once. 8/2 is chosen as the selfish option because the mean transfer in dictator games is approximately 20 % of the endowment (Camerer 2003), suggesting that many dictators feel comfortable with this division.

We conduct two treatments. Before dictators had made their decision and after being told their role (dictator / receiver), they are primed in the following ways. As in Lönnqvist & Walkowitz (2019), in the *Empathy* treatment, each dictator is asked to imagine the receiver's feelings, the well-being and how the dictator's decision would affect the receiver (perspective taking, see Coke *et al.* 1978; Fultz *et al.* 1986; Batson *et al.* 1988). Further, each dictator has to write at least three sentences about what he / she imagines. In the *Control* treatment, subjects are asked to think and write about yesterday's weather. Participants are confronted with this scenario to keep the procedure in both treatments as equal as possible. To analyze the empathy concern of the dictators, we ask them after their decision to indicate their feelings towards the receiver by rating on a scale from 0 to 6 for each of the following items whether they felt "understanding, soft-hearted, warm-hearted, compassionate, loving, moved" (Davis, 1980; 1983).

II. Procedures

Upon arrival, participants were randomly seated in computer cubicles that secured anonymity. At the beginning of the experiment, all participants received written information about the structure of the dictator game. Then each participant was assigned the role of either dictator or receiver. For reasons of cost efficiency only two participants per session were chosen to be receivers (see Walkowitz, 2021), since only decisions of dictators were of interest in our study. For payment implementation the two receivers were matched with two randomly chosen dictators and paid according to the corresponding dictator's decision. The computer-based experimental sessions were conducted at the University of Cologne (Germany) on the experimental platform z-Tree (Fischbacher 2007). The experiment lasted for about one hour including parts of the session not included in this paper. 234 participants of different majors participated.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

They were invited via the mailing list of the Laboratory of Economic Research. Eight experimental sessions were run, four per treatment. On average, dictators were paid a decision-dependent amount of 7€ and a show-up fee of 4 €.

C. Results

I. Descriptives

In total, 234 subjects participated in two treatments, 114 in the treatment *Control* and 120 in the treatment *Empathy*. The percentages of females are rather similar in both treatments (*Control*: 58.77; *Empathy*: 56.67, p= 0.745, Chi^2 test). Subjects are also comparable with regard to average age (23.41 vs. 23.27 years, p=0.788, Fisher-Pitman Permutation Test, two-tailed, FPPT in the following) and the share of German natives (86.84 % vs. 85.83 %, p=0.822, Chi^2 test). The average empathy-concern score in *Empathy* (2.74) is significantly higher than in *Control* (2.20) for the whole sample (p<0.001, FPPT) and separately for males (2.47 vs. 2.05, p=0.080) and females (2.95 vs. 2.31, p=0.004, FPPT). This suggests that perspective taking was effective and increased the empathy concern for the receiver. Further, we find a gender difference in *Empathy:* females state higher empathy concern scores than males after having been primed (2.95 vs. 2.47, p=0.026), but not in *Control* (2.31 vs. 2.05, p=0.285).

Regarding dictator behavior, reported coin use is very similar across treatments (45.83 % in *Empathy* vs. 41.23 % in *Control*, p=0.478, Chi^2 test). Although the fraction of dictators who choose an 8/2 allocation decreases in *Empathy*, and tends to be lower for females, we find no evidence that it differs significantly, neither between treatments for the whole sample (p=0.160, Chi^2 test) nor separately for men (p=0.221) or women (p=0.401). As to average monetary transfers to the receiver we find a similar picture. They increase in *Empathy*, and tend to be higher for females; yet, there is no significant difference between treatments for the whole sample, for men and women ($p\geq$0.157) separately, and also not between genders ($p\geq$0.238, both FPPT). Table 1 provides an overview over the characteristics of the dictators and the transfers in € to the receivers differentiated by gender and treatment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Gender-specific treatment effects

We now analyze how *Empathy* affects the distribution of the four outcomes for males and females. See Figure 1 for the percentages of chosen options.

Table 1: Dictators' characteristics and behavior by gender and treatment

Gender	Male		Female	
Variables/Treatment	Control	Empathy	Control	Empathy
Age	24.13 (3.66)	23.42 (4.59)	22.91 (2.72)	23.14 (4.35)
German (%)	82.98	84.62	89.55	86.76
Empathy concern (Avrg.)	2.05 (1.31)	2.47 (0.98)	2.31 (1.26)	2.95 (1.39)
Reported coin use (%)	36.17	50.00	44.78	42.65
8/2 allocations (%)	76.60	65.38	68.66	61.76
Monetary transfer (Avrg. % of 10€-pie)	27.02 (12.83)	30.38 (14.41)	29.40(14.02)	31.47(14.69)
Observations	47	52	67	68

Note: This table shows per treatment and differentiated by gender: average age, share of German natives, average empathy concern, share of reported coin use, share of 8/2 allocations and the average monetary transfer to the receiver in percent of the 10€-pie. Average empathy concern displays the average over five items of the empathic concern subscale of the Interpersonal Reactivity Index (IRI) developed by Davis (1980, 1983). The items (understanding, soft-hearted, warm-hearted, compassionate, loving, moved) are adjectives used in the stem of the Empathic Concern scale (e.g., "I am often quite touched by things that I see happen") or ratings (0 to 6) of how well each adjective describes oneself. Standard deviation in parentheses.

Result 1 (Gender-specific treatment effects):

Empathy shifts behavior of males from direct 8/2 choices towards coin-based 8/2 allocations, and, partly, to 5/5 outcomes. For females, *Empathy* induces weaker reductions in direct 8/2 allocations and instead encourages more direct 5/5 outcomes.

Support males:

Empathy priming reduces the frequency of 8/2d significantly (p=0.006, FPPT) compared to *Control*, while 8/2c and 5/5d tend to increase (p=0.157, p=0.203, FPPT, respectively). Among those dictators who implement the selfish allocation, reported coin use increases significantly in *Empathy* (p=0.014).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

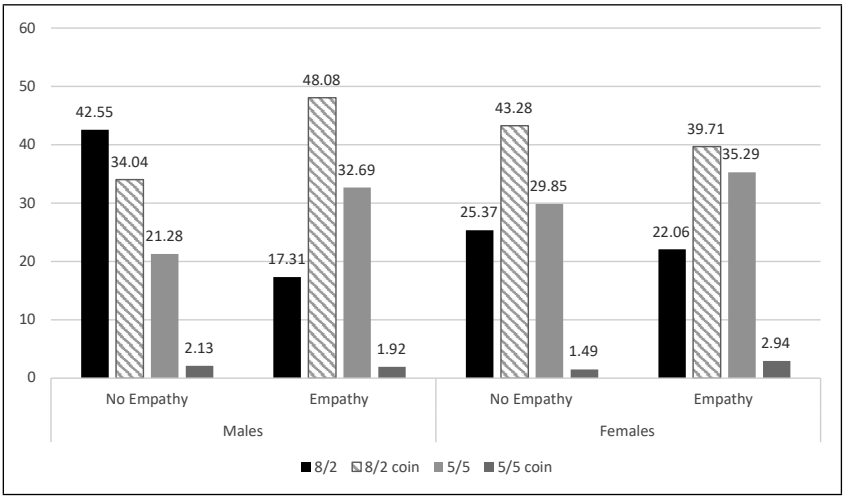


Figure 1. Allocations and Reported Coin Use across Treatments for both Genders

Support females:

Empathy lowers 8/2d and 8/2c outcomes relative to *Control*, but only slightly ($p \geq 0.651$, FPPT). Although we observe the highest increase in percentage points in 5/5d choices, this is not significant ($p=0.500$, FPPT). In contrast to males, reported coin use does not change in *Empathy* for those female dictators who choose the selfish allocation ($p=0.904$). Note that both, males and females, are cheating when reporting the coin-flip results. 8/2c occurs highly significantly more often than would be expected by the toss of a fair coin ($p<0.0001$, binomial test). 5/5c is reported very seldom and stays unchanged.

III. Mechanisms

In the following, we investigate for both genders the mechanisms behind their choices in *Empathy* compared to *Control* by regression analyses. We first analyze the determinants for reporting coin use of dictators across males and females. Secondly, we examine the drivers of selfish 8/2 choices in the two treatments for each gender.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

1. Reported coin use

We estimate a series of probit-regression models (Table 2), and restrict our analysis to those dictators who chose the selfish 8/2 outcome, as 5/5c was almost never chosen. To analyze the impact of empathy concern on coin use, we generate the variable "High-empathy-concern", which is gender specific and is coded = 1 if a male (female) dictator has a higher average empathy-concern score than the median of all average empathy-concern values for males (females). It is coded = 0 if the average empathy-concern score is lower or equal to the median of the gender-specific values.

Result 2 (Reported coin use):

We find a notable behavioral difference between genders. Whereas the variables Empathy and High-empathy-concern emerge as significant and robust predictors of coin use in the male sample, there is no discernible effect for females. Females' reported coin use appears not to be driven by treatment or high empathy concern.

Support males:

The probability of choosing 8/2c increases significantly in *Empathy* compared to *Control* (Table 2, Model 1), suggesting that for men, the empathy priming increases the probability to engage in coin use. High-empathy-concern also has a significant positive effect on the probability to use the coin compared to low-empathy concern subjects (Model 2). When both variables are included in the regression (Model 3), empathy-priming and high-empathy-concern remain positive and significant, indicating that these factors exert independent effects on the likelihood of coin use. Model 4 introduces an interaction term between *Empathy* and high-empathy-concern. Both coefficients remain positive and (weakly) significant, whereas the interaction term is negative, and not statistically significant. This indicates that the effect of empathy-priming on coin use does not vary substantially by levels of empathy concern. Age has a negative impact on reported coin use.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 2: Probability of coin use when choosing the selfish option, differentiated by gender

	Males				Females			
Variables/Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Empathy	0.779**		0.788**	0.828**	0.055		0.050	-0.032
	(0.329)		(0.335)	(0.422)	(0.278)		(0.279)	(0.333)
High-empathy-concern		0.753**	0.762**	0.815*		0.298	0.297	0.155
		(0.337)	(0.348)	(0.457)		(0.317)	(0.317)	(0.435)
Empathy x High-empathy-concern				-0.127				0.301
				(0.729)				(0.640)
Age	-0.087**	-0.099**	-0.081*	-0.084*	-0.080	-0.086*	-0.086*	-0.093*
	(0.042)	(0.045)	(0.044)	(0.046)	(0.049)	(0.049)	(0.049)	(0.050)
German native	0.469	0.381	0.646	0.620	0.029	0.073	0.079	0.061
	(0.535)	(0.607)	(0.632)	(0.643)	(0.408)	(0.408)	(0.408)	(0.417)
Constant	1.491	1.925	0.909	0.979	2.146*	2.185*	2.168*	2.379*
	(1.303)	(1.323)	(1.406)	(1.443)	(1.189)	(1.165)	(1.170)	(1.225)
Observations	70	70	70	70	88	88	88	88

Notes: Differentiated by gender, this table shows results from probit regressions on the probability of coin use when 8/2 is chosen. Standard errors clustered at subject level in parentheses. The reference category for treatment is *Control*. High-empathy-concern is a dummy coded 1 if the empathy-concern score is above the median for males (females). Empathy x high-empathy-concern is the interaction term of the two variables Empathy and-High empathy-concern. 70 males (88 females) report coin use. Age and being a German native have a negative impact on the probability to use the coin, which however, is not robust.

Support females:

In contrast to the male subsample, empathy priming does not significantly affect the probability of coin use in any specification (Models 5–8). Likewise, high-empathy-concern and its interaction with *Empathy* fail to achieve statistical significance (Models 6 to 8).

2. Selfish 8/2 choices

Next, we estimate a series of probit-regressions to study the determinants of 8/2 choices in the two treatments differentiated by gender (Tables 3 and 4). We analyze how reported coin use and high empathy concern affect the probability of choosing the selfish allocation.

Result 3 (Selfish behavior of males and females):

For males, reported coin use has little effect in *Control*, but becomes a strong predictor of selfishness in *Empathy*. For females, reported coin use is already a strong predictor of selfish allocations in *Control*. For both genders, high-empathy-concern consistently reduces the probability of choosing the selfish option. When differentiating by reported coin use,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 3: Probability of males choosing the selfish allocation 8/2 when reporting coin use, differentiated by treatment

	Control				Empathy			
Variables/Model	(1)	(2)	(3)	(4)*)	(5)	(6)	(7)	(8)
Coin use	0.820		0.814	-	2.322***		2.890***	4.873***
	(0.585)		(0.710)	-	(0.509)		(0.678)	(0.503)
High-empathy-concern		-1.202**	-1.201**	-		-1.484***	-2.172***	-2.168***
		(0.508)	(0.562)	-		(0.486)	(0.674)	(0.680)
Coin use x High-empathy-concern				-				-1.987**
				-				(0.831)
Age	0.004	0.018	0.029	-	0.044	-0.060	0.006	0.006
	(0.065)	(0.072)	(0.071)	-	(0.048)	(0.043)	(0.052)	(0.052)
German native	1.075*	1.380**	1.088**	-	-0.010	-0.371	-0.860	-0.859
	(0.555)	(0.561)	(0.538)	-	(0.631)	(0.531)	(0.784)	(0.784)
Constant	-0.429	-0.095	-0.376	-	-1.474	3.112**	1.415	1.412
	(1.721)	(1.851)	(1.778)	-	(1.520)	(1.452)	(1.863)	(1.865)
Observations	47	47	47	47	52	52	52	52

Notes: Differentiated by treatment, this table shows results from probit regressions on the probability of males choosing the selfish option 8/2 when reporting coin use. Standard errors clustered at subject level in parentheses. 47 (52) males chose 8/2 in *Control* (*Empathy*). Age and being a German native have no impact on the probability to choose 8/2. See the notes of Table 2 for further variable definitions. *)We could not estimate the full Model 4 including the interaction term "Coin use x High-empathy-concern", because all but one male dictator in the high-empathy-concern group who reported coin use, implemented the selfish allocation.

high-empathy-concern is associated with a lower probability of a selfish allocation for those who use the coin and for those who don't. For low-empathy-concern coin-users, selfish allocations are more likely.

Support males:

In *Control*, coin use does not exert a significant effect on selfishness in any of the models 1 to 3. By contrast, high empathy-concern consistently reduces 8/2 choices (Models 2 and 3, Table 3). Note that in *Control*, we could not estimate a model including the interaction term "Coin use x High empathy concern", because all but one dictator in the high-empathy group who reported coin use, implemented the selfish allocation. German natives tend to behave more selfishly. In *Empathy*, reported coin use emerges as a positive predictor of selfish behavior, with large and highly significant coefficients in Models 5, 7, and 8. High-empathy-concern significantly reduces selfishness (Models 6 to 8). Model 8 conveys that the effect of coin use on choosing the selfish allocation is weaker for people with high empathy concern, i.e., high empathy moderates the probability of choosing the 8/2 outcome when the coin is used.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 4: Probability of females choosing the selfish allocation 8/2 when reporting coin use, differentiated by treatment

	Control				Empathy			
Variables/Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coin use	2.022***		2.195***	5.405***	1.816***		2.104***	4.344***
	(0.504)		(0.552)	(0.340)	(0.413)		(0.535)	(0.547)
High-empathy-concern		-1.058***	-1.168***	-1.105**		-2.335***	-2.645***	-2.637***
		(0.346)	(0.419)	(0.455)		(0.487)	(0.572)	(0.583)
Coin use x High-empathy-concern				-3.349***				-2.246***
				(0.724)				(0.754)
Age	0.060	-0.012	0.055	0.055	0.003	0.036	0.057	0.056
	(0.074)	(0.062)	(0.073)	(0.074)	(0.035)	(0.044)	(0.041)	(0.041)
German native	0.051	-0.477	-0.214	-0.232	-0.415	-0.271	-0.635	-0.634
	(0.546)	(0.550)	(0.463)	(0.471)	(0.504)	(0.549)	(0.582)	(0.581)
Constant	-1.552	1.679	-0.670	-0.686	-0.021	1.287	0.722	0.716
	(1.702)	(1.517)	(1.722)	(1.748)	(0.939)	(1.172)	(1.204)	(1.208)
Observations	67	67	67	67	68	68	68	68

Notes: Differentiated by treatment, this table shows results from probit regressions on the probability of females choosing the selfish option 8/2 when reporting coin use. Standard errors clustered at subject level in parentheses. 67 (68) females chose 8/2 in *Control* (*Empathy*). Age and being a German native have no impact on the probability to choose 8/2. See the notes of Table 2 for further variable definitions.

Support females:

In *Control*, coin use is a strong and consistent predictor of selfish behavior, with significant positive coefficients (Table 4, Models 1, 3, and 4), which is in contrast to the male sample. In line with the male sample, however, high-empathy-concern significantly reduces selfishness (Models 2 to 4). The inclusion of the interaction term in Model 4 shows a moderation effect: The strength of the relationship between coin use and selfishness is reduced. Very similar effects can be found in the *Empathy* treatment (see Models 5 to 8 in Table 4), which is also in line with men's behavior.

D. Concluding Remarks

Taken together, we find that the empathy treatment significantly shapes dictators' allocation behavior, but in gender-specific ways. Males' choices shift away from direct selfish 8/2 choices toward coin-based 8/2 decisions and, to a lesser extent, toward equal 5/5 outcomes. For females, the reduction in direct 8/2 allocations is weaker, with empathy priming promoting more direct 5/5 choices instead. Robust predictors of reported coin use in the male sample are empathy priming and high-empathy-concern, while no such effects are observed for females, whose reported coin

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

use appears unrelated to treatment or empathy concern. Furthermore, coin use strongly predicts selfish allocations for males only in the empathy condition, whereas for females it is already associated with selfishness in *Control*. For both genders, however, higher empathy concern consistently reduces selfish behavior, mitigating the link between coin use and selfish allocations.

Our results relate to previous findings on altruistic behavior by showing that empathy influences fair choices in distinct, gender-specific ways. Prior research highlights empathy as a key driver of prosociality and reduced selfishness (Batson, 2011; Eisenberg & Miller, 1987), yet our results suggest that the way this prosociality is manifesting itself differs across men and women. For men, empathy appears to displace direct selfish allocations toward coin-based decisions. For women, empathy promotes a shift toward more 5/5 allocations. This difference could be explained by the fact, that women are generally more prosocial than men (Engel, 2011), i.e., the underlying distributions of men and women in terms of their prosociality are different. This assumption is supported for empathy priming by significantly higher empathy scores of females compared to males (see Table 1) and by the fact, that males tend to choose 8/2d (5/5d) more (less) often than females in *Control* (p=0.054), FPPT.

Due to the different underlying distributions of prosociality among men and women, empathy priming has a different effect across genders: Because many men in *Control* choose 8/2 directly, the empathy priming primarily results in a shift toward 8/2c. Men use the coin as a strategic device to eschew responsibility for their actions and to reduce/avoid cognitive dissonance (Hertwig & Engel, 2016). Thus, they preserve their image without losing money if their image is important for themselves, vis-à-vis the receiver or the experimenter. Our regressions have shown that among men, the coin report is associated with higher empathy-concern values and that coin use in the empathy treatment predicts selfish behavior. Contrarily, in *Control*, many female dictators already choose 8/2c or 5/5, so that after being primed with empathy, a shift from 8/2c to 5/5 seems feasible, but this comes at the cost of € 3 – a higher cost compared to men who switch from 8/2d to 8/2c. Woman's reluctance to take this costly step is supported by our regression results conveying that, unlike for men, empathy priming and empathy concern have no influence on coin use while the interplay of coin use and high-empathy-concern in predicting the selfish outcome is similar to men's behavior.

We have to acknowledge some limitations of our study. Arguably,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

with an increased sample size the reported effects are expected to be stronger. Moreover, our design cannot distinguish whether empathy concern is a personality trait or whether it is induced by the empathy priming, i.e., whether it is trait based, or state based. Further studies measuring empathy concern before and after priming would have to clarify this open question.

Reported coin use further connects to the literature on honesty and moral wiggle-room. Previous studies show that randomization devices allow individuals to maintain a prosocial self-image while behaving selfishly (e.g., Fischbacher & Föllmi-Heusi, 2013; Lönnqvist et al., 2014; Apffelstaedt et al., 2025). Our results indicate that reported coin use operates differently by gender: for men, its link to selfishness emerges mainly under empathy priming, whereas for women, coin use already predicts selfish allocations even absent treatment. Furthermore, we find that both men and women who claim to have used the coin to decide on the choice option, are very likely to be lying. Across both groups, however, high empathy concern reliably counteracts selfishness, reinforcing the view that empathy serves as a stable internal barrier on exploitative behavior. Taken together, these results suggest that empathy not only promotes altruism but also shapes the strategies through which individuals balance fairness, honesty, and self-interest.

Acknowledgements

The data for this project were collected by Margarita Buchholz and Gari Walkowitz as part of her master's thesis project. Financial support by the Center for Social and Economic Behavior (C-SEB) at the University of Cologne, as well as by the German Research Foundation (DFG, Subproject "Incentives and Ethical Behavior") are gratefully acknowledged.

References

Abeler, J., Becker, A., Falk, A. (2014). Representative Evidence on Lying Costs. *Journal of Public Economics*, 113, 96–104. https://doi.org/10.1016/j.jpubeco.2014.01.005.

Abeler, J., Nosenzo, D., Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, 87, 1115–1153. https://doi.org/10.3982/ECTA14673

Apffelstaedt, A., Lönnqvist, J.E., Irlenbusch,B., Walkowitz, G. (2025). "Figleafing" and Selfish Behavior. Working Paper, University of Cologne.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Andreoni, J., Bernheim, B.D. (2009). Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77, 1607–1636. https://doi.org/10.3982/ECTA7384

Andreoni, J., Vesterlund, L. (2001). Which Is the Fair Sex? Gender Differences in Altruism. *Quarterly Journal of Economics*, 116, 293–312. doi.org/https://doi.org/10.1162/003355301556419

Batson, C. D. (1991). The Altruism Question: Toward a Social-Psychological Answer. Hillsdale, NJ: Erlbaum.

Batson, C.D., Kobrynowicz, D., Dinnerstein, J.L., Kampf, H.C., Wilson, A.D. (1997). In a Very Different Voice: Unmasking Moral Hypocrisy. *Journal of Personality and Social Psychology*, 72, 1335–1348. http://dx.doi.org/10.1037/0022-3514.72.6.1335

Batson, D.C. (2011). Altruism in Humans. Oxford University Press, USA.

Bicchieri, C. (2005). The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press.

Bicchieri, C. (2016). Norms in the Wild: How to Diagnose, Measure, and Change Social Norms. Oxford University Press.

Bohnet, I., Frey, B. (1995). Ist Reden Silber und Schweigen Gold? *Zeitschrift für Wirtschafts- und Sozialwissenschaften,* 115, 169–209. https://doi.org/10.3790/schm.115.2.169

Bolton, G., Katok, E., (1995). An Experimental Test for Gender Differences in Beneficent Behavior. *Economics Letters,* 48, 287–292. https://doi.org/10.1016/0165-1765(94)00621-8

Bucciol, A., Piovesan, M. (2011). Luck or Cheating? A Field Experiment on Honesty with Children, *Journal of Economic Psychology*, 32, 73–78. https://doi.org/10.1016/j.joep.2010.12.001.

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press.

Cerrone, C., Engel, C. (2019). Deciding on Behalf of Others Does Not Mitigate Selfishness: An Experiment. Economics Letters, 183, 108616. https://doi.org/10.1016/j.econlet.2019.108616

Coke, J.S., Batson, D.C., McDavis, K. (1978). Empathic Mediation of Helping: A Two-Stage Model. *Journal of Personality and Social Psychology*, 36, 752–766. doi:10.1037/0022-3514.36.7.752

Davis, M. H. (1980). A Multidimensional Approach to Individual Differences in Empathy. *JSAS Catalog of Selected Documents in Psychology, 10, 85*. (Manuscript No. 2124).

Davis, M. H. (1983). Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach. *Journal of Personality and Social Psychology*, 44, 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Eckel, C. C., Grossman, P. J. (1998). Are Women Less Selfish than Men? Evidence from Dictator Experiments. *Economic Journal*, 108, 726–735. https://www.jstor.org/stable/2565789

Engel, C. (2011). Dictator games: A Meta Study. *Experimental Economics* 14, 583–610 (2011). https://doi.org/10.1007/s10683-011-9283-7

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Fischbacher, U. (2007). Z-tree: Zurich Toolbox For Ready-Made Economic Experiments. *Experimental Economics*, 10, 171–178. https://doi.org/10.1007/s10683-006-9159-4

Fischbacher, U., Föllmi-Heusi, F. (2013). Lies in Disguise – An Experimental Study on Cheating, *Journal of the European Economic Association*, 11, 525–547. https://doi.org/10.1111/jeea.12014.

Forsythe, R., Horowitz, J. L., Savin, N. E., Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6, 347–369. https://doi.org/10.1006/game.1994.1021

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The Truth about Lies: A Meta-Analysis on Dishonest Behavior. *Psychological Bulletin*, 145(1), 1–44. https://doi.org/10.1037/bul0000174

Hertwig, R., Engel, C. (2016). Homo Ignorans: Deliberately Choosing Not to Know. Perspectives on Psychological Science, 11, 359–372. https://doi.org/10.1177/1745691616635594

Lönnqvist, J.-E., Walkowitz, G. (2019). Experimentally Induced Empathy Does Not Affect Monetarily Incentivized Dictator Game Behavior. *Frontiers in Psychology, 10, 337*. https://doi.org/10.3389/fpsyg.2019.00337

Lönnqvist, J.-E., Irlenbusch, B. & Walkowitz, G. (2014). Moral Hypocrisy: Impression Management or Self-Deception? *Journal of Experimental Social Psychology*, 55, 53–62. https://doi.org/10.1016/j.jesp.2014.06.004

Roth, A. E. (1995). Bargaining experiments. In: Handbook of Experimental Economics, J.H. Kagel, A.E. Roth, Eds. Princeton Univ. Press, Princeton, N.Y., 235–348.

Walkowitz, G. (2021). Dictator Game Variants with Probabilistic (and Cost-Saving) Payoffs: A Systematic Test. *Journal of Economic Psychology*, 85, 102387, https://doi.org/10.1016/j.joep.2021.102387.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Vertical Mergers and Price Leadership
An Experiment

Claudia Moellers & Hans-Theo Normann

While an influential line of research maintains that vertical mergers lead to higher prices in the input market through a "raising-rival's-cost" effect, this result has been challenged on the basis that it depends on the assumption that integrated firms can commit not to compete in the input market. We argue that price leadership can work as a commitment device for integrated firms. When an integrated firm is the price leader, the prediction is that the input market will become monopolized. This outcome is subgame perfect, meaning that there is no commitment problem. We test this prediction in experiments, comparing the price-leader data to treatments involving simultaneous moves and variants without vertical integration. We find that when integrated firms act as price leaders, the input market is indeed fully monopolized. After several periods of play, we impose an unexpected change: Subjects no longer play the sequential leader-follower game but move simultaneously. Integrated firms then no longer charge higher prices. In other words, integrated firms do not learn to charge high prices and achieve Pareto-superior outcomes.

A. Introduction

One of the main arguments in the literature on vertical integration is that such mergers may be anticompetitive due to a "raising-rival's-costs" effect. This argument was first formalized by Ordover, Saloner and Salop (1990), henceforth referred to as OSS (1990), and it goes as follows. When a vertically integrated firm forecloses nonintegrated downstream firms, that is, when it withdraws from the input good market, upstream competition becomes weaker. This reduction of competition implies higher input costs for the nonintegrated downstream firms. Since the downstream unit of the integrated firm benefits when rivals' costs are raised, the integrated firm is better off pursuing such a foreclosure strategy compared to the case where it actively competes.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

While OSS (1990) has been rather influential in the theoretical industrial organization literature[1], it has also been criticized. Hart and Tirole (1990), as well as Reiffen (1992), have pointed out that the outcome hinges on the premise that integrated firms can commit not to supply nonintegrated downstream rivals. Without this commitment, vertically integrated firms will compete just like the other upstream firms. In the static Nash equilibrium without commitment, nonintegrated downstream firms are not foreclosed and equilibrium prices are the same with and without the vertical merger.

The motivation for this paper is the insight that integrated firms arguably have an incentive to *seek* commitment whereas non-integrated firms will not. Even when considering the critique by Hart and Tirole (1990) and Reiffen (1992)'s that commitment is unavailable in the static simultaneous-move model, it is evident that integrated firms may look for (or invest into) the sort of commitment which OSS (1990) assume in a *deus-ex-machina* fashion. So even if commitment is unavailable in the standard game, integrated firms might still strive for commitment in extended settings.

To illustrate this issue, we propose a model of price leadership. In this model, the integrated firm is the price leader. It chooses the input price first, and the non-integrated firm chooses second. According to the SGPN prediction, monopolization of the input market occurs. This outcome is Pareto-superior from the firms' perspective compared to a game with simultaneous moves. Non-integrated firms would also never want to become price leaders because this would result in zero profits for both firms. Therefore, integrated firms benefit from being the price leader and will seek such an opportunity.

We complement the theoretical analysis with experimental data. We implement the variant where the integrated firm is the price leader and we compare this variant to (i) a treatment without price leadership but with vertical integration and (ii) a treatment without price leadership and without vertical integration. This allows us to evaluate the impact of price leadership in conjunction with vertical integration. To study learning effects, we replace the sequential leader-follower game with simultaneous

1 Salinger (1988), Hart and Tirole (1990) and OSS are seminal among the "new" foreclosure theories. Martin, Normann and Snyder (2001) and Rey and Tirole (2007) contain reviews of foreclosure theories.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

moves after several periods of play and let subjects play the treatment without price leadership (but still with vertical integration).

We find that, when integrated firms are the price leader, they indeed charge a price above the monopoly level and the second movers best respond by charging the monopoly price. So as predicted, the input market is monopolized. While this outcome is Pareto-superior to the static Nash equilibrium, it turns out that learning effects are weak. When they are no longer the price leader, the integrated firms do not charge higher prices than non-integrated firms. In other words, integrated firms do not learn to commit.

The next section discusses the literature. Section 3 introduces the general model and reports on the simultaneous-move game and the sequential-move game. Section 4 contains the experimental design and Section 5 the procedures. Section 6 reports the experimental results. Section 7 is the conclusion.

B. Related Literature

Following the critique by Hart and Tirole (1990) and Reiffen (1992), various papers have shown that OSS's foreclosure result can be rigorously derived from game-theoretic models. OSS (1992) re-establish their result in a descending-price auction. Riordan (1998) analyses backward integration by a dominant firm with a cost advantage. Choi and Yi (2000) and Church and Gandal (2000) show the result if upstream firms can commit to a technology which makes the input incompatible to nonintegrated downstream firms. Chen and Riordan (2007) investigate the connection between vertical integration and exclusive dealing contracts. Our paper contributes to this literature by showing equilibrium vertical foreclosure in a sequential-move version of the original OSS model.

Along this line of research, various papers have shown that vertical integration can lead to collusive effects. Chen (2001) considers not only the change in incentives upstream but also in the downstream market for the case of a vertical merger. He finds collusive effects but also efficiency gains and an ambiguous result for competitive effects in general. Nocke and White (2007) analyze vertical integration in a market with two-part tariffs upstream and repeated interaction. They show that a vertical merger facilitates upstream collusion. Normann (2009) shows that collusion is easier to sustain in a vertically integrated market also with linear

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

prices. Related to our issue (but not covering issues of vertical integration), Mouraviev and Rey (2011) show that price leadership can facilitate collusion. In a theoretical model, they show that the choice of deciding simultaneously or sequentially about prices can sustain perfect collusion.

Allain et al. (2011) show that the necessity of downstream firms to share sensitive information once they trade with an upstream firm might lead to input foreclosure. In a static game with two upstream firms and vertical integration resembling OSS (1990), the non-integrated downstream firm might be reluctant to exchange information which cannot be protected by property rights. Deals between an integrated upstream supplier and non-integrated downstream firms might not occur due to the concern that information will be leaked the downstream division. Allain et al. (2016) show that vertical integration can create hold-up problems for competitors. If the integrated supplier can commit to be "greedy" or alternatively commits to offer a degraded input to the downstream competitor, hold-up problems occur. On the other hand, they show that even without commitment, foreclosure emerges if the quality of the upstream product is non-verifiable.

More recently, partial vertical ownership, where one firm controls an upstream or downstream company through a partial stake, has been in the interest of researchers. Partial vertical ownership can create stronger incentives for foreclosure than a full merger. This is the main argument put forth by Spiegel (2013) and Levy, Spiegel, and Gilo (2018). See also the extension by Hunold and Werner (2025).

The first experimental study on vertical mergers is Martin et al. (2001).[2] They analyze the commitment problem of an upstream monopolist to restrict the total quantity for downstream firms to the monopoly level (Hart and Tirole, 1990). Public contracts between downstream firms and the upstream monopolist and, alternatively, vertical integration result regularly in monopolization of the input market. In contrast, if firms are independent and contracts are secret, beliefs of downstream firms about the contract offer to the rival determine the

2 Mason and Phillips (2000) analyze the double marginalization problem in a market with two upstream and two downstream firms. They find larger outputs and a higher consumer surplus with both firms vertically integrated as compared to no integration. Durham (2000) finds support for the double marginalization problem if upstream and downstream markets are monopolized, whereas competition downstream eliminates this problem.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

outcome. In this case, monopoly power cannot be sustained and market quantity is significantly above the monopoly level. Recently, Moellers et al. (2017) extend this study and analyze the impact of communication on the commitment problem. They find that open communication leads to monopolization whereas bilateral communication between the producer and retailers do not lead to the monopoly quantity downstream.

Normann (2011) analyzes experimentally the effect of vertical integration on selling prices and market foreclosure in an OSS (1990) framework. Although he finds a significant increase in the minimum price paid by the independent downstream firm, there is little evidence for total foreclosure. The integrated firm does not withdraw completely from the input market. However, partial foreclosure (that is, the integrated firm sets a higher price than the non-integrated firm), indeed occurs. Moellers (2016) extends this analysis to one-shot games where firms can choose to reveal their choice history. She finds that integrated firms choose to reveal past prices more often than non-integrated firms. When integrated firms reveal their choice history, prices are higher, as predicted. In an intriguing experiment, Allain et al. (2021) find support for the predictions in Allain et al. (2016). Vertical integration creates hold-up problems, in particular, if commitment is possible.

C. Theory

I. Genereal setup

The basic setup of our model is as in OSS (1990) and comparable to Normann (2009) who, however, analyses the repeated game. There are $n = 2$ upstream firms and $m = 2$ downstream firms. See Figure 1. We call the upstream firms $U1$ and $U2$ and the downstream firms $D1$ and $D2$. The integrated firm will be called $U1$-$D1$. The upstream firms produce a homogeneous input. $D1$ and $D2$ transform the input on a one-to-one basis into a symmetrically differentiated final good.

Downstream firms pay a linear price for the input which constitutes their only cost. Define c_i as the price per unit firm Di pays. There is differentiated price competition at the downstream level and $Q_i\left(p_i, p_j\right)$ denotes the demand function of the final consumers. Because of the one-to-one transformation it determines at the same time the demand of Di for the input good, $i, j = 1, 2, i \neq j$. Symmetry implies that sales depend on prices

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

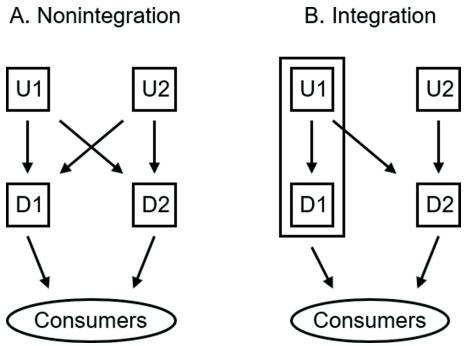


Figure 1: Market Structure

but not on the identities of the firms, that is, $Q_1(p_1, p_2) = Q_2(p_2, p_1)$. *Di*'s profits are

$$\pi_{Di} = (p_i - c_i)\, Q_i\left(p_i, p_j\right),\ i, j = 1, 2,\ i \neq j.$$

We impose the following assumptions on demand. Demand functions $Q_i\left(p_i, p_j\right)$ are twice continuously differentiable with $\frac{\partial Q_i}{\partial p_i} < 0$, $\frac{\partial Q_i}{\partial p_j} > 0$, and $\frac{\partial Q_i}{\partial p_i} + \frac{\partial Q_i}{\partial p_j} < 0$, $i, j = 1, 2$, $i \neq j$. These assumptions ensure downward sloping demand with substitute goods where the effect of a change in a firm's own prices dominates the effect resulting from a change of the rival firm's price. Further, we assume that goods are strategic complements, that is, $\frac{\partial^2 \pi_{Di}}{\partial p_i \partial p_j} > 0$. A final assumption is that $\frac{\partial^2 \pi_{Di}}{\partial p_i^2} + \frac{\partial^2 \pi_{Di}}{\partial p_i \partial p_j} < 0$. This assumption implies that own effects dominate cross effects also in terms of the slope of the demand function. Together with the other assumptions, this is sufficient to ensure the existence of a unique Nash equilibrium of the stage game.

Denote $p_i^*\left(c_i, c_j\right)$, i, j=1, 2, $i \neq j$, for the static Nash equilibrium prices at the downstream level. Since input prices $\left(c_i, c_j\right)$ sufficiently describe Nash equilibrium downstream competition, and we will employ $Q_i^*\left(c_i, c_j\right)$ as a reduced form for $Q_i\left(p_i^*\left(c_i, c_j\right), p_j^*\left(c_j, c_i\right)\right)$, and $\pi_{Di}^*\left(c_i, c_j\right)$ for $\pi_{Di} = \left(p_i^*\left(c_i, c_j\right) - c_i\right) Q_i^*\left(c_i, c_j\right)$. We assume costs are such that positive demands $Q_i^*\left(c_i, c_j\right) > 0$ occur throughout.

Given the above assumptions, OSS (1990) show that raising the cost of a downstream rival is profitable. That is,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

$$\frac{\partial \pi_{Di}^{*}\left(c_i, c_j\right)}{\partial c_j} = \left(p_i - c_i\right) \frac{\partial Q_i\left(p_i, p_j\right)}{\partial p_j} \frac{\partial p_j^{*}\left(c_i, c_j\right)}{\partial c_j} > 0, i, j = 1, 2, i \neq j, \tag{1}$$

as follows from the envelope theorem.

We finally turn to the upstream level. We assume that that upstream firms, $U1$-$D1$ and $U2$, have constant marginal cost which we normalize to zero. Let w_i denote the price upstream firm i, $i = 1, 2$, charges on the external input market.

II. Simultaneous-move game

Following OSS (1990) and subsequent literature, we begin with the assumption that $U1$ and $U2$ choose input prices simultaneously. With simultaneous moves, upstream firms are perfect Bertrand competitors. This implies that the lower of the two upstream prices constitutes the input price. Given the input price, downstream firms purchase the number of units they require, $Q_i\left(p_i, p_j\right)$, if they procure externally at all. Below, we will specify how much the downstream firms purchase from the two upstream firms when they charge the same price.

When no firm is vertically integrated (Panel A of Figure 1), upstream competition à la Bertrand implies that Nash equilibrium prices are equal to the marginal cost on the input market. This static Nash equilibrium is unique. Both downstream firms purchase the good externally on the input market and they pay the same price for it. So, we have $w_1 = w_2 = c_1 = c_2 = 0$.

Now suppose that $U1$ and $D1$ are integrated (Panel B of Figure 1). The downstream segment of $U1$-$D1$ obtains the input internally at $c_1 = 0$ (Bonanno and Vickers, 1988). For now, we will set aside the reasoning of OSS (1990) and follow the conclusions of Hart and Tirole (1990) and Reiffen (1992) that $w_1 = w_2 = c_1 = c_2 = 0$ just as in the Nash equilibrium without integration.

Proposition 1. In the simultaneous-move game, the U1-D1 merger has no impact because the input market is competitive with or without integration.

Before moving on to the sequential-move case, let us review OSS (1990)'s argument. The novel insight of OSS (1990) was that the verti-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

cal integration changes the incentives to compete in the input market. If $U1$-$D1$ can credibly commit to withdrawing from the input market, $U2$ will became the sole supplier of $D2$. $D2$ will pay some $c_2 > 0$ for the input and $U1$-$D1$ earns $\pi^*_{D1}(0, c_2)$. Now, suppose instead that $U1$-$D1$ competes in the input market, as argued above. Then we obtain $c_2 = 0$ and $U1$-$D1$ earns $\pi^*_{D1}(0, 0) < \pi^*_{D1}(0, c_2)$. In other words, $U1$-$D1$ makes no upstream profit in either case, but, when it withdraws from the input market, it makes a higher downstream profit, $\pi^*_{D1}(0, c_2) > \pi^*_{D1}(0, 0)$, due to the raising-rivals'-costs effect (see equation (1)). Hence, $U1$-$D1$ prefers to withdraw from the input market. In the static simultaneous-move game, there is no commitment, however, and $U1$-$D1$ has an incentive to deviate in this situation (as emphasized by Hart and Tirole, 1990, and Reiffen, 1992). It will re-enter the input market, contrary to its claim to withdraw, and undercut $U2$'s input price. As $U2$ will anticipate this deviation, the static Nash equilibrium has both $U1$-$D1$ and $U2$ charging a price equal to marginal cost, that is, $c_1 = c_2 = 0$ just as in the case without integration. $U2$ earns zero profits and $U1$-$D1$ earns $\pi^*_{D1}(0, 0)$ in the static Nash equilibrium.

III. Sequential moves

Now assume that $U1$ is the price leader. $U1$ sets its price, w_1, first and, knowing w_1, $U2$ moves second. Otherwise, the setup is as above.

In any subgame perfect Nash equilibrium of the sequential-move game, $U2$ will play a best response to w_1. To define this best response, we need to define a benchmark, namely the price of the input, w_2, that maximizes $U2$'s profits if it is a monopolist in the input market and $U1$-$D1$ plays its myopic best reply at the downstream level, $p^*_1(0, w_2)$ where $c_2 = w_2$. Denote this price by w_2^{mon} and define formally

$$w_2^{mon} := arg \quad max_{w_2} w_2 Q^*_2(w_2, 0).$$

Accordingly, define $\pi_2^{mon} := w_2^{mon} Q^*_2(w_2^{mon}, 0)$.[3] We can now define $U2$'s best response as

$$w_2^{BR}(w_1) = \min\{\max\{0, w_1 - \varepsilon\}, w_2^{mon}\}.$$

3 The input-good price w_2^{mon} is also central to the analysis in OSS (1990).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In words, $U2$ it will undercut $U1$ price by an infinitesimally small margin, ε, unless w_1 is strictly larger than $U2$'s preferred monopoly price, or unless w_1 is negative.

It is now straightforward to see the impact of price leadership. We know that $U2$ will undercut any input price w_1 set by $U1$-$D1$, so $U1$-$D1$ will not earn anything on the input market. But $U1$-$D1$ earns $\pi^*_{D1}(0, w_2)$. Due to the raising-rivals' cost effect (see (1)), it follows that $U1$-$D1$ will prefer $U2$ to charge a price as high as possible. In any SGPNE, $U1$-$D1$ best responds to $w_2^{BR}(w_1)$ by charging some $w_1 > w_2^{mon}$ strictly. Any price pair $(w_1, w_2 | w_1 > w_2^{mon}, w_2 = w_2^{BR}(w1))$ constitutes a SGPNE. This implies that in any SGPNE the market will be monopolized through price leadership.

It is equally important to show that price leadership will not have any impact when $U1$ and $D1$ are not integrated. Without integration, the price leader will sell a zero amount. Indeed, there exists a SGPNE in which the price leader charges $w_1 > w_2^{mon}$ and $U2$ marginally undercuts this price, taking home the entire profit. (This is an equilibrium for $U1$-$D1$ because the firm is indifferent as it obtains zero profits for all other w_1 values.) Actually, any outcome with $w_1 \geq 0$ is a SGPNE. However, as also argued by Mouraviev and Rey (2011) the equilibria with $w_1 > 0$ and in particular those with $w_1 > w_2^{mon}$ are rather knife edge and not robust with respect to a potential cost of moving first or second, for example. In extended games like those suggested by Hamilton and Slutsky (1990), a cost of moving first would destroy the equilibria with $w_1 \geq 0$ in our game and firms would choose simultaneously in such extended games. We conclude that price leadership will not have any impact without integration.

Proposition 2. In the sequential-move game with $U1$ as the price leader, the $U1$-$D1$ merger results in monopolization of the input market. In contrast, the input market is competitive without the integration.

D. Design

First, let us clarify the scope of our experiment. It focuses solely on the price-setting behavior of upstream firms. The decision to integrate is not part of the experiment – it is given exogenously. Neither the decisions of downstream firms nor consumer decisions are part of the experiment – they are assumed to behave according to Nash equilibrium. The equilib-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

rium behavior of the downstream firms and consumers determines the payoff tables given to the subjects.

Our treatment design is as follows: There are experimental treatments with and without price leadership, and there are treatments with and without vertical integration. We do not conduct a full 2×2 design because the variant with price leadership but without integration is likely to result in an undesirable payoff for the price leader. Table 1 summarizes the design. The treatment labels read as follows. "Sequent-Integ" refers to the sequential-move variant with vertical integration, we conducted "Simul-Integ-After" subsequently and *within subjects*. "Simul-Integ-After" and "Simul-Integ" are the simultaneous-move variant with vertical integration and "Simul-Separ" is the simultaneous-move variant with vertical separation.

Table 1. Treatments and number of participants

Label	Sequent-Integ	Simul-Integ-After	Simul-Integ	Simul-Separ
order of moves	sequential	simultaneous	simultaneous	simultaneous
integrated?	yes	yes	yes	no
# participants	20		40	40

Our experiments are designed to analyze the effects of price leadership, but we also intend to check for learning effects. Specifically, we will examine whether experience as a price leader results in higher prices when the leader-follower sequence is removed. After playing Sequent-Integ for ten periods in the first phase, subjects immediately play Simul-Integ for 15 periods in the second phase (referred to as Simul-Integ-After). The other treatments involve only one phase.

In all treatments, two subjects representing the two upstream firms have to make one single choice in every period. They set a price. Prices must be integers between one and nine, inclusive.

The treatments differ based on the payoff table that is given to the subjects. These payoff tables are derived from a parametrized model (see Appendix). The tables are fully consistent with the analysis in OSS (1990) who indeed use the same parametrized model for some of their analysis.

In the Simul-Separ treatment, the basis for decision-making is the profit table shown in Table 2. Participants play a simple Bertrand duopoly experiment. The firm that charges the lowest price will earn the profits in the "Profit" cells. The high-price firm receives nothing. In the case of a tie,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

both firms receive half of the profit. The instructions illustrate this with two examples. One example reads, "If you charge a price of 7 and the other firm charges a price of 4, you will get zero, and the other firm will get 81".[4]

Table 2. The payoff table for treatment Simul-Separ (vertical separation)

Price	1	2	3	4	5	6	7	8	9
Profit	39	54	69	81	90	99	90	72	51

In the treatments with vertical integration, Table 3 represents the payoffs. The integrated firm, $U1$-$D1$, is the row player, $U2$ plays columns. $U1$-$D1$

Table 3. Payoffs for $U1$-$D1$ (top) and $U2$ (bottom) for corresponding price pairs (w_1, w_2).

w_2 / w_1	1	2	3	4	5	6	7	8	9
1	85.5, 19.5	105, 0	105, 0	105, 0	105, 0	105, 0	105, 0	105, 0	105, 0
2	66, 39	101, 27	128, 0	128, 0	128, 0	128, 0	128, 0	128, 0	128, 0
3	66, 39	74, 54	118.5, 34.5	153, 0	153, 0	153, 0	153, 0	153, 0	153, 0
4	66, 39	74, 54	84, 69	136.5, 40.5	177, 0	177, 0	177, 0	177, 0	177, 0
5	66, 39	74, 54	84, 69	96, 81	150, 45	195, 0	195, 0	195, 0	195, 0
6	66, 39	74, 54	84, 69	96, 81	105, 90	181.5, 49.5	231, 0	231, 0	231, 0
7	66, 39	74, 54	84, 69	96, 81	105, 90	132, 99	204, 45	249, 0	249, 0
8	66, 39	74, 54	84, 69	96, 81	105, 90	132, 99	159, 90	216, 36	252, 0
9	66, 39	74, 54	84, 69	96, 81	105, 90	132, 99	159, 90	180, 72	223.5, 25.5

4 The experimental markets were designed such that firms still make a positive profit when they both charge 1 (the Nash equilibrium price). The reason is that subjects might be biased against an action with zero profit (Dufwenberg et al., 2007).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

always gets a positive payoff, which corresponds to it's the profit of its downstream affiliate and is affected by the raising-rivals-cost effect: the higher min $\{w_1, w_2\}$, the higher this payoff. Otherwise, firms compete for the profit in the input market which is determined by Bertrand rules, as indicated above.

The stage games were repeated several times for each treatment. Specifically, the Simul-Integ and Simul-Separ treatments were repeated 15 times. The Sequen-Integ treatment was repeated ten times, followed by fifteen repetitions of Simul-Integ After. Subjects knew the number of periods and the final period from the instructions in all treatments.

All treatments were conducted under a random matching scheme. This moves the lab setting closer to the static game and precents reputation building, which is possible in the repeated game.

E. Procedures

All treatments were conducted in sessions with ten participants. Five of the participants acted as "firm 1", the integrated firm, while the remaining five participants acted as "firm 2". These roles remained fixed throughout the experiment. In the Simul-Separ treatment, firms are symmetric but the "firm 1" and "firm 2" labels are still given to maintain a comparable matching scheme and set of instructions.

A total of 100 subjects took part in this experiment. There were four sessions of ten participants each for the Simul-Integ treatment, and four for the Simul-Separ treatment. There were two sessions for the Sequen-Integ and Simul-Integ-After treatments. The lower number of sessions for Sequen-Integ and Simul-Integ-After is due to the extremely low heterogeneity in these treatments.

The experiments were computerized and programmed in Z-Tree, developed by Fischbacher (2007). The experiments were conducted at Royal Holloway, University of London. To avoid exchange rates, the payoffs in Tables 2 and 3 represent cash payments in real currency. Subjects' average monetary earnings were £12.50, including a flat payment of £5. Subjects were primarily undergraduate students, many of whom were from faculties other than economics or business studies.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

F. Results

Tables 4–6 and Figures 2–4 summarize the results. Table 4 reports the average minimum and posted prices and profits for each firm type in the four treatments. Tables 5 and 6 report results from linear probability regressions: Table 5 analyzes the minimum prices w_{min}, and the prices w_1 and w_2. Table 6 focuses on the probability of monopolization ($w_{min} = 6$), foreclosure strictly defined (probability of $w_1 > 6$), and foreclosure broadly defined (probability of $w_1 > w_2$).

Table 4. Descriptive Statistics

Treatment	Sequent-Integ	Simul-Integ-After	Simul-Integ	Simul-Separ
w_{min}	5.27	2.61	2.33	1.53
	(1.55)	(1.38)	(1.26)	(0.76)
w_1	7.01	3.70	3.36	2.04
	(2.10)	(2.05)	(2.11)	(1.21)
w_2	5.31	3.19	2.90	2.22
	(1.50)	(1.57)	(1.55)	(1.69)
π_1	62.25	54.43	103.44	22.69
	(13.67)	(16.85)	(32.32)	(19.24)
π_2	43.14	17.91	32.74	24.13
	(11.76)	(14.90)	(28.07)	(19.84)

I. Average prices and profits

Price leadership in Sequent-Integ effectively monopolizes the upstream market. The minimum price w_{min} Sequent-Integ does not differ much from the monopoly price, on average it equals 5.27 (the upstream monopoly price is 6). The integrated firm chooses on average a price above monopoly level (7.01) and the non-integrated firm sets 5.31 (see Table 4). As is evident also from Figure 2, the minimum price w_{min} virtually identical to w_2, the price of the non-integrated firm, suggesting that $U2$ indeed best responds.

The effect of price leadership in Sequent-Integ does not carry over to the second phase with simultaneous price choices (Simul-Integ-After), though. When price leadership is removed, the price level drops strongly

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

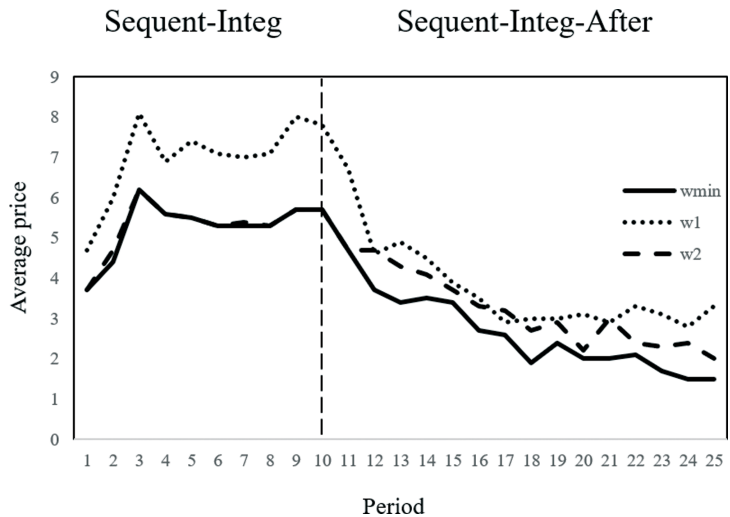


Figure 2. Sequential-Integ and Sequential-Integ-After treatments

(see the right part of Figure 2). The minimum price is less than half as high, we obtain a mean price of 2.61 in Simul-Integ-after. Comparing w_{min}, w_1 and w_2, the treatments Sequent-Integ and Simul-Integ-After differ significantly at the 1 % level (see Table 5, regressions (1), (2) and (3) and the corresponding Wald tests for coefficient differences).

Price leadership is profitable for both firms, but more so for the non-integrated firm. The profit π_2 is more than twice as high in the Sequent-Integ treatment compared to the Simul-Integ-After treatment. The difference is not as pronounced but still substantially positive for the integrated firm. It earns on average 62.25 in Sequent-Integ and 54.43 in Simul-Integ-After (see Table 4).

The experience of price leadership in the Sequent-Integ treatment is not even strong enough to have an impact compared to the treatment with simultaneous moves. The Simul-Integ data are illustrated in Figure 3. When we compare the treatment Simul-Integ-After with the results of treatment Simul-Integ, we do not find significant differences. The minimum price is very similar (2.61 in Simul-Integ-After and 2.33 in Simul-Integ) and average price choices of both firms are only slightly higher in Simul-Integ-After (see also Figures 2 and 3). If we exclude periods 1–3 in Simul-Integ-After and Simul-Integ, the difference gets even smaller. Averages are 2.28 for Simul-Integ-After and 2.25 and the difference between

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 5. Linear regression on w_{min}, w_1 and w_2

	(1)	(2)	(3)
	w_{min}	w_1	w_2
Sequent-Integ	5.49***	7.33***	5.50***
	(0.13)	(0.03)	(0.14)
Simul-Integ-After	2.28***	3.28***	2.85***
	(0.03)	(0.23)	(0.01)
Simul-Integ	2.25***	3.12***	2.78***
	(0.36)	(0.35)	(0.37)
Simul-Separ	1.37***	1.86***	1.95***
	(0.15)	(0.19)	(0.19)
Periods 1–3 × Sequent-Integ	-0.72	-1.06***	-0.63
	(0.60)	(0.22)	(0.69)
Periods 1–3 × Simul-Integ-After	1.66***	2.12***	1.72***
	(0.02)	(0.18)	(0.09)
Periods 1–3 × Simul-Integ	0.44*	1.23***	0.62**
	(0.23)	(0.37)	(0.24)
Periods 1–3 × Simul-Separ	0.78***	0.89***	1.35***
	(0.14)	(0.17)	(0.10)
Obs.	850	850	850
R^2	0.86	0.82	0.81
Coefficient differences			
Sequent-Integ – Simul-Integ-After	3.21***	4.05***	2.65***
Simul-Integ-After – Simul-Integ	0.03	0.16	0.07
Simul-Integ – Simul-Separ	0.88*	1.26**	0.83*

Notes: Each column represents an OLS regression with the dependent variables: minimum price w_{min}, prices w_1 and w_2. We regress on treatment dummies Sequent-Integ, Simul-Integ-after, Simul-Integ, and Simul-Separ as well as treatment specific time trends, that is, interaction terms of a dummy for periods 1–3 and treatment variables. We omit the constant and cluster standard errors at session level. A Wald test on the coefficient differences is performed and the significance level is indicated by the stars, in particular *** 1 % level significance, ** 5 % level significance and * 10 % level significance.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

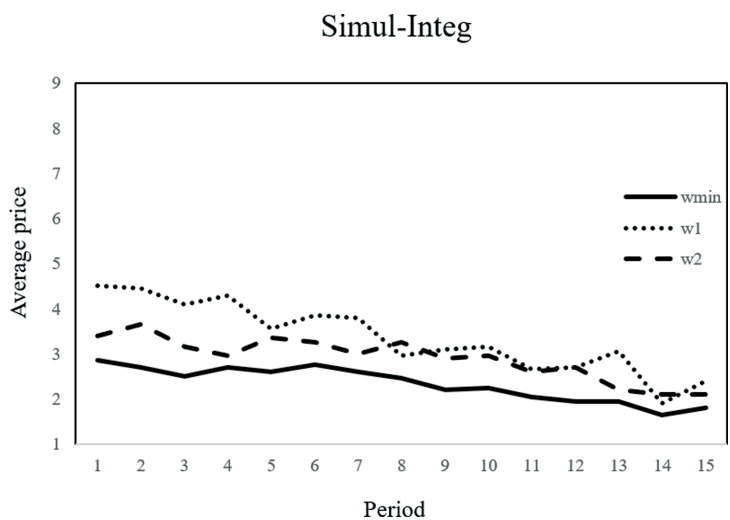


Figure 3. Simul-Integ treatment

the treatments is insignificant (see Table 5, regressions (1), (2) and (3) and the corresponding Wald tests for coefficient differences).

However, vertical integration still has an effect. We find that treatment Simul-Integ increases the minimum price compared to Simul-Separ. The average minimum price equals 2.33 with integration while it is 1.53 without. Regression (1) in Table 5 confirms that the difference is significant at the 10 % level. Price choices of both firms are significantly larger with integration (see also Figures 3 and 4) and the non-integrated firm benefits from the integration of its competitor in Simul-Integ (see Table 3).

II. Foreclosure and monopolization

We now turn to the issues of foreclosure and monopolization of the market. To some degree, the results on the average prices already indicate the extent of foreclosure and monopolization. It will, however, be useful to discuss these issues separately. We analyze monopolization defined as $w_{min} = 6$, foreclosure strictly defined ($w_1 > 6$), and foreclosure broadly defined (probability of $w_1 > w_2$).

We begin with *monopolization* which is OSS's (1990) prediction. For the Sequent-Integ treatment after period three (Table 6, regression (4)),

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

we can see that the probability of monopolization is 69%. The probability of monopolization then decreases substantially: After the first three periods in Simul-Integ-after we do not observe any monopolization (see Table 6, regression (4)). Monopolization is also practically zero with simultaneous moves. In the Simul-Integ treatment, only 0.67% of markets are monopolized.[5] Thus, in contrast to OSS's (1990) predictions, vertical integration alone does not lead to monopolization of the upstream market. In Simul-Separ, foreclosure never results in monopolization.

We now turn to the *strictly defined concept of foreclosure*, which is the withdrawal of an integrated firm from the upstream market. In strict foreclosure, the integrated firm sets a price above the monopoly level – that is, a price greater than six.[6] If the integrated firm moves first (Sequent-Integ), it forecloses (in the sense of $w_1 > 6$) after period 3 in 76% of observations (see Table 6). However, in Simul-Integ-After the integrated firm forecloses only in 4% of observations. That is, the foreclosure in the first phase does not have a long-term impact on pricing in the second phase. Foreclosure in Simul-Integ-After is insignificant and even slightly below the percent-

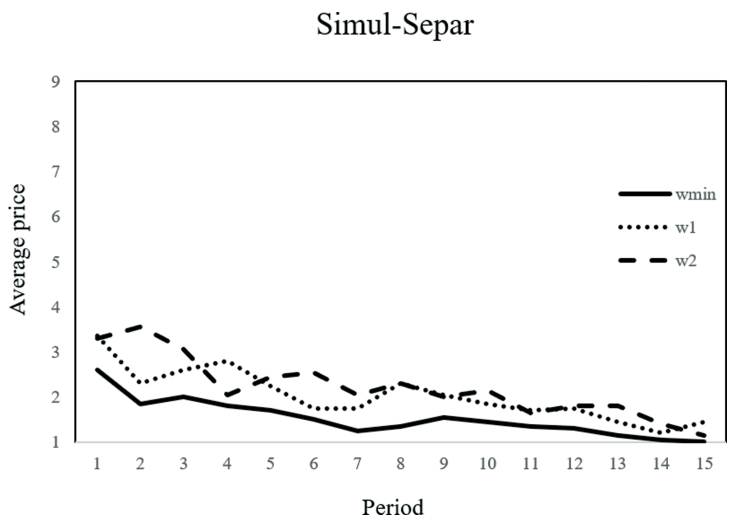


Figure 4. Simul-Separ treatment

5 In Simul-Integ, the foreclosure rate is largely influenced by a single participant who selected it in 12 out of 15 periods.

6 Monopolization occurs when there is strict foreclosure and, in addition, the non-integrated firm charges the monopoly price.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 6. Linear probability regressions for monopolization and foreclosure

	(4)	(5)	(6)
	$w_{min} = 6$	$w_1 > 6$	$w_1 > w_2$
Sequent-Integ	0.69***	0.76***	0.93***
	(0.02)	(0.01)	(0.01)
Simul-Integ-After	0.00	0.04*	0.43***
	(0.00)	(0.02)	(0.04)
Simul-Integ	0.01	0.08*	0.42***
	(0.01)	(0.03)	(0.01)
Simul-Separ	0.00	0.00	0.30***
	(0.00)	(0.00)	(0.02)
Periods 1–3 × Sequent-Integ	–0.29***	–0.22***	–0.06
	(0.02)	(0.04)	(0.06)
Periods 1–3 × Simul-Integ-After	0.27***	0.22***	0.10
	(0.05)	(0.02)	(0.09)
Periods 1–3 × Simul-Integ	0.00	0.11*	0.13**
	(0.01)	(0.06)	(0.05)
Periods 1–3 × Simul-Separ	0.00	0.00	0.08*
	(0.00)	(0.00)	(0.04)
Obs.	850	850	850
R^2	0.55	0.26	0.53
Coefficient differences			
Sequent-Integ – Simul-Integ-after	0.69***	0.72***	0.50***
Simul-Integ-after – Simul-Integ	–0.01	–0.03	0.01
Simul-Integ – Simul-Separ	0.01	0.07**	0.12***

Notes: Each column represents a linear probability model with the dependent variables: dummies on monopolization ($w_{min} = 6$), foreclosure strictly defined ($w_1 > 6$), and foreclosure broadly defined ($w_1 > w_2$). We regress on treatment dummies Sequent-Integ, Simul-Integ-after, Simul-Integ, and Simul-Separ as well as treatment specific time trends, that is, interaction terms of a dummy for periods 1–3 and treatment variables. We omit the constant and cluster standard errors at session level. A Wald test on the coefficient differences is performed and the significance level is indicated by the stars, in particular *** 1% level significance, ** 5% level significance and * 10% level significance.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

age of foreclosure in Simul-Integ. The integrated firm withdraws in 8 % of markets in Simul-Integ (after period three). In Simul-Separ there is no incentive for foreclosure, and there is virtually no foreclosure. The difference between Simul-Integ and Simul-Separ is significant at the 5 % level (see Table 6 regression (5)).

If we look more *broadly at foreclosure strategies*, a similar picture emerges. We consider prices of the integrated firm that are larger than the prices of the non-integrated firm ($w_1 > w_2$). When prices are chosen sequentially, the integrated firm chooses a larger price in 93 % of markets. In Simul-Integ-After the percentage drops by roughly 50 %, the integrated firm sets a higher price in 43 % of markets. In Simul-Integ 42 % of integrated firms choose a higher price than their opponent which is almost identical. Again, we do not observe any effect from Sequent-Integ on Simul-Integ-After (see Table 6, regression (6)). With no vertical integration in Simul-Separ, firm $U1$ sets a higher price in 30 % of markets. Considering the difference between Simul-Integ and Simul-Separ from period four, it is significant at the 1 % level (see Table 6, regression (6)).[7]

III. Effects over Time

We analyze the effects over time by adding dummy variables for periods 1 through 3 to the regressions in Tables 5 and 6. In periods 1 to 3 of Sequent-Integ, we observe a slight upward trend of prices followed by a stable price level over later periods. The difference is between periods 1 to 3 and the following periods is highly significant for w_1, however, it is insignificant for w_{min} and w_2 (see Figure 2, Table 5, regressions (1) to (3)). In the first period of Sequent-Integ, average prices are similar to the treatment Simul-Integ. But in period 3, we observe an average price of 8.1 of the integrated firm and 6.2 of the opponent. In contrast, the overall price level slightly decreases after the first three periods of Simul-Integ. The difference is significant for w_{min} at the 10 % level), w_1 (at the 1 % level), and w_2 (at the 5 % level). Overall, there seems to be an upward trend of prices in Sequent-Integ and a downward trend of prices in Simul-Integ.

7 Note that the fraction of identical prices in markets with integration decreases while $U1$ undercuts the opponent just as often with and without integration. The integrated firm undercuts the competitor in 31.33 % of markets in Simul-Integ-after and 32 % in Simul-Integ. Firm $U1$ chooses a lower price than $U2$ in 29 % of markets in Simul-Separ.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The price level decreases exceptionally in Simul-Integ-after (see Figure 2). While in the first period of Simul-Integ-after (period 11) the integrated firm sets on average a price of 6.7, it drops already in period 12 to an average of 4.6. The difference between periods 1–3 and the following periods is highly significant for w_{min}, w_1, and w_2 (see Table 5, regressions (1) – (3)). As conclude before, that is, the stable price level in Sequent-Integ does not carry over to the phase with simultaneous moves.

Although we find an overall downward trend in Simul-Integ, the observed price difference between Simul-Integ and Simul-Separ holds also with more periods. Normann (2011) shows in a robustness check, that the difference persists if the game is repeated 25 instead of 15 times.

G. Conclusion

One key argument against vertical mergers is that they lead to higher prices in the input market through a raising-rival's-cost effect (OSS, 1990). Whereas Hart and Tirole (1990) and Reiffen (1992) argue that this argument is flawed because it requires commitment, OSS (1992) counter that "*[t]he notion that vertically integrated firms behave differently from nonintegrated ones in supplying inputs to downstream rivals would strike a business person, if not an economist, as common sense*". Who is right and who is wrong? Our paper shows and provides evidence for both diametrically opposed positions.

First, we argue in favor of the OSS (1990) model by suggesting that price leadership can function as a commitment device for integrated firms. Here, we build upon the existing literature, which includes the work of Choi and Yi (2000), Church and Gandal (2000), and OSS (1992), who demonstrate that commitment is effective in extended versions of the OSS (1990) model. Our argument is broader in that we claim the integrated firm will strive to become the price leader, as in a model of endogenous timing (Hamilton and Slutsky, 1990; Normann, 2002). By contrast, a non-integrated firm would never want to be a price leader.

Our experimental evidence shows that price leadership indeed leads to monopolization, but it also shows that learning effects from experiences a phase of price leadership are limited. Once they are no longer the price leader, integrated firms do not charge higher prices than non-integrated firms, nor are prices higher than in simultaneous-move experiments. This supports the findings of Hart and Tirole (1990) and Reiffen (1992) that,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

in the absence of price leadership, the commitment problem is severe and vertical mergers do not have anticompetitive effects. The Nash equilibrium is strong – it seems that players cannot learn to achieve non-Nash outcomes in this case.

References

Allain, Marie-Laure, Claire Chambolle, and Patrick Rey (2011) "Vertical Integration, Information and Foreclosure," *TSE Working Paper, 11–237*.

Allain, Marie-Laure, Claire Chambolle, and Patrick Rey (2016) "Vertical Integration as a Source of Hold-up," *The Review of Economic Studies*, 83(1): 1–25.

Allain, Marie-Laure, Claire Chambolle, Patrick Rey and Sabrina Teyssier (2021) "Vertical Integration as a Source of Hold-up: an Experiment," *European Economic Review,* 137, 103783.

Bonanno, Giacomo, and John Vickers (1988) "Vertical Separation," *The Journal of Industrial Economics*, 36(3): 257–265.

Chen, Yongmin (2001) "On Vertical Mergers and Their Competitive Effects," *RAND Journal of Economics*, 32: 466–496.

Chen, Yongmin, Michael H. Riordan (2007). "Vertical Integration, Exclusive Dealing and Ex Post Cartelization." *RAND Journal of Economics* 38, 1–21.

Choi, Jay Pil, and Sang-Seung Yi (2000) "Vertical Foreclosure with the Choice of Input Specifications," *RAND Journal of Economics*, 717–743.

Church, Jeffrey and Neil Gandal (2000) "System Competition, Vertical Merger and Foreclosure," *Journal of Economics & Management Strategy*, 9(1): 25–51.

Durham, Yvonne (2000) "An Experimental Examination of Double Marginalization and Vertical Relationships," *Journal of Economic Behavior and Organization*, 42(2): 207–229.

Dufwenberg, Martin, Uri Gneezy, Jacob K. Goeree, and Rosemarie Nagel (2007) "Price Floors and Competition," *Economic Theory*, 33(1): 211.

Fischbacher, Urs (2007) "ZTree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 10: 171–178.

Greiner, Ben (2015) "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 1(1): 114–125.

Hamilton, Jonathan H., and Steven M. Slutsky (1990) "Endogenous Timing in Duopoly Games: Stackelberg or Cournot equilibria," *Games and Economic Behavior*, 2(1): 29–46.

Harrison, Glenn W. (1989) "Theory and Misbehavior of First-Price Auctions," *The American Economic Review*, 79: 749–762.

Hart, Oliver and Jean Tirole (1990) "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity (Microeconomics)*, 205–276.

Hunold, Matthias and Vasilisa Werner (2025) "Foreclosure and Tunneling with Partial Vertical Ownership", Journal of Economics and Management Strategy, 2025.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Levy, Nadav, Yossi Spiegel, and David Gilo (2018) "Partial Vertical Integration, Ownership Structure, and Foreclosure," *American Economic Journal: Microeconomics* 10.1, 132–180.

Martin, Stephen, Hans-Theo Normann and Christopher M. Snyder (2001) "Vertical Foreclosure in Experimental Markets," *Rand Journal of Economics*, 32: 466–496.

Mason, Charles F. and Owen R. Phillips (2000) "Vertical Integration and Collusive Incentives: an Experimental Analysis" *International Journal of Industrial Organization*, 18: 471–496.

Moellers, Claudia (2016) "Reputation and Foreclosure with Vertical Integration – Experimental Evidence," *mimeo*.

Moellers, Claudia, Hans-Theo Normann and Christopher M. Snyder (2017) "Communication in Vertical Markets: Experimental Evidence," *International Journal of Industrial Organization*, 50: 214–258.

Mouraviev, Igor and Patrick Rey (2011) "Collusion and Leadership," *International Journal of Industrial Organization*, 29: 705–717.

Nocke, Volker and Lucy White (2007) "Do Vertical Mergers Facilitate Upstream Collusion?" *American Economic Review*, 97: 1321–1339.

Normann, Hans-Theo (2002) "Endogenous Timing with Incomplete Information and with Observable Delay", *Games and Economic Behavior*, 39: 282–291.

Normann, Hans-Theo (2009) "Vertical Integration, Raising Rivals' Costs and Upstream Collusion," *European Economic Review*, 53(4): 461–480.

Normann, Hans-Theo (2011) "Vertical Mergers, Foreclosure and Raising Rivals' Costs – Experimental Evidence," *The Journal of Industrial Economics*, 59: 506 –527.

Ordover, Janusz A., Garth Saloner and Steven C. Salop (1990) "Equilibrium Vertical Foreclosure," *American Economic Review*, 80: 127–142.

Ordover, Janusz A., Garth Saloner and Steven C. Salop (1992) "Equilibrium Vertical Foreclosure: Reply," *American Economic Review*, 82: 698–703.

Reiffen, David (1992) "Equilibrium Vertical Foreclosure: Comment," *American Economic Review*, 82: 694–697.

Rey, Patrick and Jean Tirole (2007) "A Primer on Foreclosure," *M. Armstrong, Porter, R. (Eds.), Handbook of Industrial Organization*, North Holland, 3: 2145–2220.

Riordan, Michael H. (1998). "Anticompetitive Vertical Integration by a Dominant firm." *American Economic Review* 88: 1232–48.

Salinger, Michael A. (1988) "Vertical Mergers and Market Foreclosure," *The Quarterly Journal of Economics*, 103: 345–356.

Spiegel, Y. (2013). "Backward Integration, Forward Integration, and Vertical Foreclosure," CEPR Discussion Paper.

Appendix A: Linear Model

This appendix presents the model underlying the payoff tables of the experiment. See OSS (1990) and Normann (2011) for similar linear

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

parametrizations. As in the general model, there are two upstream firms ($U1$ and $U2$) and two downstream firms ($D1$ and $D2$).

Downstream firm Di's demand is given by the linear demand function

$$q_i\left(p_i, p_j\right) = a - bp_i + dp_j; i, j = 1, 2; i \neq j,$$

where p_i and p_j are the prices the downstream firms i and j charge ($i, j = 1, 2; i \neq j$). Assume downstream firm i purchases the input good at a linear price of c_i per unit. As the D firms incur no other costs, they produce at constant marginal costs of c_1 and c_2, respectively. It follows that the downstream level is a standard model with product differentiation and asymmetric cost. We solve for downstream Nash equilibrium prices

$$p_i^*\left(c_i, c_j\right) = \frac{\left(2b + d\right)a + 2b^2 c_i + bdc_j}{4b^2 - d^2},$$

outputs

$$q_i^*\left(c_i, c_j\right) = b\frac{\left(2b + d\right)a - \left(2b^2 - d^2\right)c_i + bdc_j}{4b^2 - d^2},$$

and profits $\pi_{Di}^* = \frac{\left(q_i^*\left(c_i, c_j\right)\right)^2}{b}$.

Upstream firms have constant marginal cost which are assumed to be zero as in the general model. The upstream firms compete for each of the two downstream markets in a Bertrand fashion. Employing slightly different notation than above, we assume upstream firm k sets two prices, c_1^{Uk} and c_2^{Uk}, for downstream firms 1 and 2, respectively. Bertrand competition implies that the downstream firm i buys from the upstream firm with the lowest price, formally $c_i = \min\left\{c_i^{U1}, c_i^{U2}\right\}, i = 1, 2$. Put it another way, an upstream firm will sell a positive amount to Di only if it charges the lowest price. Formally, when upstream firm k bids c_i^{Uk} to downstream firm i, it will make the following profit with Di

$$\pi_i^{Uk}\left(c_i^{Uk}, c_i^{Ul}\right) = \begin{cases} c_i^{Uk} \cdot q_i^*\left(c_i^{Uk}, c_i^{Uk}\right) & if\, c_i^{Uk} < c_i^{Ul} \\ c_i^{Uk} \cdot q_i^*\left(c_i^{Uk}, c_i^{Uk}\right)/2 & if\, c_i^{Uk} = c_i^{Ul} \\ 0 & if\, c_i^{Uk} > c_i^{Ul} \end{cases}$$

where $k, l = 1, 2; k \neq l; i = 1, 2$.

When neither firm is integrated, in the general model, $U1$ and $U2$ set prices $\left(c_1^{U1}, c_2^{U1}\right)$ and $\left(c_1^{U2}, c_2^{U2}\right)$, respectively. For the derivation of the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

payoff tables, c_1 is set equal to zero. The reason is that $c_1 = 0$ with vertical integration. Thus, in order to keep treatments comparable and avoid wealth effects, one also needs $c_1 = 0$ without integration. Essentially, this is implies that firms only compete for $D2$ also absent integration. This is without loss of generality of the qualitative features of Bertrand competition are unaffected by this. $D2$ buys at the lower of the two prices such that $c_2 = \min\left\{c_2^{U1}, c_2^{U2}\right\}$. Next, given $c_1 = 0$ and c_2, $D1$ and $D2$ set the final good prices. In equilibrium, $D1$ and $D2$ charge $p_1^*(0, c_2)$ and $p_2^*(c_{2,0})$, respectively. Downstream profits are $\pi_{D1}^* = \frac{(q_1^*(0,c_2))^2}{b}$ and $\pi_{D2}^* = \frac{(q_2^*(c_{2,0}))^2}{b}$, and upstream profits are π_2^{U1}, π_2^{U2} and $\pi_1^{Uk} = 0$.

A vertical merger of $U1$ and $D1$ implies that the integrated firm's true input price is $U1$'s marginal cost (Bonanno and Vickers, 1988). Hence, $D1$ will be delivered internally and efficiently at $c_1 = 0$ and $U2$ cannot compete for the $D1$ business any more. For both upstream firms, only the $D2$ market remains a source for potential business. Profits are as follows. $D2$ earns $\pi_{D2}^* = \frac{(q_2^*(c_{2,0}))^2}{b}$, $U2$ earns π_2^{U2}, and the integrated firm $U1$-$D1$ makes a profit of $\pi_2^{U1} + \pi_{D1}^* = \pi_2^{U1} + \frac{(q_1^*(0,c_2))^2}{b}$.

Tables 2 and 3 can be derived from these closed-form solutions for the parameters $a = \frac{35}{2}$, $b = 4$, $d = 2$. The actual price parameters used to derive the profits in the table differ from the prices labels "1" to "9". In particular, profits around the joint-profit maximizing prices are quite flat. Hence, prices were increased in steps larger than one in this range to avoid the "flat-maximum" critique (Harrison, 1989). The actual price parameters underlying the values in the table are {1.1, 1.6, 2.2, 2.9, 3.5, 5.0, 6.5, 7.6, 8.5}. Additionally, profits were multiplied by three and rounded to yield the payoff in real currency subjects received.

Appendix B: Instructions

Below are the instructions for Sequen-Integ, phase 1. These include phase 1, phase 2 and the additional oral instructions.

*

This is an experiment on market decision-making. Funds for this experiment have been provided by an external research foundation. Take the time to read carefully the instructions. A good understanding of the instructions and well thought out decisions during the experiment can earn

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

you a considerable amount of money. All earnings from the experiment will be paid to you in cash at the end of the experiment.

YOUR ROLE AND TASK IN THE EXPERIMENT

There are a total of 10 participants in this experiment (you and 9 others). Each participant will represent a firm. There are two types of firms, firm 1 and firm 2. The computer randomly assigns five participants the role of firm 1 and the other five participants the role of firm 2. Your role as firm 1 or firm 2 will remain fixed throughout the experiment, and you will learn whether you are firm 1 or firm 2 before we begin the experiment.

The experiment takes place over 25 rounds. In each round, a firm 1 and a firm 2 will meet in a market for a fictitious commodity, called market A. Firm 1 operates also in market B but firm 2 does not. The computer will randomly match the firm 1-firm 2 pairs in every round. The matching is completely random, meaning that there is no relation between the participant you have been matched with last round (or any other previous round) and the participant to whom you will be assigned this round.

Your task is the same in every round, no matter whether you are firm 1 or firm 2. You have to decide on a price, and this single price is valid in both markets A and B. The price can be any (whole) number from 1 to 9.

Importantly, firm 1 will have to choose its price first. Firm 2 is informed about firm 1?s choice, and then firm 2 chooses its price.

The profit you can make by charging the price is as follows.

PROFIT CALCULATION

In market A, the firm that charges the lowest price will receive the profit corresponding to this price in the following table. The firm with the higher price gets zero profit in that round, and, if both firms set the same price, they share the profit equally. Price profit in market A in pence profit in market B in pence

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Price 1	2	3	4	5	6	7	8	9
Profit in Market A 39	54	69	81	90	99	90	72	51
Profit in Market B 66	74	84	96	105	132	159	180	198

Consider two examples for market A. If you charge a price of 7 and the other firm charges a price of 4, you will get zero and the other firm gets 41 pence in market A. Or, if both firms charge a price of 3, both will get 35 /2 = 17.5 pence in market A.

In market B, firm 1 only receives a profit – the profit of the "market B" column in the table. As in market A, it is the lowest of the two prices which determines the profit, no matter whether firm 1 or firm 2 (or both) charged the lowest price.

Consider an example for market B. If firm 1 charges a price of 7 and firm 2 charges a price of 4, firm 1 only gets 48 pence in market B.

Taking both markets into account, firm 1 receives the profit it made in market A plus the profit it made in market B. Firm 2 does not get any profit in market B, only in market A.

TWO PHASES

There are two phases in this experiment. The first phase lasts 10 periods, the second phase lasts 15 periods. So, in total, the experiment will go over 25 periods. After phase 1 is over, we will inform you about a change in the market relevant for the second phase.

EACH ROUND

A round ends when all firms have chosen their price. At the end of the round, each firm sees its own price, the price of the other firm and the own profit from the round.

At the completion of 25 rounds, you will be paid your earnings in pounds you have accumulated during the experiment. In addition to these earnings, each participant will receive a payment of 3£. While the earnings are being counted for distribution, you will be asked to complete a questionnaire related to the experiment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

QUESTIONS?

If you have any questions about the instructions, please raise your hand and an experimenter will come to assist you. Thank you for your participation.

*

The instructions for phase 2 merely emphasized that the same game would be played but that firm 1 would no longer move first. Instead, both firms would move simultaneously. The second phase lasted for 15 periods.

*

The following summary of the instructions was read aloud at the beginning of the treatment and after participants had finished reading the instructions.

TASK

- Your task is the same in every round: You have to decide on a price.
- The price can by any (whole) number from 1 to 9.
- You set only one price every period
- There are two phases in this experiment. The first phase lasts 10 periods, the second phase lasts 15 periods. After phase 1 is over, we will inform you about a change in the market relevant for the second phase.

PROFIT

As you read in the instructions, there are two markets, market A and market B. The profit you can make by charging the price is as follows.

- In market A, the firm which charges the lower price will receive the profit in the table.
- The firm with the higher price gets zero profit in that round
- and if both firms set the same price, they share the profit equally.
- In market B, firm 1 only receives a profit
- As in market A, it is the lowest of the two prices which determines the profit, no matter whether firm 1 or firm 2 (or both) charged the lowest price
- Again, firm 1 receives the profit it made both in market A and B.
- Firm 2 gets a profit only in market A.]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

MATCHING

- There are two types of firms, firm 1 and firm 2.
- The computer will randomly match a firm 1 with a firm 2 pairs in every round.
- In the beginning, the computer randomly assigns five participants the role of firm 1 and the other five participants the role of firm 2.
- Your role as firm 1 or firm 2 will remain fixed throughout the experiment
- Firm 1 has to choose its price first. Firm 2 is informed about firm 1?s choice, and then firm 2 chooses its price.

ROUNDS

- Phase 1 takes place over 15 rounds.
- At the end of phase 2, you will be paid your earnings in pence
- In addition, you will receive a payment of 5

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Institutions, Experiments, and Tax Morale
An Integrated Perspective on Tax Compliance

Azzura Morreale & Luigi Mittone

This chapter outlines an institutional–behavioral account of tax compliance that extends beyond classic deterrence. Guided by Christoph Engel's institutional lens – and using enforcement-free laboratory experiments co-designed with him – we show that people pay taxes even without audits or sanctions evidencing rule-following. Compliance, however, declines when income heterogeneity was made salient and when high-income taxpayers were expected to evade. A second set of experiments isolates legitimacy effects: compliance rises when revenues fund broadly acceptable public purposes (output legitimacy) and when taxpayers can help decide allocations (input legitimacy), with participation valued intrinsically even when outcomes were not individually optimal. Taken together, these results identify the institutional foundations of tax morale and point to policy: enforcement is necessary but insufficient without fairness and legitimacy. The chapter concludes with a reflection on the insights gained through collaboration with Christoph Engel, whose interdisciplinary vision and commitment to experimental rigor were instrumental to the success of this work.

A. Introduction

Institutions shape tax behavior along two channels. The first is deterrence – the formal obligations, audits, and sanctions emphasized by classic deterrence models (Becker, 1968; Allingham & Sandmo, 1972). The second is legitimacy: citizens' judgments about whether the system is acceptable in what it funds (output legitimacy) and in how choices are made (input legitimacy) (Easton, 2017; Scharpf, 1999). Deterrence alone, however, has long struggled to explain the high levels of real-world compliance, pointing to the role of tax morale – intrinsic motivation, fairness norms, trust, and social expectations (Luttmer & Singhal, 2014; Torgler, 2007). Consistent with this view, extensive research shows that tax compliance emerges from the interaction of behavioral and institutional factors

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

(see e.g., Alm, McClelland & Schulze, 1999; Kirchler, Hoelzl & Wahl, 2008, Cummings et al., 2009; Casal & Mittone, 2016; Bott et al. 2020).

This chapter advances an integrated institutional–behavioral perspective developed in close collaboration with Christoph Engel. Guided by Engel's institutional lens, we designed enforcement-free experiments that strip away audits and fines to isolate two mechanisms. Section 2 presents evidence that individuals comply with tax regulations even when there is no enforcement, although this compliance can decrease when differences in income are highlighted and when doubts about the honesty of others, particularly among higher-income groups, are introduced (Engel, Mittone & Morreale, 2020). Section 3 tests legitimacy more directly by varying both what taxes fund (outcomes) and how allocations are decided (procedures). We find that acceptable uses of revenue and opportunities for voice each raise compliance, even when outcomes are not individually optimal (Engel, Mittone & Morreale, 2024). Section 4 synthesizes these results to locate the institutional foundations of tax morale, and Section 5 draws implications for policy and research.

Christoph Engel's contribution has been foundational to this project. Throughout his career, he has bridged disciplines that often remain apart – law, economics, psychology, and political theory – and his role has been pivotal in translating normative concerns into testable designs, bridging law's structural reasoning with economics' experimental method. The upshot is simple but far-reaching: the same formal tax rule can elicit different behaviors depending on how it aligns with fairness norms and legitimacy beliefs.

B. Rule-Following and Fairness

Why do individuals pay taxes when evasion is free? We run an experiment where participants had to pay taxes with no chance of enforcement: no audits, no fines, and no sanctions whatsoever (Engel, Mittone & Morreale, 2020). If the classical deterrence model were correct, compliance should have dropped to zero. Instead, participants paid significant amounts of taxes.

The result confirms that rule-following is a genuine motivation: people often comply simply because a rule exists. The result aligns with norm-based accounts (Guala & Mittone, 2010) and evidence from other areas showing that people follow norms even when there are no material incentives involved (Kimbrough & Vostroknutov, 2016). In the context of

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

taxation, this insight suggests that deterrence is not the only reason people pay their taxes.

The same experiment also shed light on the interaction between rule-following and fairness concerns. After several periods of play, participants were told that incomes were heterogeneous: some had low endowments, others medium, and others high. Compliance immediately fell by about 16 percent, driven mainly by low-income participants. Importantly, this drop occurred under uniform, proportional, and progressive tax regimes alike. What mattered was not the specific regime but the salience of heterogeneity itself.

Beliefs about others' compliance further shaped behavior. Participants' willingness to pay was strongly correlated with their expectations about the honesty of high-income individuals. When the wealthy were perceived as likely to evade, tax morale weakened across all income groups. This finding mirrors field evidence showing that perceptions of elite noncompliance undermine tax morale in the broader population.

The methodological innovation of the study was its careful design: no enforcement, no redistribution, and no confounding incentives. This allowed us to isolate the psychological mechanisms at work. The findings indicate that rule-following does occur, yet it can be affected by fairness concerns and skepticism about others' behavior.

C. Legitimacy: Outcomes and Participation

A second line of experiments examined legitimacy more directly, asking whether taxpayers comply more when they approve of the uses of tax revenues (outcome legitimacy) or when they can participate in deciding how revenues are used (procedural legitimacy) (Engel, Mittone & Morreale, 2024).

Participants completed a real-effort task to earn income, then faced an unenforced tax (no audits). Taxes were allocated to real-world public entities – such as hospitals, fire brigades, and the Red Cross – under four different regimes: Imposed (random recipient), Winner-Takes-All (the recipient with the most approvals gets all funds), Proportional (funds split in proportion to approvals), and Gold–Silver–Bronze (funds split 1/2, 1/3, 1/6 among the top three overall recipients).

The results were clear. *Outcome legitimacy* mattered: the more entities a participant found acceptable, the more tax they paid. But *procedural*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

legitimacy mattered as well: participants paid more taxes when they had a voice in the allocation process, and they were even willing to pay to avoid random allocation. Interestingly, participation was valued not only instrumentally – because it increased the chance of securing preferred outcomes – but also intrinsically: even when outcomes were not individually optimal, the act of participation itself legitimized the process.

Two implications follow. First, citizens care about what their taxes fund: compliance rises when revenues are channeled toward acceptable purposes. Second, they care about how decisions are made: giving people a voice increases compliance even when the outcome is not their preferred one. This aligns with evidence that direct-democratic institutions are associated with higher tax morale (Torgler, 2005) and with political-science distinctions between output and input legitimacy (Easton, 2017; Scharpf, 1999), here demonstrated with measurable behavior in the laboratory.

D. The Institutional Foundations of Tax Morale

Taken together, the experiments on rule-following and on legitimacy illuminate the institutional foundations of tax morale. Compliance is not merely an individual choice; it emerges from the interaction of psychological motives with institutional design.

Rule-following supplies a baseline willingness to comply even without enforcement. Yet this motive is conditional: when inequality becomes salient or doubts arise about others' honesty – especially among the wealthy – morale declines. Fairness norms thus condition rule-following, and beliefs about others' compliance become pivotal.

Legitimacy then acts as an amplifier. Compliance rises when taxes finance acceptable outcomes and when taxpayers have a voice in allocating revenues. Crucially, outcomes and procedures each matter on their own: what taxes fund and how choices are made jointly strengthen morale.

Together, the studies by Engel, Mittone, and Morreale (2020; 2024) are pivotal in isolating these dynamics and in showing how carefully structured experiments can evaluate them with rigor.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

E. Conclusion

The study of tax compliance has evolved far beyond the deterrence model. Evidence now shows that compliance is shaped by the interplay of rule-following, fairness, and legitimacy. These forces are not substitutes but complements: together they sustain the voluntary willingness of citizens to finance collective goods.

For policymakers, the message is clear. Enforcement remains necessary but cannot stand alone. Tax systems must also be perceived as fair, transparent, and legitimate. Citizens need to see that revenues are used for valuable purposes, that procedures are inclusive, and that others – especially those with high incomes – are also paying their share. Only then can compliance be durable and widespread.

For researchers, the lesson is equally clear. Experimental economics, when combined with institutional theory, offers powerful tools to uncover the subtle mechanisms that drive compliance. By stripping away confounds and carefully manipulating institutional variables, experiments make it possible to test theories that would otherwise remain abstract.

Throughout this journey, we have been privileged to collaborate with Christoph Engel. His vision, rigor, and intellectual generosity have shaped our thinking. We learned a great deal from him, especially how to connect different disciplines and ask meaningful questions. Collaborating with him has been an enriching and enjoyable experience, both academically and personally. We are deeply grateful and hope that this collaboration will continue in the years to come, as there is still much to explore together at the intersection of law, economics, and psychology.

References

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. Journal of public economics, 1(3–4), 323–338.

Alm, J., McClelland, G. H., & Schulze, W. D. (1999). Changing the social norm of tax compliance by voting. Kyklos, 52(2), 141–171.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. Management science, 66(7), 2801–2819.

Becker, G. S. (1968). Crime and punishment: An economic approach. Journal of political economy, 76(2), 169–217.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Casal, S., & Mittone, L. (2016). Social esteem versus social stigma: The role of anonymity in an income reporting game. Journal of Economic Behavior & Organization, 124, 55–66.

Cummings, R. G., Martinez-Vazquez, J., McKee, M., & Torgler, B. (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. Journal of Economic Behavior & Organization, 70(3), 447–457.

Easton, D. (2017). A systems analysis of political life. In Systems research for behavioral science (pp. 428–436). Routledge.

Engel, C., Mittone, L., & Morreale, A. (2020). Tax morale and fairness in conflict an experiment. Journal of Economic Psychology, 81, 102314.

Engel, C., Mittone, L., & Morreale, A. (2024). Outcomes or participation? Experimentally testing competing sources of legitimacy for taxation. Economic Inquiry, 62(2), 563–583.

Guala, F., & Mittone, L. (2010). How history and convention create norms: An experimental study. Journal of Economic Psychology, 31(4), 749–756.

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. Journal of the European Economic Association, 14(3), 608–638.

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. Journal of Economic psychology, 29(2), 210–225.

Luttmer, E. F., & Singhal, M. (2014). Tax morale. Journal of economic perspectives, 28(4), 149–168.

Scharpf, F. (1999). Governing in Europe: Effective and democratic?. Oxford University Press.

Torgler, B. (2005). Tax morale and direct democracy. European Journal of Political economy, 21(2), 525–531.

Torgler, B. (2007). Tax compliance and tax morale: A theoretical and empirical analysis. In Tax Compliance and Tax Morale. Edward Elgar Publishing.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

An Application of the Panel Hurdle Model to Experimental Data

Peter G. Moffatt

Using a user-written STATA program developed by Engel and Moffatt (2014), we apply the panel hurdle model to experimental data from a public goods game. Subjects are assumed to be one of two types: free-riders, and potential contributors. We focus on a test of the "house money effect". This effect is seen to be significant in the first hurdle: specifically, house money makes a subject more likely to be a potential contributor. Hence we find that the effect of house money is more than just an effect on behaviour; it has the effect of changing a subject from one type to another. This result is potentially important in the external validity debate.

A. Introduction

It is standard practice in economic experiments to provide subjects with an initial endowment, for the obvious reason that it is unreasonable to expect them to use their own money in carrying out the tasks that they are required to perform during the experiment. However, an important question is whether the provision of this endowment has any effect on behaviour in the tasks, since this would imply that the endowment is causing behaviour in the experiment to diverge from behaviour that would be expected in real decisions. This effect has come to be known as the "house money effect".

There is some evidence of a house money effect in risky choice experiments, the effect being to make choices more risk-seeking (Thaler and Johnson, 1990; Keasey and Moon, 1996). There is also evidence from dictator game experiments that subjects are much more willing to give money away when the endowment is unearned rather than earned (Cherry et al., 2002). In a metastudy, Engel (2011) finds that dictators give an average of 29.52 % of their endowment if this is "manna from heaven", while this proportion falls to 12.63 % if they had to earn the money in some real effort task.

Here, we are interested in interactive experiments. Clark (2002) has already suggested a number of possible ways in which house money might

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

have an effect in such settings. First, house money might be used by subjects to "purchase" greater quantities of non-pecuniary goods such as "public-spiritedness", or "reciprocal fairness", than they would do in real life. Second, a shift towards more risk seeking may induce subjects to experiment with "strategic giving", that is, to bet on their co-players being reciprocators (Kreps et al., 1982). What is important about these hypotheses is that they all embody the idea that the provision of house money makes it cheaper, and therefore arguably more likely, that participants trade some pecuniary gain for increasing the payoff of other participants. Critically, (some) subjects are expected to condition their choices on the presence or absence of house money, i.e. on a general feature of the treatment, irrespective of their experiences from repeated interaction with other subjects. This yields a directed hypothesis: if they receive house money, some subjects no longer play Nash.

Now, even subjects who hold social preferences have been shown, in the main, not to be unconditionally cooperative. Rather they condition their willingness to cooperate on their beliefs about the cooperativeness of their interaction partners and, if the game is repeated, on their experiences (Fischbacher and Gächter 2001, 2010). Provided they are pessimistic, or they have experienced little cooperativeness in the current experiment, such participants will not make positive contributions either. This provides a challenge for parametric estimation. One needs to distinguish those participants who do not contribute as a consequence of their type, from other participants who do not contribute for reason of the context being insufficiently favourable.

One statistical model that allows for estimating the relative importance of both causes for seemingly selfish behaviour is the double hurdle model. The double hurdle model allows for two types of zero contribution: zeros which result from the subject being a free-rider, or a "zero-type"; and zeros which result from the circumstances of the task, in spite of the subject being a potential contributor. Hurdle models have a number of advantages. They make it possible to incorporate the "zero-type", and therefore do not collapse zero-type with zero-censoring. Moreover, the hurdle model allows estimation of the proportion of the population who are of the zero-type. Better still, it allows the probability of a subject being of zero-type to depend on observables.

Since the data set that we shall be working with consists of repeated tasks for each subject, we shall be estimating the panel version of the double hurdle model (Dong, et al. 2004; Dong and Kaiser, 2008; Engel and Moffatt, 2014). The extension of the model to panel data is demanding because

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the outcome of the "first hurdle", that is, the determination of whether a subject is of the zero-type, must apply to that subject for the whole experiment. Switching in and out of the zero-type is ruled out. In contrast, the outcome of the second hurdle, that is, the amount actually contributed in any particular task, is determined at the level of individual observations. In principle, subjects classified as "free-riders" must necessarily contribute zero in *every* task. This may be seen as too strict a definition though; we might like to make it possible for subjects to be classified as free-riders even if they contribute positive amounts on only a small number of occasions. For this purpose, we introduce a "tremble parameter", which allows such observations to be attributed to lapses of concentration.

We use a data set that has kept experimental economists engaged for more than two decades. The data was originally generated by Clark (2002). In his experiment, there are 10 rounds, and subjects are in groups of 5, with the make-up of groups changing between rounds. In each round, the endowment is 80 tokens, which subjects are required to allocate between an individual account and a group account. There are 150 subjects, of whom 75 are in the "house money" treatment (given 80 tokens at the start of each round), and 75 are in the "own money" treatment (nothing given to them).

Clark (2002) himself did not find a significant house money effect. Yet this non-result might follow from the fact that he used unconditional non-parametric tests with low statistical power. Harrison (2007) re-examined the same data set parametrically and at the individual level, and found evidence of a house money effect. Harrison discussed the important role of zero observations in this data set, and in particular, found that house money impacts on the proportion of free-riding.

The objective of this paper is to continue where Harrison (2007) left off, and to estimate a fully parametric panel double hurdle model with dependence between the two hurdles, and to test for the house money effect within this framework. The dependent variable is the amount contributed to the group account in a particular round. In order to allow house money to impact on type, a dummy for the house money treatment is included in the first hurdle. Key explanatory variables appearing in the second hurdle are the round number, and the mean contribution of the other group members in the previous round.

The remainder of the paper is organized as follows. In section B, we introduce the statistical model. In section C we describe Clark's data. In section D, we present our results. Section E concludes.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

B. The panel hurdle model

Our statistical model assumes that two hurdles need to be passed in order for a positive contribution to be observed. The first hurdle determines whether the subject is a free-rider; if they fall at the first hurdle, they are a free-rider, and their contribution will be zero in all tasks. If the first hurdle is passed, a "second hurdle" is encountered for each task, and whether this is crossed determines whether the contribution in that task is zero or positive. This data generating process consists of repeated observations, which is why we develop a panel estimator. Another feature of the data that we need to deal with is upper censoring, which arises because there is an upper limit to the contribution.

I. Previous Literature

The basic form of the panel hurdle model has been introduced, and applied to data on food demand, by Dong, et al. (2004) and Dong and Kaiser (2008). STATA code to estimate the panel hurdle model has been provided by Engel and Moffatt (2014), and further detail on practical issues of estimation was provided by Moffatt (2020). The model has been applied in many other settings, including public goods games (Cheung, 2014), tax evasion (Guerra and Harrington, 2018), and corporate emissions trading behaviour (Abrell et al., 2021).

Our application is to data on a public goods experiment originally published by Clark (2002). A previous version of the present paper appeared as Engel and Moffatt (2012).

II. Basic Panel hurdle Model with upper censoring

Let i index subject ($i=1,\ldots,n$) and t index task ($t=1,\ldots T$). Let y_{it} be the contribution of subject i in task t, and let y_{max} be the maximum possible contribution (80 in the experiment). Let z_i be a vector of characteristics of subject i relevant in explaining whether the first hurdle is passed. Let x_{it} be a vector of characteristics of subject i and/or task t that determine the contribution in that task.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

First Hurdle:

$$\begin{aligned} d_i^* &= z_i'\alpha + \varepsilon_{1,i} \\ d_i &= 1 \text{ if } d_i^* > 0;\ d_i = 0 \text{ otherwise.} \\ \varepsilon_{1,i} &\sim N(0, 1) \end{aligned} \tag{1}$$

Second Hurdle:

$$\begin{aligned} y_{it}^{**} &= x_{it}'\beta + u_i + \varepsilon_{2,it} \\ y_{it}^* &= 0 \text{ if } y_{it}^{**} \leq 0; \\ y_{it}^* &= y_{it}^{**} \text{ if } 0 < y_{it}^{**} < y_{\max}; \\ y_{it}^* &= y_{\max} \text{ if } y_{it}^{**} \geq y_{\max}. \\ \varepsilon_{2,it} &\sim N\left(0, \sigma^2\right);\ u_i \sim N\left(0, \sigma_u^2\right) \end{aligned} \tag{2}$$

Observed:

$$y_{it} = d_i\, y_{it}^* \quad i = 1, \cdots, n \quad t = 1, \cdots, T \tag{3}$$

The central feature of this model is that the first hurdle has only one outcome per subject, and that outcome applies to all observations for that subject. If subject *i* falls at the first hurdle (d_i=0), then all observations on *y* for subject *i* must be zero (y_{it}=0, t=1,...,*T*). If subject *i* does not fall on the first hurdle, i.e. if she makes a positive contribution in the first round, this suffices to determine the subject's type. She is not a free-rider. If the subject contributes 0 in any later period, even in all later periods, this must be in reaction to observed or expected circumstances, not to type.

Note also that there is a subject-specific random effect term (u_i) appearing in the second hurdle that allows between-subject heterogeneity. Conditional on d_i=1, we obtain the random-effects 2-limit Tobit likelihood:

$$\begin{aligned} (L_i | d_i = 1, u_i) = \prod_{t=1}^{T} &\left[1 - \Phi\left(\frac{x_{it}'\beta + u_i}{\sigma}\right)\right]^{I(y_{it}=0)} \\ &\left[\frac{1}{\sigma}\phi\left(\frac{y_{it} - x_{it}'\beta - u_i}{\sigma}\right)\right]^{I(0<y_{it}<y_{\max})} \\ &\times \left[1 - \Phi\left(\frac{y_{\max} - x_{it}'\beta - u_i}{\sigma}\right)\right]^{I(y_{it}=y_{\max})} \end{aligned} \tag{4}$$

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Note that, at this stage, we are also conditioning on the value of u_i.

Conditional on d_i=0, the likelihood is trivial, and depends simply on whether or not all observations are zero for subject i:

$$\begin{aligned}(L_i|d_i = 0) = 0 \text{ if } \sum_{t=1}^{T} y_{it} > 0 \\ = 1 \text{ if } \sum_{t=1}^{T} y_{it} = 0\end{aligned} \tag{5}$$

The marginal likelihood for subject i is then obtained as:

$$L_i = \int_{-\infty}^{\infty} \left[\Phi\left(z_i'\alpha\right)\left(L_i|d_i = 1, u\right) + \left[1 - \Phi\left(z_i'\alpha\right)\right]\left(L_i|d_i = 0\right)\right] f\left(u; \sigma_u\right) du \tag{6}$$

where $\Phi(z_i'\alpha)$ is the probability of passing the first hurdle (1), the two conditional likelihoods are as defined in (4) and (5), and $f(u; \sigma_u)$ is the $N\left(0, \sigma_u^2\right)$ density function for u. Note that in (6), the i subscript is absent from u; this is because here u is the random variable over which integration is being performed.

III. Panel hurdle model with tremble

A potential problem with the model described in II above is that it is too rigid. In particular, an individual can only be classified as falling at the first hurdle (i.e. a free-rider, or selfish type) if their observations on y are zero in every task. We might wish to relax this assumption and make it possible to also classify a subject as a free-rider who reports zero in nearly every task, with the rationalisation that the subject may have suffered a lapse of concentration on the small number of tasks for which the observation was positive. Such "lapses of concentration" might be considered to be particularly likely in early rounds, where they might be better interpreted as the result of a misunderstanding of the experiment.

To allow this, we introduce a tremble parameter, ω. This is the probability of a lapse of concentration on any individual task. When this occurs, the distribution of y is assumed to be uniform on $[0, y_{max}]$ where y_{max} is the upper limit of the action space (or, in a setting in which there is no upper limit, the highest observed value of y in the data).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

With the introduction of ω, (4) and (5) become:

$$(L_i|d_i = 1, u_i) = \prod_{t=1}^{T} \left[(1-\omega) \left(1 - \Phi \left(\frac{x_{it}'\beta + u_i}{\sigma} \right) \right) \right]^{I(y_{it}=0)} \times \left[(1-\omega) \frac{1}{\sigma} \phi \left(\frac{y_{it} - x_{it}'\beta - u_i}{\sigma} \right) + \frac{\omega}{y_{\max}} \right]^{I(0<y_{it}<y_{\max})} \times \left[(1-\omega) \left(1 - \Phi \left(\frac{y_{\max} - x_{it}'\beta - u_i}{\sigma} \right) \right) \right]^{I(y_{it}=y_{\max})} \quad (7)$$

$$(L_i|d_i = 0) = \omega^{N_i(y_{it}>0)} (1-\omega)^{N_i(y_{it}=0)} \quad (8)$$

where $N_i(.)$ is the number of occurrences of the event contained in parentheses, for subject i.

The marginal likelihood is then constructed as in (6), but using (7) and (8) in place of (4) and (5).

A technical problem with the model defined in (7) and (8) is that the log-likelihood function is monotonically increasing in the parameter ω. Intuitively, this can be interpreted in terms of a tendency for all of the positive observations to be attributed to the tremble. This problem does not appear to arise in other classes of model in which tremble parameters are introduced (see Moffatt and Peters, 2001).

One solution to this estimation problem is simply to assume the presence of a tremble, rather than to attempt to estimate the tremble parameter. That is, we propose that a reasonably small value for ω is chosen, say 0.01, and the other parameters of the model are estimated on the assumption that this is the true tremble probability.

IV. Panel hurdle model with dependence

The models developed in 2.2 and 2.3 assume that there is no correlation between the error terms in the two hurdles. In this section, this assumption is relaxed. Subject i's idiosyncratic propensity to pass the first hurdle is represented by the error term $\varepsilon_{1,i}$; her idiosyncratic propensity to contribute, conditional on passing the first hurdle, is represented by u_i. It is between these two terms that we introduce a correlation:

$$\text{corr}(\varepsilon_1, u) = \rho \quad (9)$$

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

How is the correlation parameter ρ incorporated in estimation? Let us return to the first hurdle:

$$\begin{aligned} d_i^* &= z_i'\alpha + \varepsilon_{1,i} \\ d_i &= 1 \text{ if } d_i^* > 0;\ d_i = 0 \text{ otherwise.} \\ \varepsilon_{1,i} &\sim N(0,1) \end{aligned} \tag{10}$$

Since $\text{corr}(\varepsilon_1, u) = \rho$ we may represent ε_1 as (suppressing the i subscript):

$$\varepsilon_1 = \rho\frac{u}{\sigma_u} + \sqrt{1-\rho^2}\,\xi \tag{11}$$

where $\xi \sim N(0,1)$ and $\xi \perp u$. The requirement for passing the first hurdle becomes:

$$d_i = 1 \text{ if } \xi > -\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-\rho^2}}. \tag{12}$$

whence the probability of passing the first hurdle (conditional on u) becomes:

$$\Phi\left(\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-\rho^2}}\right) \tag{13}$$

The subject i log-likelihood function becomes:

$$\begin{aligned} L_i = \int_{-\infty}^{\infty} \Bigg[&\Phi\left(\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-\rho^2}}\right)(L_i|d_i=1,u) \\ &+ \left[1-\Phi\left(\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-\rho^2}}\right)\right](L_i|d_i=0)\Bigg] f(u;\sigma_u)\,du \end{aligned} \tag{14}$$

where $(L_i \mid d_i, u)$ and $(L_i \mid d_i = 0)$ are defined in either (4) and (5), or (7) and (8), depending on whether a tremble is assumed.

In a double hurdle model with just one observation per subject, there are known to be problems identifying the correlation coefficient ρ (Smith, 2003). By contrast, with panel data, and using the estimation approach outlined in this section, the parameter may be estimated.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

V. Estimation

Estimation is performed using the method of Maximum Simulated Likelihood (MSL, Train, 2003). This requires the use of Halton draws, which, when converted to normality, represent simulated realisations of the random effect term u. In the model with dependence, in accordance with (13), the simulated values also appear in the probability of passing the first hurdle. Maximisation of the simulated likelihood function is performed using the ml routine in STATA. All programs are available from the authors upon request.

C. Data

The data set we use is from the experiment conducted by Clark (2002). In this experiment, there are 10 rounds, and subjects are in groups of 5, with the make-up of groups changing between rounds. In each round, the endowment is 80 tokens, which subjects are required to allocate between an individual account and a group account. There are 150 subjects, of whom 75 are in the "house money" treatment (given 80 tokens at the start of each round), and 75 are in the "own money" treatment (nothing given to them).

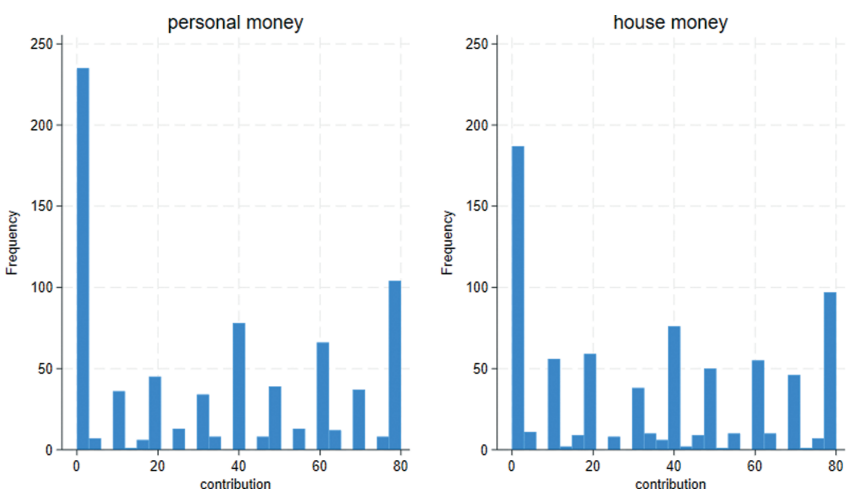


Figure 1: Histogram of Contributions by treatment.

Figure 1 shows that, both with and without house money, the data is left and (less intensely) right censored. The key difference between the two

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

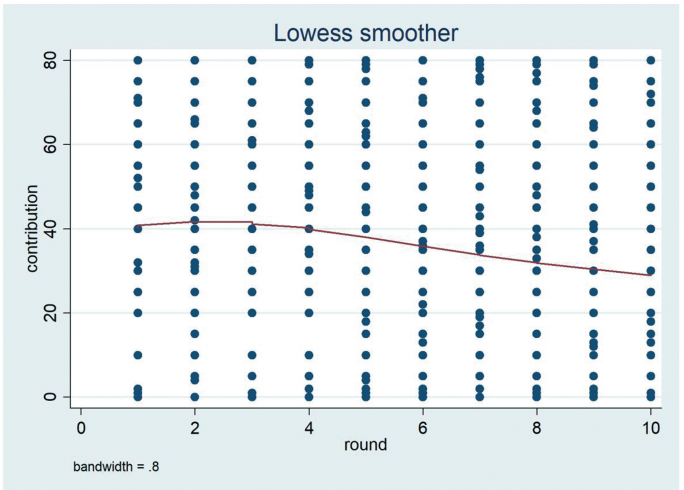


Figure 2: Plot of contribution against round number.

histograms is that participants appear to be more likely to contribute zero if they are using their own money.

Figure 2 shows a plot of contribution against round number, with a lowess smoother. The smoother reveals that contributions slightly decline as the experiment progresses. Hence the inclusion of round number in the second hurdle is justified.

Figure 3 shows a plot of contribution against the mean of contributions of the other group members in the previous round. An important feature of the experiment is that in each round subjects are made aware of the group mean in the previous round. It is clear from the smoother that participants are sensitive to this measure of recent experience, justifying its inclusion as an explanatory variable in the second hurdle. The reason for using the lagged mean contribution of *other* group members, rather than the group mean, is to guard against the consistency problems inherent in a dynamic panel. At the same time, this variable explains, rather than just neutralizes, the additional dependence resulting from the fact that choices are nested in participants, who are nested in groups. Since groups are randomly and anonymously composed, group effects collapse with the experiences participants encounter within their group.

When using the lagged mean contribution of other group members as an explanatory variable, a slight problem that arises is that this variable contains missing values in the first round. Contributions from the first round

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

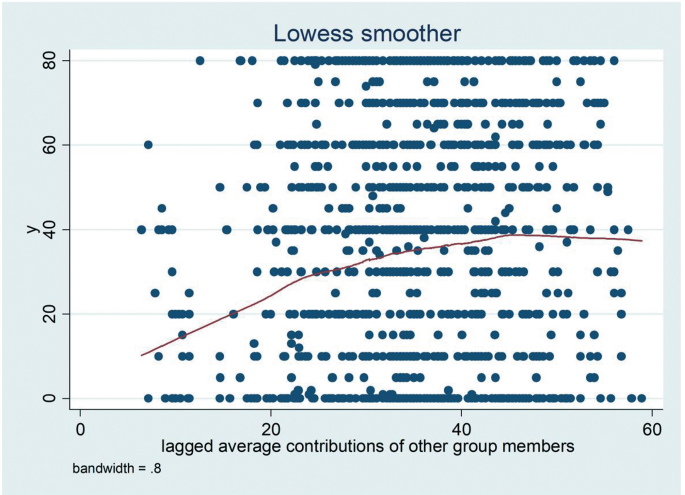


Figure 3: Plot of contribution against mean of contribution of the *other* group members in previous round.

could be excluded from estimation. However, rather than doing this, we follow Bardsley and Moffatt (2007, see their Table II, note 4) by assuming that, in the first round, subjects make a judgement (perhaps based on lifetime experience prior to the experiment) of what a typical contribution of others is likely to be. This value is then used as the value of the lagged mean in the first round. The value we use for this purpose is, in fact, 66 units, and this is obtained, via a grid search, as the value which maximises the sample log-likelihood. This relatively high value implies that subjects tend to start the experiment optimistically. While this assumption may be viewed as somewhat arbitrary, the main result of the paper (relating to the house money effect) is in fact robust to the choice of value made here.

D. Results

Results from various models are shown in Table 1. The first is a binary probit model estimated using the cross section of 150 subjects, with dependent variable taking the value 1 if at least one of the subject's contributions is positive, and zero otherwise. This model gives estimates of the first hurdle, which are used as starting values in the estimation of the corresponding part of the hurdle model. The second model is a panel tobit

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 1: Results of hurdle and related models.

	probit	Panel tobit	Hurdle (ρ=0; ω=0)	Hurdle with dep. (ω=0)	Hurdle dep. (ω=0.001)
First hurdle					
Const	0.941		0.968	0.967	0.673
	(0.171)		(0.177)	(0.161)	(0.148)
House money	0.379		0.393	0.256	0.497**
	(0.263)		(0.276)	(0.249)	(0.236)
Second hurdle					
Const		27.251	42.608	35.027	29.283
		(10.453)	(10.388)	(10.364)	(10.010)
round		–2.986***	–3.094***	–3.111***	–2.948***
		(0.394)	(0.399)	(0.396)	(0.393)
previous		1.016**	0.954**	0.941**	1.177***
		(0.392)	(0.398)	(0.392)	(0.376)
previous^2		–0.013***	–0.013***	–0.013***	–0.015***
		(0.004)	(0.004)	(0.004)	(0.004)
σ		25.943	26.090	26.053	25.097
		(0.681)	(0.689)	(0.688)	(0.696)
σ_u		43.562	29.986	35.457	34.753
		(3.191)	(2.375)	(2.950)	(3.344)
ρ			0	0.901	0.940
				(0.091)	(0.065)
n	150	150	150	150	150
T	–	10	10	10	10
LogL	–57.85	–4722.89	–4707.01	–4705.19	–4688.21
AIC (2k-2LogL)		9457.78	9426.38	9428.38	9416.42

* mildly significant (p<0.10); ** significant (p<0.05); *** strongly significant (p<0.01).

model, which gives estimates of the second hurdle; these are also used as starting values. Then, three panel hurdle models are estimated: the basic model; the model with dependence; the model with dependence and a non-zero tremble probability. The last of these is the best model on the basis of the Akaike Information Criterion (AIC).

The effects seen in the second hurdle are as expected. Firstly, In agreement with other public good studies, and with Figure 2 above, the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

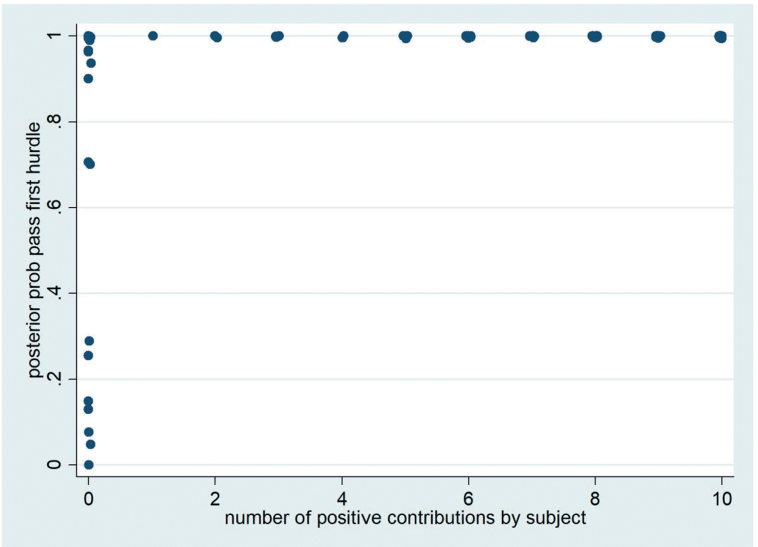


Figure 4: Posterior probability of passing first hurdle against number of positive contributions (posterior probability obtained from model with no tremble).

negative coefficient of round number provides evidence of a decline in contributions over the course of the experiment. Secondly, the positive coefficient of the mean previous contribution by others, and the negative coefficient of its square, confirm the conclusion drawn from Figure 3, that contributions rise but level off with others' contributions. This amounts to evidence of reciprocity.

The key result is the evidence of the house money effect in the final column: house money increases the probability of passing the first hurdle, that is, it increases the probability of being a "potential contributor". Note that this result is only seen in the final column.

The role of the tremble parameter is crucial in reaching this conclusion. In order to see why, we first plot the posterior probability of passing the first hurdle (obtained using Bayes' Rule) from the model with *no* tremble (whose results are presented in the penultimate column of Table 1). Note that all subjects with positive contributions are certain to pass the first hurdle.

Next, we present the same plot, but for the model in which the tremble probability is set to 0.001 (i.e. the model whose results are reported in the final column of Table 1). Note that some subjects with a small number

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

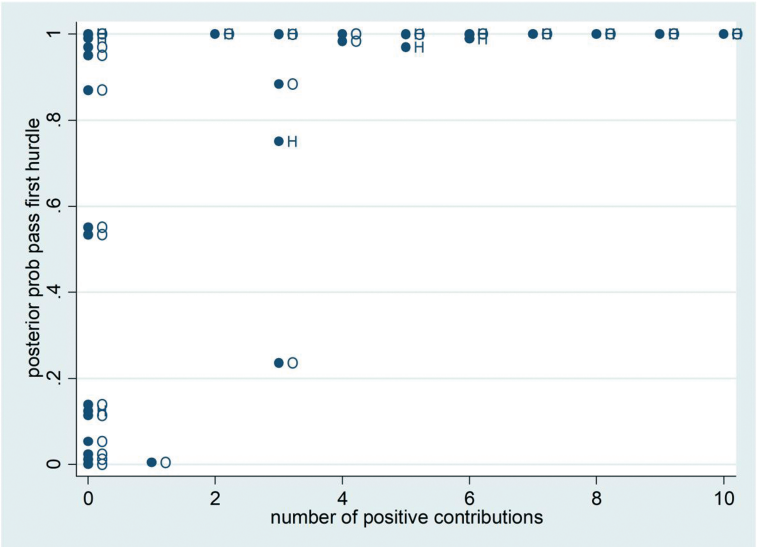


Figure 5: Posterior probability of passing first hurdle against number of positive contributions (posterior probability obtained from model with tremble set to 0.001). H=house money treatment; O=own-money treatment.

of positive contributions have high posterior probability of being a free-rider (being low down in the plot). Note further that these subjects tend to be in the "own-money" treatment, and it is this that drives the difference in results between the penultimate and final columns of Table 1. The small number of subjects in the own-money treatment who are being attributed to the "free-rider" type, despite their occasional contributions, are tilting the balance between no house money effect and a house money effect.

E. Conclusion

In experimental practice, it is common to give participants an endowment. This routine makes it easier to run experiments, and goes largely unquestioned. Yet in reality, people are clearly not given money when deciding how much to donate to charity, or when deciding where to invest. It is their own earned income that is at stake.

Clark's (2002) non-result was therefore a relief for experimentalists. Seemingly, giving participants some money upfront did not endanger ex-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ternal validity. Our paper cautions against such complacency. If only one uses an estimator that matches the complex data generating process, it turns out that, already in Clark's original data, house money moves some individuals out of the free-riding type and into the contributor type.

Given our finding, it seems that the house money debate ended prematurely. More experiments are needed to show under which conditions and, in particular with which experimental designs, the house money effect is negligible, and under which conditions it creates a serious obstacle for extrapolation from the lab to the field.

To enable the thorough investigation of the house money effect, an "own-money" treatment is necessary, and this presents obvious ethical issues. A very useful complement would be field experiments in which subjects are operating in their natural environment where "own money" is the default.

References

Abrell, J., Cludius, J., Lehmann, S., Schleich, J., & Betz, R. (2022). Corporate emissions-trading behaviour during the first decade of the EU ETS. *Environmental and Resource Economics*, *83*(1), 47–83.

Bardsley N. and Moffatt P.G. (2007), "The experimetrics of public goods: inferring motivations from contributions", *Theory and Decision*, 62, 161–193.

Cherry T.L., Frykblom P. and Shogren J.F., (2002) "Hardnose the Dictator", *American Economic Review*, 92 (4), 1218–1221.

Cheung, S. L. (2014). New insights into conditional cooperation and punishment from a strategy method experiment. *Experimental Economics*, *17*(1), 129–153.

Clark J.(2002), "House money effects in public good experiments", *Experimental Economics*, 5, 223–231.

Dong, D., Chung, C., and Kaiser, H.M. (2004), Modeling milk purchasing behavior with a panel data double-hurdle model. *Applied Economics*, 36(8), 769–779.

Dong D. and H.M. Kaiser (2008), "Studying Household Purchasing and Nonpurchasing Behaviour for a Frequently Consumed Commodity: Two Models", *Applied Economics*, 40(15), 1941–1951.

Engel, C. (2011): "Dictator Games: A Meta Study", *Experimental Economics,* 14(4), 583–610.

Engel, C., & Moffatt, P. G. (2012). Estimation of the house money effect using hurdle models. *MPI Collective Goods Preprint*, (2012/13).

Engel, C., & Moffatt, P. G. (2014). dhreg, xtdhreg, and bootdhreg: Commands to implement double-hurdle regression. *The Stata Journal*, *14*(4), 778–797.

Fischbacher, U. and S. Gächter (2001): "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters* 71, 397–404.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Fischbacher, U. and S. Gächter (2010): "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments", *American Economic Review,* 100, 541–566.

Guerra, A., & Harrington, B. (2018). Attitude–behavior consistency in tax compliance: A cross-national comparison. *Journal of Economic Behavior & Organization*, *156*, 184–205.

Harrison G.W., 2007, "House money effects in public good experiments: comment", *Experimental Economics*, 10, 429–437.

Keasey K. and Moon P. (1996) "Gambling with the house money in capital expenditure decisions", *Economics Letters*, 50, 105–110.

Kreps, D.M., P. Milgrom et al., 1982, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma", *Journal of Economic Theory*, 27, 245–252.

Moffatt, P.G. (2020). *Experimetrics: Econometrics for experimental economics*. Bloomsbury Publishing.

Moffatt, P.G. and Peters S.A., 2001, "Testing for the presence of a tremble in economic experiments", *Experimental Economics*, 4, 221–228.

Smith M.D., 2003, "On dependency in double hurdle models", *Statistical Papers*, 44, 581–595.

Thaler R. and Johnson E. (1990) "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice", *Management Science*, 36, 643–660.

Train, K.E., 2003, *Discrete Choice Methods with Simulation*, Cambridge.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Anticipated Third-Party Rewards
Experimental Evidence on Social Control and Cooperation

Lilia Wasserka-Zhurakhovska

This contribution honors Christoph Engel's pioneering research on rule-following, legitimacy, and the institutional foundations of cooperation. Building on his experimental law and economics agenda, I investigate how anticipated third-party rewards influence trust and reciprocity in strategic interaction. The laboratory design combines a trust game with a subsequent helping game, in which Observers reward Trustees based on their behavior. Treatments vary depending on whether Investors and Trustees are informed in advance about the presence of Observers. Consistent with Engel's findings on conditional rule-following and the motivating power of institutions, the results show that Observers reward Trustees more generously when return transfers are higher, and that Trustees return more when they anticipate being observed. However, this strategic adjustment does not lead to higher investments by Investors, highlighting both the reach and the limits of anticipated indirect reciprocity. The study thus illuminates the subtle mechanisms through which institutions shape cooperation, while paying tribute to Engel's lifelong commitment to understanding how norms, roles, and third parties sustain social order.

A. Introduction

Christoph Engel has repeatedly asked how fragile social order can nonetheless persist when individuals face clear incentives to act selfishly. In his reflections on law as an "impossible discipline" (Engel, 2024), he stresses that rules and institutions do not operate mechanically. Their effectiveness depends on how they are perceived, interpreted, and legitimized. This line of questioning has shaped his experimental law-and-economics research: individuals are neither unconditional egoists nor simple altruists, but respond systematically to institutional and social cues that frame their interactions.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The present study takes up this challenge by examining whether the mere anticipation of institutional observation can alter behavior in a trust game. Specifically, I investigate a two-stage design in which two players interact in an investment game, while a third party (Observer) subsequently decides on rewards for the trustee, conditional on the trustee's behavior. My treatments vary whether trustees are aware from the outset that they may later be observed and rewarded (Announced treatment) or whether this possibility remains undisclosed until after their decisions (Unannounced treatment).

By embedding this mechanism into a controlled laboratory setting, the study connects directly to Engel's concern with conditional rule following, legitimacy, and the subtle role of institutional signals. Trustees behave more generously when they anticipate third-party evaluation and reward, suggesting that even minimal institutional cues are sufficient to trigger more cooperative behavior. Yet investors do not respond with higher levels of trust. This asymmetry highlights both the reach and the limits of anticipated indirect reciprocity: institutional interventions may strongly affect immediate reciprocity but fade as interactions become more indirect and cognitively demanding.

In this way, my findings resonate with Engel's broader research program. They echo his emphasis on conditionality, legitimacy, and the role of institutions in sustaining cooperation, while also illustrating the boundaries of such mechanisms. The study thus exemplifies Engel's call to accept complexity rather than simplifying it away and contributes to his intellectual legacy by advancing my understanding of how institutions shape behavior in layered social interactions.

B. Related Literature

Engel's research has repeatedly demonstrated that individuals condition their behavior on cues about norms and the expected actions of others. Rule compliance has been shown to depend on perceived compliance in one's reference group (Engel and Desmet, 2021). Similarly, even minimal social information has been shown to trigger peer effects in rule violation (Engel, 2023). These insights resonate strongly with my finding that trustees adapt when the existence of a monitoring third party is announced in advance. Even a minimal institutional cue suffices to trigger more cooperative behavior.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The institutional framing of decisions is another hallmark of Engel's research. The simple assignment of a judicial office led participants to enforce norms more strongly, suggesting that institutional roles activate a sense of duty (Engel and Zhurakhovska, 2017). My design parallels this mechanism: trustees act more generously once they recognize the evaluative role of a third party, even when the observer has no material stake. Similarly, legitimacy has been shown to derive both from outcomes and from the opportunity to participate in decision-making (Engel, Mittone and Morreale, 2024). The anticipatory effect of a third party in my study can be seen as a minimal form of participatory legitimacy.

my own work has likewise engaged with the institutional determinants of cooperation and fairness. Participation procedures increase subsequent kindness toward authorities, irrespective of outcomes (Kleine, Langenbach and Zhurakhovska, 2017). Similarly, impartial decision-makers have been shown to be swayed by stakeholder communication, though in asymmetric ways (Kleine, Langenbach and Zhurakhovska, 2016). Other studies emphasize how dishonesty and norm violations spread in distinct ways across genders (Böhm, Goerg and Wasserka-Zhurakhovska, 2023) and how leaders face ethical dilemmas that interact with gender differences in dishonesty (Grosch, Müller, Rau and Wasserka-Zhurakhovska, forthcoming). These contributions highlight the power of procedural and contextual signals to affect compliance and reciprocity, much as the presence of a third party in the current experiment alters trustee behavior.

Engel's research has also emphasized the importance of fairness perceptions and their fragility in social dilemmas. Rule-following declined once income heterogeneity was revealed, underscoring the role of fairness expectations (Engel, Mittone and Morreale, 2020). In my experiment, third-party rewards can be interpreted as fairness signals that amplify cooperative choices among trustees. Yet, in line with evidence showing that deciding on behalf of others does not mitigate selfishness (Engel and Cerrone, 2019), my investors remain unresponsive. This may be because the chain of reasoning is too indirect, or because the fairness signal fades with distance from the immediate decision context.

Finally, the broader literature on third-party enforcement underscores both the promise and the limits of institutional interventions. Norm enforcement has been identified as a cornerstone of large-scale cooperation (Fehr and Schurtenberger, 2018), and both direct and indirect sanctions have been shown to sustain cooperation (Balafoutas, Nikiforakis and Rockenbach, 2016). Third-party punishment can operate as

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a costly signal of trustworthiness (Jordan et al., 2016), while theories of third-party reward and punishment highlight their role as mechanisms of norm enforcement (Chang et al., 2022). My findings add nuance to this literature by showing that anticipation of third-party reward shifts trustee behavior but does not necessarily cascade into increased trust. This asymmetry highlights a boundary condition for institutional interventions: they may bolster compliance at one stage of interaction while leaving upstream trust decisions unaffected.

C. Experimental Design

Each subject was randomly assigned to one of three roles: Investor/Observer (Player A), Trustee (Player B), or Outsider (Player C). The study combined a one-shot trust game with a subsequent one-shot helping game. The trust game builds on Berg, Dickhaut and McCabe (1995), while the helping game is a variant of the dictator game first introduced by Nowak and Sigmund (1998) and later adapted to study indirect reciprocity and strategic reputation building (Engelmann and Fischbacher, 2009). Both stages relied on the strategy method (Selten, 1967), requiring participants to make conditional choices for every possible scenario rather than only for the realized case.

In the trust game, both Player A (Investor) and Player B (Trustee) received an endowment of 100 Taler. The Investor could transfer up to 60 Taler, choosing any multiple of ten from the set $X \in \{0, 10, 20, 30, 40, 50, 60\}$, to Trustee. The transferred amount was then tripled by the experimenter. The Trustee, having received $3 * X$, then chose a return transfer $Y \in \{0, 1, 2, 3\}$ times the original investment X, using the strategy method. Thus, for every possible transfer by Investor, Trustee had to indicate whether they would return nothing, the transfer itself, double the transfer, or triple the transfer. In case of zero investment, no return was possible. The Outsiders did not take any decision in the trust game, nor were they informed about the decisions of the other players. Outsiders received a fixed payoff of 100 Taler in the trust game, identical to what Investors and Trustees earned if no transfer had been made.

In the helping game, each Player A was matched with a Player B from another group. In this game Player A became an Observer. Importantly, Observers did not evaluate the Trustee from their own trust game. Instead, they observed the complete strategy of an Investor–Trustee interaction

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

from a different group (i.e., the investment and the corresponding return transfer) and then decided how much to transfer. This ensured that Observers' transfers could not be interpreted as direct reciprocity toward their own partner, but only as third-party reward. Observers were endowed with 100 Taler and could transfer any amount between 0 and 100 to the Trustee (Player B). Transfers were tripled by the experimenter, so the Trustee received $3 * T$. Observers' transfer decisions were elicited using the strategy method: they specified transfers for Trustees conditional on each possible history from the first stage (i.e., every combination of Investor transfer X and Trustee return decision Y). Additionally, Observers were asked how much they would transfer to a Passive Trustee (who had received 0 in the trust game) and to an Outsider.

Two treatments were implemented. In the Unannounced treatment, Investors and Trustees were not aware of the existence of Observers when making their decisions in the trust game. Only afterward were Observers introduced in the helping game. In the Announced treatment, by contrast, Investors and Trustees were informed ex ante that their trust game behavior could be observed by a third party, who might later reward Trustees in the helping game. For Observers, the two treatments were strategically identical: they always observed the Trustee's behavior and decided on transfers. However, they did know whether their existence was announced beforehand, which in principle could have influenced their rewarding behavior.

The experiment was run at the University of Bonn. It was programmed and conducted with z-Tree (Fischbacher, 2007). Four sessions with 96 subjects were held, resulting in 22 independent observations per treatment for the roles of Investor/Observer and Trustee. Not every group was assigned an Outsider, and while the instructions avoided explicitly stating that each group had one, they were carefully worded to prevent any deception in this regard. Subjects were recruited via ORSEE (Greiner, 2015). None had participated in prior trust or dictator/helping games. The subject pool consisted mainly of students: 19 from economics, 20 from law, and the remainder from other disciplines; 46 percent were female. Average earnings were 12.52 Euro (including a 4 Euro show-up fee), corresponding to about $ 15–16 depending on the exchange rate at the time. Sessions lasted about 70 minutes. Control questions ensured comprehension before the start, and participants filled out a short post-experiment questionnaire including demographics.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

D. Behavioral Predictions

Anticipation functions similarly to backward induction. First, one must form expectations about the choices of Observers. Then, Trustees' reactions to these choices can be predicted. In the helping game, assuming individuals aim solely to maximize their own payoffs, the theoretical prediction is that no transfers will occur. A self-interested Trustee has no monetary incentive to return anything, regardless of the information available about the helping game. Similarly, a rational, self-interested Investor anticipates this and chooses not to invest in any treatment. Under the assumption of pure payoff maximization and common knowledge, the unique Nash equilibrium predicts zero transfers across all games and treatments.

However, empirical evidence contradicts this prediction. In helping games, positive transfers are observed (Forsythe et al., 1994), and in trust games, transfers occur in both directions (Berg, Dickhaut, and McCabe, 1995). The former is often attributed to social preferences such as the "warm glow" effect (Andreoni, 1990), while the latter is explained by strong direct reciprocity. Models by Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Levine (1998), and Rabin (1993) suggest that individuals derive utility from rewarding kind actions and punishing unkind ones. What counts as "kind" varies across models.

Levine's (1998) model is particularly relevant here, as it is not limited to two-player interactions. It assumes that individuals derive utility not only from their own payoffs but also from rewarding others based on their perceived altruism. In this framework, a person's utility increases when they reward someone who is seen as altruistic. In the present experiment, Observers do not need to form independent beliefs about Trustees' preferences; they can infer them from Trustees' return transfers to Investors. If some Observers are altruistic and value others' altruism, then:

> **Prediction 1a:** The higher Trustees' return transfers in the trust game, the more help will be transferred by Observers.

The second key question is whether Observers evaluate Trustees' altruism differently in the Unannounced and Announced treatments. Levine's model is a signaling model, where credible signals of trustworthiness should be rewarded. In the Announced treatment, high return transfers may signal trustworthiness, but they may also be interpreted as strate-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

gic actions aimed at gaining rewards. Therefore, Observers may perceive these actions as less genuine.

Falk, Fehr, and Fischbacher (2008) show that intentions matter: actions driven by intrinsic motives are rewarded more than those driven by strategic ones. This suggests that Observers' transfers should be higher in the Unannounced treatment, where Trustees' return transfers cannot be strategically motivated. Similarly, Stanca, Brunni, and Corazzini (2009) find stronger indirect reciprocity when strategic motivations are ruled out.

> **Prediction 1b:** For any given level of Trustees' return transfers, Observers' transfers are lower in the Announced treatment than in the Unannounced treatment.

Following Levine's model, I can also predict Trustees' behavior:

> **Prediction 2a:** Trustees in the trust game make higher return transfers when they receive higher investments.

In the Announced treatment, Trustees know that Observers may reward them based on their behavior. If Trustees believe that some Observers value altruism and are willing to reward it, they may increase their return transfers to signal trustworthiness. If the expected reward outweighs the cost of the transfer, Trustees have an incentive to behave more generously.

> **Prediction 2b:** Trustees in the trust game make higher return transfers in the Announced treatment than in the Unannounced treatment.

Finally, Costa-Gomes, Huck, and Weizsäcker (2014) find that Investors' optimism about return transfers correlates positively with their investment levels. If Investors anticipate that Trustees will behave more generously in the Announced treatment, they may invest more.

> **Prediction 3:** Investors make higher investments in the Announced treatment than in the Unannounced treatment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

E. Results

We now turn to the experimental results. This section is structured along the behavioral predictions derived above. I begin by analyzing Observers' transfers in the helping game, followed by Trustees' behavior in the trust game, and finally Investors' behavior. Throughout, I use both non-parametric tests and regression analysis to evaluate treatment effects.

I. Observers' Transfers

Prediction 1a stated that Observers transfer more the higher the Trustees' relative return transfers are in the trust game. Figure 1 illustrates mean Observers' transfers across all possible combinations of investments and Trustees' relative return transfers, separately for the Unannounced and the Announced treatment. The pattern clearly supports Prediction 1a: the higher the relative return transfer, the more Observers transfer. This holds across both treatments.

A Spearman rank correlation confirms this result: in the Unannounced treatment, Observers' transfers and Trustees' relative return transfers are positively correlated ($\rho = 0.44$, $p < 0.01$), and similarly in the Announced treatment ($\rho = 0.53$, $p < 0.01$). These findings provide strong support for Prediction 1a.

Turning to Prediction 1b, I expected that Observers would transfer less in the Announced treatment for any given level of Trustees' relative return transfer. Yet the regression results in Table 1 show a positive and significant interaction between return transfers and the Announced treatment. This indicates that while baseline transfers are lower in Announced, the sensitivity of Observers to Trustees' generosity is actually stronger when Trustees could anticipate observation. In other words, Observers reward high-return Trustees more intensely if their actions were taken under announced observation. This contrasts with prior findings (e.g., Stanca, Brunni, and Corazzini, 2009) and suggests that factors beyond perceived intentions, such as the salience of reputational incentives, shape strong indirect reciprocity.

Taken together, these results strongly support Predictions 1a but not 1b. Observers reward Trustees more when they return more to Investors, but they do so even more when Trustees could anticipate that their behavior would be observed.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 1: Explaining Observers' Transfers – Comparison of Unannounced and Announced Treatment

	Model 1	Model 2	Model 3	Model 4
Announced Treatment	10.05	5.66	9.87	–1.24
	(10.59)	(10.38)	(8.52)	(9.43)
Relative Return Transfers	16.43***	16.43***	16.41***	14.82***
	(.72)	(.72)	(.72)	(1.00)
Investment	.12***	.12***	.12***	.044
	(.04)	(.04)	(.04)	(.06)
*Announced Treatment *Relative Return Transfer*				2.94**
				(1.37)
*Announced Treatment *Investment*				.14*
				(.08)
Own Investment		.03	–.23	–.23
		(.26)	(.23)	(.26)
Own Profit in TG		.36**	.12	.12
		(.17)	(.15)	(.15)
Transfer to Passive Trustee			.34	.71
			(.47)	(.45)
Transfer to Outsider			.32	.32
			(.45)	(.47)
Constant	–47.22***	–82.30***	–62.71***	–56.65***
	(7.99)	(20.41)	(17.12)	(17.25)
N	1056	1056	1056	1056
N of group	44	44	44	44
P model	< .001	< .001	< .001	< .001
Wald Chi2	528.48	530.67	539.19	550.51

Notes: Random effects Tobit regressions. Standard errors are presented in parentheses. Observations are clustered on individual level. The Announced dummy equals 1 for all observations of the Announced treatment, relative return transfers controls for the relative return of a trustee (Y) for a given investment, investment controls for the investment the trustee has received, own investment is the investment the helper has transfers in the trust game himself to her trustee, Announced* relative return transfer and Announced *investment are interaction terms, own profit in TG controls for the helper's profit from part 1 of the experiment. Transfer to passive trustee and Transfer to outsider are the levels transferred to passive players. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively. Left-censored = 577; right-censored = 14.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

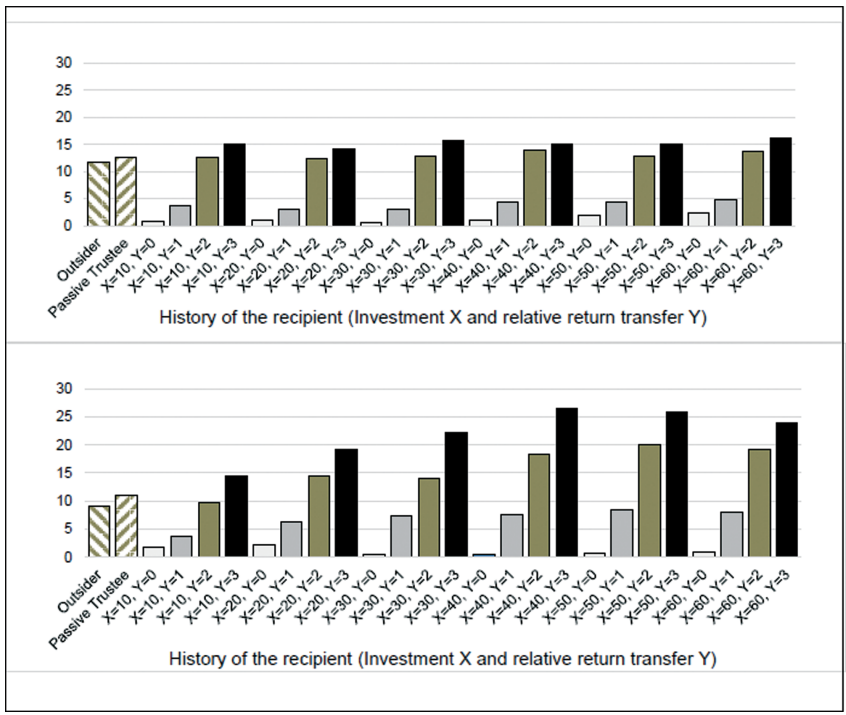


Figure 1: Mean Observers' transfers by Trustees' return transfers and treatment (Announced vs. Unannounced)
Notes: The upper figure displays Observers' transfers in the Unannounced treatment and the lower figure in the Announced treatment. On the x-axis, the exact condition is displayed, i.e., one can see if the recipient is either an Observer or a passive Trustee (no investment and thus no opportunity for a return transfer) or an active Trustee who has received an investment of X and has made a relative return transfer of Y (Y=0: Trustees keep full transfer; Y=1: Trustees return transfer and keep rest; Y=2: equal split; Y=3: full return). On the y-axis, mean Observers' transfers are displayed for the particular situation. Standard errors are indicated.

II. Trustees' Transfers

Prediction 2a stated that Trustees make higher return transfers the more they receive from Investors. Figure 2 shows mean relative return transfers for each level of investment. In the Unannounced treatment, Trustees' transfers increase with investment, consistent with Prediction 2a. The correlation between investments and return transfers is positive and significant (Spearman $\rho = 0.15$, $p < 0.05$).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

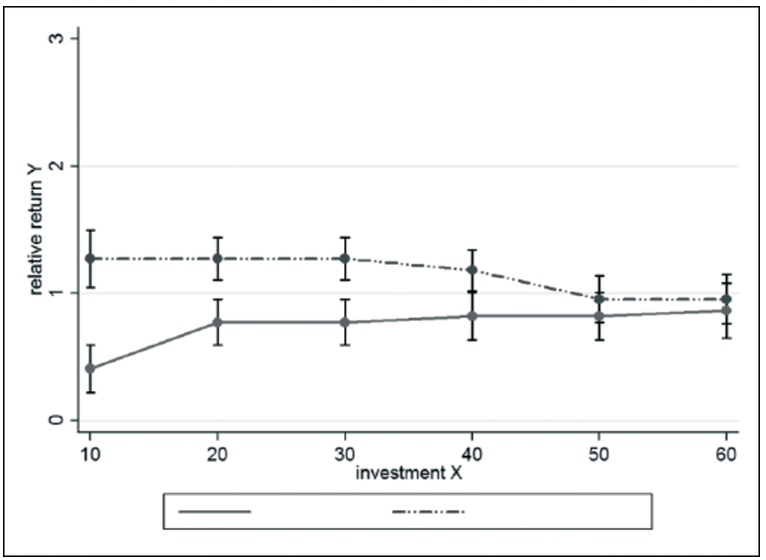


Figure 2: Trustees' relative return transfers by investment and treatment
Notes: On the x-axis, the investment the Trustee has received is depicted; on the y-axis, mean relative return transfers are displayed. The figure compares the Unannounced treatment and Announces treatments. Standard errors are indicated.

For the Announced treatment, however, the relationship is weaker and even negative ($\rho = -0.15$, $p < 0.10$), suggesting that Trustees do not condition as strongly positively on the level of investment when they anticipate being observed.

Prediction 2b stated that Trustees would make higher return transfers in the Announced treatment than in the Unannounced treatment. Figure 2 provides some support: average return transfers are indeed higher in the Announced treatment. A Mann–Whitney U test confirms that the difference is significant ($p < 0.05$).

Ordered Probit regressions with random effects (Table 2) further support this interpretation. The treatment dummy for Announced is positive and highly significant across all specifications, indicating that Trustees return more when they know Observers may reward them. At the same time, the interaction term between investment and the Announced treatment is negative and significant, consistent with the flatter slope observed in Figure 2. Trustees return more overall, but their responsiveness to investment is reduced when they can anticipate being observed.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Overall, Predictions 2a and 2b are supported: Trustees return more when they receive higher investments, and they return more in the Announced treatment, although their conditioning on investment is weaker. This pattern also indicates that trustees' behavior in this setting departs from the standard reciprocity usually observed in trust games, and is instead primarily driven by strategic concerns when anticipating potential rewards from Observers.

III. Investors' Transfers

Finally, Prediction 3 stated that Investors would invest more in the Announced treatment than in the Unannounced treatment. However, this prediction is not supported. Statistically, the investments do not differ between treatments (Mann–Whitney rank-sum: $p > 0.1$). The cumulative distribution functions in Figure 3 suggest some visual differences in the distribution, but these are not statistically significant. Similarly, parametric tests using ordered probit regressions (not reported here) confirm that

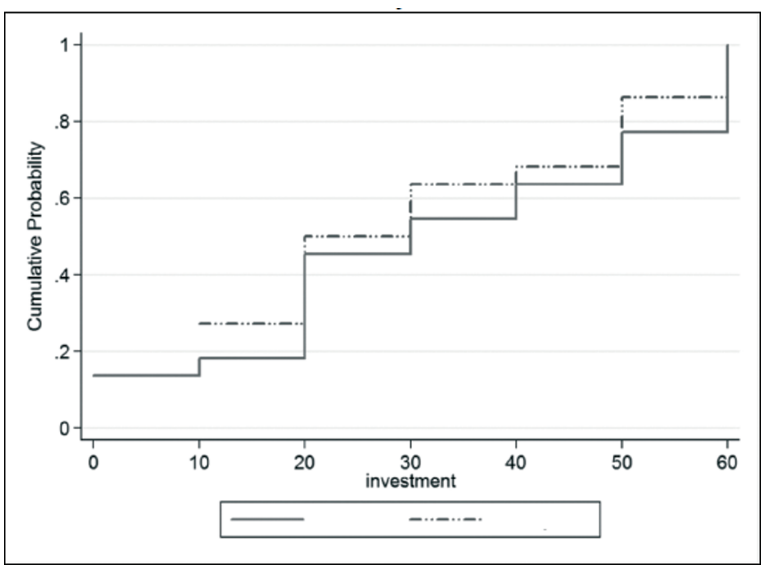


Figure 3: Cumulative distributions of Investors' investments across treatments
Notes: The graph shows an empirical cumulative distribution function of the investments for both treatments. On the x-axis, the investment is depicted; on the y-axis, the estimated probability for each investment is displayed.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 2: Explaining Trustees' Transfers – Comparison of Unannounced and Announced Treatment

	Model 1	Model 2	Model 3	Model 4
Announced Treatment	2.45***	2.45***	4.16***	4.20***
	(.27)	(.27)	(.50)	(.52)
Investment	–.00	–.00	.02***	.02***
	(.00)	(.00)	(.01)	(.01)
*Announced Treatment *Investment*			–.04***	–.05***
			(.01)	(.01)
*Announced Treatment *Female*				–1.11**
				(.44)
Cut 1 Constant	.29	.25	1.13***	1.06***
	(.18)	(.25)	(.34)	(.33)
Cut 2 Constant	2.29***	2.25***	3.28***	3.31***
	(.21)	(.27)	(.38)	(.39)
Cut 3 Constant	4.54***	4.50***	5.65***	5.51***
	(.39)	(.41)	(.52)	(.50)
Rho Constant	.86***	.86***	.87***	.86***
	(.02)	(.02)	(.02)	(.02)
N	264	264	264	264
N of groups	44	44	44	44
P model	0.033	0.101	< .001	< .001
LR Chi2	4.52	4.58	23.73	22.74

Notes: Random effects ordered Probit regressions. Standard errors are presented in parentheses. Observations are clustered on individual level. The Announced dummy equals 1 for all observations of the Announced treatment, investment controls for the investment (X) the trustee has received, the gender dummy equals 1 for all female observations. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

treatment has no statistically significant effect, and no gender–treatment interaction is found. Thus, Investors' behavior remains unaffected by the anticipation of third-party observation.

This pattern suggests that Investors are somewhat sensitive to the strategic implications of observation but less so than Trustees or Observers. One possible explanation, in line with Camerer, Ho, and Chong (2004), is that anticipating higher-order reasoning about indirect reciprocity is cognitively demanding, leading to weaker treatment effects on Investors' behavior.

F. Discussion

This study sets out to investigate how anticipated indirect reciprocity shapes trust, trustworthiness, and third-party helping. Building on a two-stage design with an investment game followed by a helping game, I compared two conditions: one in which the presence of Observers was announced to Investors and Trustees ex ante (Announced treatment), and one in which their presence remained unannounced until after the trust game (Unannounced treatment).

The behavioral predictions derived from models of social preferences and indirect reciprocity received substantial support. Consistent with Prediction 1a, I find a strong positive relationship between Trustees' return transfers and Observers' subsequent transfers. This result aligns with Levine's (1998) framework, where individuals derive utility from rewarding others perceived as altruistic. The Tobit regressions reported in Table 1 confirm that Observers' transfers increase systematically with the generosity of Trustees, both across and within treatments.

Prediction 1b, which anticipated lower transfers by Observers in the Announced treatment, is not supported. Instead, Observers reward generous Trustees even more strongly when Trustees act under announced observation. In other words, Observers do not penalize Trustees for behaving strategically, but still give up part of their own endowment to reward them. This suggests that strong indirect reciprocity operates even when cooperative behavior may be strategically motivated.

Turning to Trustees, I find support for both Predictions 2a and 2b. Trustees return more when investments increase, consistent with conditional reciprocity in the Unannounced treatment. Moreover, average return transfers are significantly higher in the Announced condition, as

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

confirmed by the Mann–Whitney tests and the ordered probit regressions in Table 2. At the same time, the negative interaction between investment and the Announced treatment shows that Trustees' responsiveness to investment is reduced once they anticipate third-party evaluation and potential reward. Overall, this pattern suggests that higher return transfers in the Announced treatment are driven less by reciprocity and more by strategic signaling aimed at attracting Observer rewards.

Finally, Prediction 3 receives less robust support. Investors' transfers are not systematically higher in the Announced treatment compared to the Unannounced treatment. While Figure 3 shows some indication of higher investments when Observers' presence was announced, the effect is not statistically significant. One explanation is that higher-order reasoning about Observers' potential transfers and Trustees' incentives proved too demanding for Investors, echoing Camerer, Ho, and Chong (2004). Alternatively, the attenuation of treatment effects across multiple layers of interaction may have limited Investors' responsiveness.

Taken together, these findings demonstrate that anticipated indirect reciprocity powerfully shapes behavior, but its effects are concentrated at the level of Trustees. Observers reward generosity; Trustees, in turn, strategically increase reciprocity when under observation. Investors, however, remain comparatively unaffected. This layered pattern underscores the importance of institutional signals and expectations of monitoring but also highlights the limits of such mechanisms in complex multi-stage interactions.

My results connect directly to Engel's broader research program. First, they speak to his long-standing interest in conditional rule following: Trustees adapt their behavior when institutionalized observation changes the perceived "rules of the game." Second, they resonate with his emphasis on legitimacy and third-party control: Observers effectively act as informal enforcers. Finally, they illustrate the importance of experimental methods for disentangling the subtle interplay of motives in institutional settings.

Of course, some limitations must be acknowledged. The study relied on a one-shot, laboratory setting with anonymity and strategy method elicitation. While these features increase control and internal validity, they may attenuate or amplify certain effects compared to repeated or naturalistic settings. Future research could explore how repeated interactions, reputation building, or heterogeneous observer roles affect the dynamics documented here. In making these points, I also follow Engel's broader in-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

vitation to embrace the law as an "impossible discipline" (Engel, 2024). My experiment exemplifies his call to combine rigorous empirical methods with sensitivity to institutional context, and to accept complexity rather than simplifying it away. In this sense, the study not only extends the experimental law-and-economics toolkit, but also pays tribute to Engel's intellectual stance of taking risks in search of deeper insight.

G. Conclusion

This study provides evidence that anticipated indirect reciprocity can alter trust game behavior in systematic ways. Observers reward Trustees' generosity but penalize behavior that may appear strategically motivated. Trustees respond strategically to announced observation by increasing return transfers, while Investors remain comparatively unaffected.

The findings highlight both the power and the limits of anticipated indirect reciprocity. Strong effects emerge when the link between action and evaluation is immediate and transparent, as for Trustees, but weaken when reasoning requires higher-order inference, as for Investors. These insights contribute to understanding how institutions relying on external evaluation and monitoring may shape cooperative behavior, even in the absence of formal enforcement mechanisms.

For Engel's broader research agenda, the results underscore the role of conditional rule following and legitimacy in sustaining cooperation. They show how institutionalized observation can change incentives, but also how actors distinguish between genuine and strategic behavior. In this sense, the experiment speaks directly to Engel's interest in the foundations of social order: institutions are effective not only because they alter payoffs, but also because they shape perceptions of legitimacy and authenticity.

From a policy perspective, these findings suggest that institutional designs relying on third-party evaluation must carefully balance transparency with credibility. If evaluation mechanisms make actions appear overly strategic, their legitimacy may be undermined and their effectiveness reduced.

In closing, the study illustrates the value of combining formal models of social preferences with careful experimental design to capture the complex interplay of motives in social interactions. It is a contribution not only to the experimental literature on trust and reciprocity but also to

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the intellectual legacy of Christoph Engel, whose work has consistently illuminated the mechanisms that underlie social order. In celebrating Christoph Engel's 70th birthday, this contribution aims to echo his commitment to risk-taking, interdisciplinarity, and the relentless pursuit of understanding how institutions shape behavior. If my results help sharpen the puzzle of cooperation under observation, they are indebted to the scholarly path Engel has charted.

References

Andreoni, J. (1990) 'Impure altruism and donations to public goods: a theory of warm-glow giving', *The Economic Journal*, 100(401), pp. 464–477.

Berg, J., Dickhaut, J. and McCabe, K. (1995) 'Trust, reciprocity, and social history', *Games and Economic Behavior*, 10(1), pp. 122–142.

Böhm, K.L., Goerg, S.J. and Wasserka-Zhurakhovska, L. (2023) 'How does unethical behavior spread? Gender matters!', *CESifo Working Paper*, No. 10314.

Camerer, C.F., Ho, T.-H. and Chong, J.-K. (2004) 'A cognitive hierarchy model of games', *The Quarterly Journal of Economics*, 119(3), pp. 861–898.

Cerrone, C. and Engel, C. (2019) 'Deciding on behalf of others does not mitigate selfishness: an experiment', *Economics Letters*, 183, p. 108616.

Costa-Gomes, M., Huck, S. and Weizsäcker, G. (2014) 'Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect', *Games and Economic Behavior*, 88, pp. 298–309.

Dufwenberg, M. and Kirchsteiger, G. (2004) 'A theory of sequential reciprocity', *Games and Economic Behavior*, 47(2), pp. 268–298.

Engel, C. (2023) 'How little does it take to trigger a peer effect? An experiment on crime as conditional rule violation', *Journal of Research in Crime and Delinquency*, 60(4), pp. 455–492.

Engel, C., (2024). 'The Law–An Impossible Discipline,' *Review of Law & Economics*, 20(2), pp. 287–322.

Engel, C. and Desmet, P. (2021) 'People are conditional rule followers', *Journal of Economic Psychology*, 85, p. 102367.

Engel, C., Mittone, L. and Morreale, A. (2020) 'Tax morale and fairness in conflict: an experiment', *Journal of Economic Psychology*, 81, p. 102314.

Engel, C., Mittone, L. and Morreale, A. (2024) 'Outcomes or participation? Experimentally testing competing sources of legitimacy for taxation', *Economic Inquiry*, 62(2), pp. 563–583.

Engel, C. and Wasserka-Zhurakhovska, L. (2018) 'Do explicit reasons make legal intervention more effective? An experimental study', *MPI Collective Goods Preprint*, No. 2013/16.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Engel, C. and Zhurakhovska, L. (2014) 'Conditional cooperation with negative externalities: an experiment', *Journal of Economic Behavior & Organization*, 108, pp. 252–260.

Engel, C. and Zhurakhovska, L. (2016) 'When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives', *Applied Economics Letters*, 23(16), pp. 1157–1161.

Engel, C. and Zhurakhovska, L. (2017) 'You are in charge: Experimentally testing the motivating power of holding a judicial office', *The Journal of Legal Studies*, 46(1), pp. 1–50.

Engelmann, D. and Fischbacher, U. (2009) 'Indirect reciprocity and strategic reputation building in an experimental helping game', *Games and Economic Behavior*, 67(2), pp. 399–407.

Falk, A. and Fischbacher, U. (2006) 'A theory of reciprocity', *Games and Economic Behavior*, 54(2), pp. 293–315.

Falk, A., Fehr, E. and Fischbacher, U. (2008) 'Testing theories of fairness – intentions matter', *Games and Economic Behavior*, 62(1), pp. 287–303.

Forsythe, R., Horowitz, J.L., Savin, N.E. and Sefton, M. (1994) 'Fairness in simple bargaining experiments', *Games and Economic Behavior*, 6(3), pp. 347–369.

Fischbacher, U. (2007) 'z-Tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics*, 10(2), pp. 171–178.

Greiner, B. (2015) 'Subject pool recruitment procedures: organizing experiments with ORSEE', *Journal of the Economic Science Association*, 1(1), pp. 114–125.

Grosch, K., Müller, S., Rau S. and Wasserka-Zhurakhovska, L. (forthcoming) 'Ethical dilemmas in leadership: gender differences in dishonesty under responsibility', *The Leadership Quarterly*.

Kleine, M., Langenbach, P. and Zhurakhovska, L. (2016) 'Fairness and persuasion: how stakeholder communication affects impartial decision making', *Economics Letters*, 141, pp. 173–176.

Kleine, M., Langenbach, P. and Zhurakhovska, L. (2017) 'How voice shapes reactions to impartial decision-makers: an experiment on participation procedures', *Journal of Economic Behavior & Organization*, 143, pp. 241–253.

Levine, D.K. (1998) 'Modeling altruism and spitefulness in experiments', *Review of Economic Dynamics*, 1(3), pp. 593–622.

Nowak, M.A. and Sigmund, K. (1998) 'Evolution of indirect reciprocity by image scoring', *Nature*, 393, pp. 573–577.

Rabin, M. (1993) 'Incorporating fairness into game theory and economics', *American Economic Review*, 83(5), pp. 1281–1302.

Selten, R. (1967) 'Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments', in Sauermann, E. (ed.) *Beiträge zur experimentellen Wirtschaftsforschung*. Tübingen: Mohr, pp. 136–168.

Stanca, L., Bruni, L. and Corazzini, L. (2009) 'Testing theories of reciprocity: Do motivations matter?', *Journal of Economic Behavior & Organization*, 71(2), pp. 233–245.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Public Goods, Teams and Institutions

Oliver Kirchkamp

We use an experiment to study unverifiable contributions to a public good in a small team of two players with asymmetric productivity. We compare two institutions: In one, players use a simple bidding mechanism to choose a contract before investing. After having chosen ownership of the output, they contribute. In the other institution, players rely on ex-post coordination. First, they contribute and then they play a Nash demand game to split profits. In the second institution, theoretically, coordination could lead to an efficient outcome. The team could follow a convention which would allow players to make efficient equilibrium investments. Such an outcome cannot be obtained in the equilibrium of the first institution, with a predetermined contract. However, we find the opposite. Joint production is higher when ownership is determined before contributions are made.

A. Introduction

Consider two players who invest in a common good. Investments are not verifiable; hence, players can't write contracts contingent on their investments. They might be able to condition contracts on the output. We can call such a situation a prisoners' dilemma or a public goods problem. Following Coase (1937) and, more specifically, Grout (1984), contracts based solely on an ex-ante determined division of joint output may yield underinvestment.

Taking a behavioural perspective, we might hope that ex-post negotiation helps reach efficient contributions. One group of behavioural studies considers the problem of a single player (e.g., Sonnemans, Oosterbeek, and Sloof 2001; Oosterbeek, Sonnemans, and Velzen 2003; R. Sloof, Sonnemans, and Oosterbeek 2004; Ellingsen and Johannesson 2004a). Here, we consider a situation where all members of the team participate in the investment, even though their productivity can be asymmetric.

A setting where both team members can contribute is more similar to Hackett (1993), Hacket (1994), and Gantner, Güth, and Königstein (2001). Gantner, Güth, and Königstein (2001) use a linear and additive

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

production function. They allow players to distribute profits ex-post using either an ultimatum game or a Nash demand game. The linear and additive production function might yield corner solutions, which is indeed what they find in their experiment. Here, to avoid corner solutions, we use a non-linear production function. Unlike Hackett (1993) and Hacket (1994), we do not focus on the incomplete contract aspect of the team interaction. In our experiment, the only source of uncertainty is the choices of the participants.

In this paper, we ask whether ex-post negotiations can help the players find a more efficient allocation. We consider a small team of only two players. Both players invest in a joint production process and then split joint benefits. We compare two institutions: In one (contract), players choose a fixed division of the produced amount, before they decide investments. In the other (hold-up), players first invest, and then play a Nash demand game to divide joint profits.[1]

With selfish and rational players, the institution with a fixed division of the produced amount (contract) is expected to yield underinvestment from the players. Instead, the institution where players first invest and then play a Nash demand game to divide profits (hold-up) has efficient equilibria. The commitment that is inherent in the Nash demand game should allow players to choose efficient investment levels which are not an equilibrium in the institution with a fixed division of the produced amount.

But, as Engel (2006) points out, players are not necessarily selfish and rational. Real players, who are not selfish and who are boundedly rational, might benefit from these two institutions in a different way. In a slightly different context, for unilateral investments, this question has been addressed by Ellingsen and Johannesson (2004a). While Ellingsen and Johannesson (2004a) observe that ex-post negotiation can be beneficial, they also find that players may follow different standards and, thus, miscoordinate. Here, we look at a situation where investments are bilateral. We use the strategy method to elicit the norms that are used by

1 A different possible approach is to implement ex-post negotiation as a bargaining game with alternating offers as, e.g., in Sonnemans, Oosterbeek, and Sloof (2001), Oosterbeek, Sonnemans, and Velzen (2003), R. Sloof, Sonnemans, and Oosterbeek (2004), Davis and Leider (2018), or as an ultimatum game, e.g., Ellingsen and Johannesson (2004b), Randolph Sloof, Oosterbeek, and Sonnemans (2007), Morita and Servátka (2013), Morita and Servátka (2018), Haruvy et al. (2019).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

players directly. We find that players seem to coordinate on an inefficient norm that does not take into account players' prior investments.

In the institution with fixed divisions of profits, we observe the type of the division and the impact of the status quo contract. An "efficient" division provides second-best incentives, an "egalitarian" contract splits the joint production evenly. Although, in equilibrium both players benefit from an efficient contract, weak (less productive) players expect in the experiment to gain more from the egalitarian contract. More productive players instead, bid for a more "efficient" contract which would provide second-best incentives.

B. Experiment

I. Contracts and Hold-ups

In our experiment[2] we compare two treatments:

Hold-up

In the 'hold-up' situation, players choose their investments first. Thereafter, players play a Nash demand game to negotiate a division of their joint profit. We use the strategy method to implement the Nash demand game. Since many equilibria of the Nash demand game are efficient, selfish players could obtain an efficient outcome.

Contracts

In the 'contract' situation, players first determine the terms of a contract, i.e., how to divide the surplus of a joint investment that is to occur at a later stage. Here we restrict ourselves to sharing rules, i.e., to contracts that divide the joint profit according to some pre-specified share. Sharing rules are common and have been used already in the

2 The experiment was implemented in z-Tree (Fischbacher 2007). For the statistical analysis we use R (R Core Team 2025). Bayesian analysis is done with JAGS (Plummer 2003). Data and methods are available at https://www.kirchkamp.de/research/pubGood.html.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

medieval Islamic world in the institution of qirad (see Sapuan 2016). Profit sharing rules can be found today in contracts of suppliers for the automotive industry, in company-community partnerships (e.g., in fair trade) but also in crop-sharing rules.
Given a sharing rule, both players simultaneously choose investments, and then profits are distributed according to the negotiated contract.
To simplify matters, players will only choose between two contracts, one asymmetric efficient contract and one egalitarian contract. This allows us to use the strategy method and to elicit investment levels for both contracts, even if only one of the two contracts is implemented.
Allowing only two possible sharing rules is not unrealistic. Contracting parties often choose only between a small number of contracts. E.g., a majority of farmers in Ohio use either a 50:50 sharing rule or a $^2/_3$:$^1/_3$ sharing rule (Breece and Forster 2001).

In both treatments participants are randomly matched for 20 periods.

Güth et al. (2003) also study an experiment where hold-up is possible and where players can choose a contract. Unlike Güth et al. (2003) we do not ask whether players give up bargaining power to gain more efficiency. While in their experiment asymmetric productivities and bargaining powers are chosen by players, in our experiment bargaining power and productivities are fixed.

In Fehr, Kremhelmer, and Schmidt (2008), players decide sequentially. The investment of a first player is observed by the second player who then determines their investment. Although an egalitarian division of profits would in equilibrium be inefficient, Fehr, Kremhelmer, and Schmidt (2008) find that in their experiment the egalitarian contract yields the highest investments.

Hacket (1994) and Oosterbeek, Sonnemans, and Velzen (2003) both study hold-up situations where, after production, players have no commitment power at all but engage in bargaining with alternating offers. In our hold-up treatment profits are divided with a Nash demand game, i.e., players can commit to a demand. Theoretically this should remove the hold-up problem.

Zhang and Bayer (2024) study a hold-up situation where players use a contest to divide profits. In their experiment, Zhang and Bayer (2024) find contracts to be more efficient than wasteful investments into a contest.

Our experiment is closer to Ellingsen and Johannesson (2004a) who also use a Nash demand game. While they put a focus on the effect of com-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

munication, we, instead, compare the hold-up treatment with a treatment where contracts are possible.

Our research question is also similar to Bruttel and Eisenkopf (2012) who ask whether an unfair contract can enhance welfare over a situation without contracts.

II. Investments and Payoffs

In this experiment we study the situation of two players, A and B. Players differ in productivity and status quo rights. Each player has an own initial endowment M_A=M_B=500. Simultaneously, both players decide to invest a part of their endowment. The joint output $Q(I_A, I_B)$ depends on investments I_A and I_B.

$$Q(I_A, I_B) = S \cdot I_A^{\alpha} \cdot I_B^{\beta} \tag{1}$$

In our experiment, productivities of the two players A and B are α=0.315, β=0.427. The parameter S=21.89.[3] We call player B the one who is more efficient, i.e., the one whose investment has the larger exponent, β=0.427. We call player A the one who is less efficient, i.e., the one whose investment has the smaller exponent, α=0.315. We compare two treatments. One, we call 'hold-up', the other we call 'contracts'.

1. The hold-up treatment

In the 'hold-up' treatment players choose their investments I_A and I_B first. Thereafter, players play a Nash demand game to negotiate a division of their joint profit. Players A and B demand shares D_A and D_B of their joint profit. If the sum of the demands $D_A + D_B$ is smaller or equal than the total profit, then players obtains their demands. Otherwise players obtain nothing. Payoffs are as follows:

$$u_A = \begin{cases} D_A - I_A & \text{if } D_A + D_B \leq M_A + M_B + S \cdot I_A^{\alpha} \cdot I_B^{\beta} \\ -I_A & \text{otherwise} \end{cases} \tag{2}$$

3 These numbers are chosen to exclude multiple equilibria and to maximise the difference between the equilibrium of the 'egalitarian' and the 'efficient' contract.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

$$u_B = \begin{cases} D_B - I_B & \text{if } D_A + D_B \leq M_A + M_B + S \cdot I_A^{\alpha} \cdot I_B^{\beta} \\ -I_B & \text{otherwise} \end{cases} \tag{3}$$

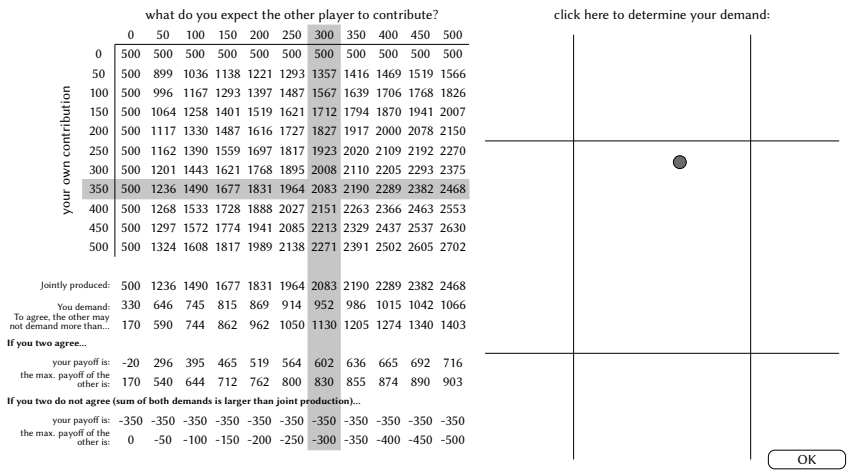
Demand functions are as follows:

$$D_i = \phi_i \left(M_A + M_B + S \cdot I_A^{\alpha} \cdot I_B^{\beta} \right) + \rho_i \left(I_{\text{own}} - I_{\text{other}} \right) \tag{4}$$

Parameter values ϕ_1=ϕ_2=½ and ρ=0 would describe a 'split the revenue' rule. Parameter values ϕ_1=ϕ_2=½ and ρ=½ would describe a 'split revenue–cost' rule.

In each round, players choose their own investment I_i and the two parameters, ϕ_i and ρ_i, of the demand function in Equation (4). Players make these choices simultaneously, i.e., players can't condition their investment on the outcome of the demand game.

To learn more about expectations and in line with the 'contract' treatment, we also ask for the expected contribution of the opponent.



your own contribution	0	50	100	150	200	250	300	350	400	450	500
0	500	500	500	500	500	500	500	500	500	500	500
50	500	899	1036	1138	1221	1293	1357	1416	1469	1519	1566
100	500	996	1167	1293	1397	1487	1567	1639	1706	1768	1826
150	500	1064	1258	1401	1519	1621	1712	1794	1870	1941	2007
200	500	1117	1330	1487	1616	1727	1827	1917	2000	2078	2150
250	500	1162	1390	1559	1697	1817	1923	2020	2109	2192	2270
300	500	1201	1443	1621	1768	1895	2008	2110	2205	2293	2375
350	500	1236	1490	1677	1831	1964	2083	2190	2289	2382	2468
400	500	1268	1533	1728	1888	2027	2151	2263	2366	2463	2553
450	500	1297	1572	1774	1941	2085	2213	2329	2437	2537	2630
500	500	1324	1608	1817	1989	2138	2271	2391	2502	2605	2702
Jointly produced:	500	1236	1490	1677	1831	1964	2083	2190	2289	2382	2468
You demand:	330	646	745	815	869	914	952	986	1015	1042	1066
To agree, the other may not demand more than...	170	590	744	862	962	1050	1130	1205	1274	1340	1403
If you two agree...											
your payoff is:	-20	296	395	465	519	564	602	636	665	692	716
the max. payoff of the other is:	170	540	644	712	762	800	830	855	874	890	903
If you two do not agree (sum of both demands is larger than joint production)...											
your payoff is:	-350	-350	-350	-350	-350	-350	-350	-350	-350	-350	-350
the max. payoff of the other is:	0	-50	-100	-150	-200	-250	-300	-350	-400	-450	-500

Figure 1: The experimental interface for the 'hold-up'-treatment.[4]

4 Participants were guided to specify first their expectation about the other player by choosing a column and their own intended contribution by choosing a row in the table. Then they were asked to specify the parameters of their demand function by clicking in the area on the right. They could adjust their strategies (and see how these adjustments would affect the outcomes on the bottom part of the table) as often as they wanted, before moving to the next stage by clicking 'OK'.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Figure 1 shows the experimental interface for the 'hold-up' treatment. In each round of the experiment, participants would first only see the upper left part of the interface, that is only the table with mutual payoffs. First they had to provide their expectations for the choice of the other player and choose an own contribution. Once they had made their choice, participants would also see the 'demand-calculator' in the right part of the screen as shown in Figure 1. Each click into the field on the right sets the parameters of the demand function. Their own choice would be shown as a red dot, lines would give some orientation and would allow them to start with similar choices in the next period. The vertical axis would describe ϕ_i and the horizontal axis would describe ρ_i. The lower horizontal line denotes ϕ_i=0 while the upper line denotes ϕ=½. The vertical line on the left corresponds to ρ=0 while the line on the right corresponds to ρ=½.

Upon each click participants would see in the bottom part of the left table their actual demand and individual payoffs depending on the choices of the other player. Participants could adjust their own choices and expectations and the parameters of their demand as often as they wanted.

2. The contract treatment

In the treatment with contracts, these contracts are determined by a sharing rule π. Payoffs are as follows:

$$u_A = M - I_A + \pi \cdot S \cdot I_A^{\alpha} \cdot I_B^{\beta} - p \tag{5}$$

$$u_B = M - I_B + (1 - \pi) \cdot S \cdot I_A^{\alpha} \cdot I_B^{\beta} + p \tag{6}$$

Here p is the price paid or received by A for the contract. Different from the hold-up situation, in this contract investments are fully borne by the investing player. Since π is determined ex-ante, players can condition their investment on the terms of the contract.

To keep the experiment simple, we consider only two sharing rules: π=0.5 (the egalitarian contract) and π=0.44 (the efficient contract).[5] If players' investments were the same, I_A=I_B, then with the egalitarian contract π=0.5 their payoffs were the same. With the efficient contract π=0.44

5 For our parameter values the contract that in equilibrium maximises joint surplus is $\pi \approx 0.43995$. Hence, the 'efficient' contract we use in the experiment, π=0.44, is a close approximation.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

their payoffs were different in this situation. However, since in equilibrium under the efficient contract the more productive person invests more, total profit might be higher for both players. Selfish players would prefer the equilibrium outcome under the efficient contract over the outcome under the egalitarian contract.

Before players choose investments, they determine the contract. To do this, players submit bids b_A and b_B for one of the two contracts (either for π=0.44 or for π=0.5).

- If both players bid for the same contract, then this contract is used and the price is p=0.
- If the two players bid for different contracts, then the contract with the higher bid is implemented. The winning bidder pays the own bid to the other player.

$$p = \begin{cases} b_A & \text{if } b_A > b_B \text{ and bids are for different contracts.} \\ -b_B & \text{if } b_A < b_B \text{ and bids are for different contracts.} \\ 0 & \text{otherwise.} \end{cases}$$

- If both bids are the same and for different contracts or if both bids are zero, then a so called 'default' contract is used. Our implementation of the default contract is framed in a neutral way.

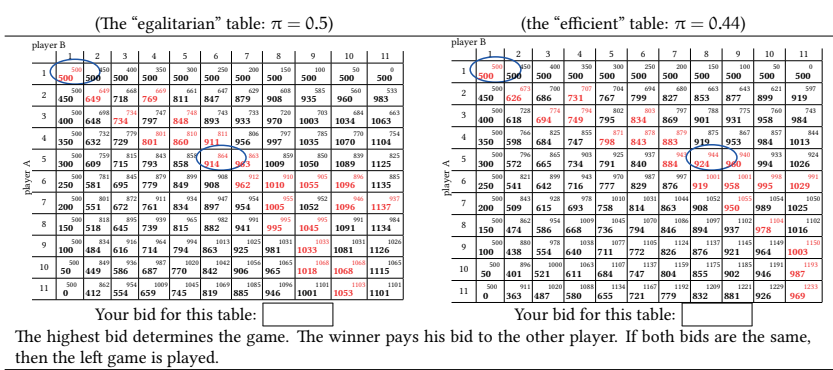


Figure 2: The experimental interface for the 'contract'-treatment.[6]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Figure 2 shows the decision screen in the 'contract' treatment. The screen shows two tables, each table representing one 'contract'. Which of the two tables shows the 'efficient' and which the 'egalitarian' contract was determined randomly for each experimental session. To submit choices (investments) for the two contracts, players click on a row for each of the two tables. Players are also asked to submit expectations by clicking on a column for each table. To determine which table will actually be used, players can make bids for either the left or the right table on the screen. The player with the higher bid wins and determines the game that is played, i.e., essentially whether π=0.44 or whether π=0.5. If players do not bid or submit the same bid as the other player (this is what 23.6 % of all bidders do) then the default contract is implemented. The default contract was determined randomly for each session and known to participants.

Table 1 provides an overview of the treatments.

Table 1: List of treatments.

treatment	obs	participants
contract	20	182
holdup	45	404
total	65	586

C. Hypotheses

I. Hypotheses for the contract game

Hypothesis 1 (Efficient sharing rule): *Both players should prefer the theoretically efficient sharing rule π=0.44.*

In a different game where players decide sequentially, Fehr, Kremhelmer, and Schmidt (2008) find that the egalitarian sharing rule, although not efficient in equilibrium, maximises investments in the experiment.

6 In each table, best replies are shown in red. Nash equilibria are denoted with a blue circle. This information was not provided in the experiment. All numbers in the experiment were shown in black. Also, no information about the values of π and 'egalitarian' and 'efficient' was shown. In about half of our sessions the layout was reversed, i.e., the table that we show here on the right was shown on the left and vice versa.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Engel and Kurschilgen (2013) explain that customs and normative expectations shape behaviour. The default sharing rule in our game can be seen as such a normative expectation. This rule is, however, an expectation that players can easily override by making a very small bid.

Hypothesis 2 (Irrelevance of default): *We expect that the default sharing rule has no influence on the sharing rule that is chosen by the players.*

II. Hypotheses for the hold-up game

Since strategies can be characterised as cost shares ρ and revenue shares ϕ_A and ϕ_B, Figure 3 shows how these parameters determine investments and profits in equilibrium. Since only eleven different investment levels are allowed, the contour lines are not smooth. Small values of ρ lead to underinvestment. Larger values of ρ allow sharing investment costs and help obtain an efficient outcome.

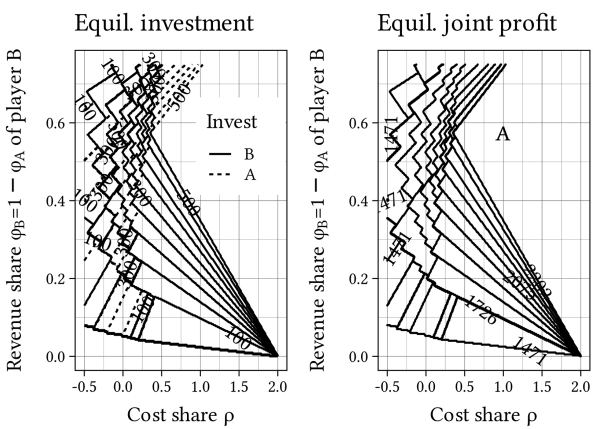


Figure 3: Contours of equilibrium investments (left) and joint profits (right) in the hold-up game in the experiment depending on fairness norms.

To obtain efficient investments in equilibrium, ρ can be as small as ρ=0.347, provided that the strong player obtains a small extra incentive to invest to compensate for the weak players less efficient technology, i.e., the strong player's share is 0.574 and the weak player's share is 0.426.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

If, instead, players use a 50:50 norm, they have to take into account more of their investments in the Nash demand game, i.e., to obtain equilibrium efficient investments with ϕ=½ we need $\rho \geq 0.56$.

Hypothesis 3 (Efficient norms in the hold-up game): *We expect that participants coordinate on fairness norms, expressed as ϕ and ρ, that allow them to obtain the maximum joint profit, i.e., norms in range 'A' in Figure 3.*

If players are interested in a sharing norm which is robust to different levels of ρ then they should give extra incentives to the strong player.

Hypothesis 4 (Strongly efficient norms in the hold-up game): *The strong player's share will be larger than 1/2 and close to 0.574.*

In the hold-up treatment players can make efficient investments in an equilibrium. Equilibrium investments in the contract treatment are lower.

Hypothesis 5 (Efficient investment in the hold-up game): *We expect that participants obtain higher joint profits in the hold-up treatment than in the contract treatment.*

III. Comparison of both treatments

If players select the efficient equilibrium in the Nash-demand game they will choose higher investment levels and, hence, have higher joint profits than in the equilibrium of the contract game.

Hypothesis 6 (More efficiency in the hold-up treatment): *We expect that investments and payoffs in the hold-up treatment are higher than those in the contract treatment.*

D. Results

I. Some descriptives

Of the 586 participants 326 were female, 206 were male, for 54 we do not know their gender. The left graph in Figure 4 shows the distribution of the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

age of the participants. The right graph in Figure 4 shows a histogram of the perceived difficulty of the experiment. Apparently, most participants found the experiment easy to understand.

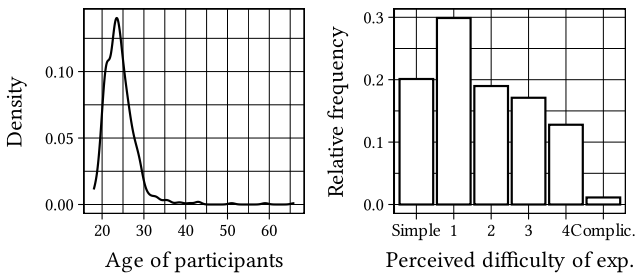


Figure 4: Age of participants and perceived difficulty of the experiment

II. Results for the contract game

1. Bids for contracts

According to Hypothesis 1, we expect both players to prefer the efficient sharing rule $\pi=0.44$. Figure 5 shows the distribution of bids for egalitarian and efficient contracts.

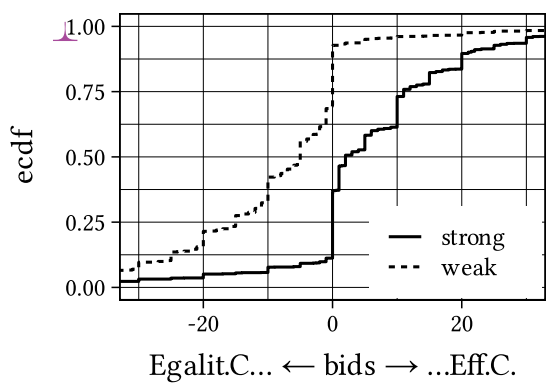


Figure 5: Distribution of bids

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Positive bids in Figure 5 denote bids for the efficient contract (π=0.44). Negative bids denote bids for the egalitarian contract (π=0.5). In line with Hypothesis 1, the strong player bids most of the time for the efficient contract (most of the dotted distribution is to the right of 0). Different from what we expect in Hypothesis 1, the weak player bids most of the time for the egalitarian contract (most of the solid line distribution is to the left of 0).

To study this more formally, we estimate the following mixed effects regression

$$b_{ijk} = \beta_{\text{weak}} d_{\text{weak}} + \beta_{\text{strong}} d_{\text{strong}} + \eta_{kj} + \nu_j + u_{ijk} \tag{7}$$

where η_{kj} is a random effect for player k in group j, ν_j is a random effect for the matching group in the experiment, u_{ijk} is the residual, d_{weak} and d_{strong} are dummies for the two types of players and b_{ik} is player i's bid for the efficient contract in round k.

We estimate β_{weak}=-8.68, CI_{95}=[-11.7, -5.59]. The odds of $\beta_{\text{weak}} < 0$ are larger than 1000:1. We have, hence, "very strong" evidence that the weak player prefers the egalitarian contract.[7]

2. Effect of the default contract

According to Hypothesis 2, we expect the default sharing rule to have no influence on the choice of the sharing rule. Indeed, for the strong player, the effect of a default efficient contract on the bid for the efficient contract is, in line with Hypothesis 2, only -0.906, CI_{95}=[-7.32, 5.6].[8] For the weak player the effect of a default efficient contract on the bid for the efficient contract is 3.53, CI_{95}=[-2.98, 9.94]. The odds that the effect of the default are positive are 6.06:1, i.e., we have at least "positive" evidence that the weak player's bids for the egalitarian contract are larger if the egalitarian contract is the default.

7 We follow the terminology suggested by Kass and Raftery (1995): odds $\in [1:1, 2.72:1]$: only anecdotal evidence, odds $\in [2.72:1, 20.1:1]$: positive evidence, odds $\in [20.1:1, 148:1]$: strong evidence, odds $\in [148:1, \infty:1]$: very strong evidence.

8 We use a specification similar to Equation (7), with an interaction for the default contract.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. Results for the hold-up game

1. Norms

According to Hypothesis 3, we expect that participants coordinate on fairness norms that are compatible with maximising joint profits. These norms are shown as range 'A' in Figure 3. Figure 6 shows the norms that we observe in the experiment. Contour lines show the distribution of norms in the experiment. The shaded area shows the range of norms that would yield a profit maximising equilibrium.

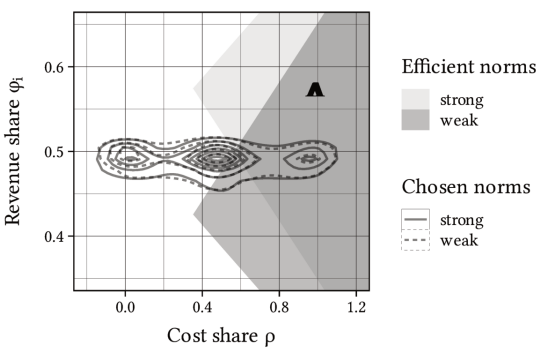


Figure 6: Norms in the hold-up game.

Different from Hypothesis 3 the majority of players does not take into account investments in a way that would support efficient investments in equilibrium.

We find that the norms of weak and strong players are very similar. There is no asymmetry as expected in Hypothesis 4. We estimate[9] an average ϕ_i=0.46, CI_{95}=[0.45, 0.47]. We do not find a substantial difference between ϕ_A and ϕ_B. The odds that $\phi_B > \phi_A$ are only 1.86:1. We estimate an average ρ=0.444, CI_{95}=[0.407, 0.482]. Again, we do not find a substantial difference between ρ_A and ρ_B. The odds that $\rho_B > \rho_A$ 1:2.04. We can neither confirm Hypothesis 3 nor Hypothesis 4.

9 We follow the specification of Equation (7), only with ρ_{ijk} and ϕ_{ijk} as dependent variables.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2. Efficiency of contracts versus hold-up

According to Hypothesis 5, we should expect higher joint profits in the hold-up treatment.

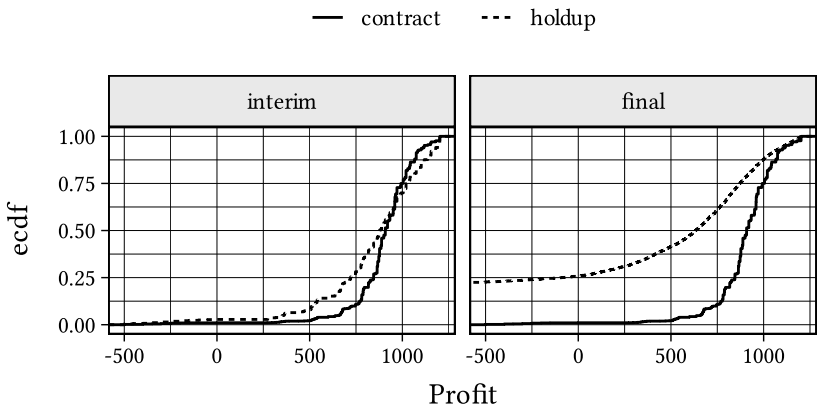


Figure 7: Distribution of produced output in the hold-up and contract treatment.

The left part of Figure 7 shows the distribution of the amounts produced. We see that the amount produced in the hold-up treatment is slightly lower than the amount produced in the contract treatment. While no agreement is necessary after production in the contract treatment, only 79% of the players find an agreement in the hold-up treatment. This lack of agreement in the hold-up treatment leads to even lower profits. The right part of Figure 7 shows the distribution of final profits. Contrary to Hypothesis 5, actual profits in the hold-up treatment are much lower than profits in the contract treatment.

E. Discussion

In this paper, we have compared two treatments: one treatment describes a hold-up situation where players first invest; then they divide the proceeds of their joint investment with the help of a Nash demand game. The other treatment describes a contract situation where players negotiate a contract before their investment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Rational players should be able to coordinate on an efficient equilibrium in the hold-up situation. In equilibrium, payoffs can be higher in the hold-up treatment than in the contract treatment. However, in the experiment players fail to coordinate in the hold-up game. This failure to coordinate primarily makes the hold-up game less efficient than the contract game.

We expected players to adopt different norms for weak and strong players. Instead, the norms used in the experiment were rather egalitarian and did not seem to depend on the players' strength. Such egalitarian norms made coordination on efficient investments more difficult.

In the contract treatment, we compared two contracts: an asymmetric one and a symmetric one. In equilibrium, the asymmetric contract should provide better incentives for the more productive player, so that, in equilibrium, both players would earn more with the asymmetric contract than they would with the symmetric one. In the experiment, however, the less productive player preferred the symmetric contract. To some degree, this choice can be justified through concerns for efficiency. To some extent, however, the less productive player clearly overestimates the investment made by the other player.

In this paper, we explored how players' contributions to a public good depend on institutions. Perhaps unsurprisingly, we found that contributions often deviate from equilibrium predictions. Importantly, we found that the direction of these deviations depends on the institution. Consequently, institutions that are theoretically less efficient in equilibrium sometimes proved to be more efficient in the experiment. Here, we did not include social preferences in the derivation of theoretical benchmark solutions. Future research should delve into the specific cognitive and behavioral factors that influence players' deviations from equilibrium predictions. Employing experimental designs that isolate players' expectations and concerns for overall efficiency could provide valuable insights. Additionally, understanding these dynamics could have significant implications for the design of institutions and policies aimed at enhancing efficiency in public goods scenarios.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

References

Breece, Donald J., and D. Lynn Forster. 2001. 'Crop Share Leasing in Ohio'. Extension Fact Sheet FR-0005–01. The Ohio State University.

Bruttel, Lisa, and Gerald Eisenkopf. 2012. 'No contract or unfair contract: What's better?' Journal of Behavioral and Experimental Economics 41 (4): 384–90.

Coase, R. H. 1937. 'The Nature of the Firm'. Economica 4 (16): 386–405.

Davis, Andrew M., and Stephen Leider. 2018. 'Contracts and Capacity Investment in Supply Chains'. Manufacturing & Service Operations Management 20 (3): 403–21.

Ellingsen, Tore, and Magnus Johannesson. 2004a. 'Is There a Hold-up Problem?' Scandinavian Journal of Economics 106 (3): 475–94.

–. 2004b. 'Promises, Threats and Fairness'. The Economic Journal 114 (495): 397–420.

Engel, Christoph. 2006. 'Social Dilemmas Revisited from a Heuristics Perspective'. In Heuristics and the Law, edited by Gerd Gigerenzer and Christoph Engel. Dahlem Workshop Reports. The MIT Press.

Engel, Christoph, and Michael Kurschilgen. 2013. 'The Coevolution of Behavior and Normative Expectations: An Experiment'. American Law and Economics Review 15 (2): 578–609.

Fehr, Ernst, Susanne Kremhelmer, and Klaus M. Schmidt. 2008. 'Fairness and the Optimal Allocation of Property Rights'. The Economic Journal 118 (531): 1262–84.

Fischbacher, Urs. 2007. 'Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments'. Experimental Economics 10 (2): 171–78.

Gantner, Anita, Werner Güth, and Manfred Königstein. 2001. 'Equitable Choices in Bargaining Games with Joint Production'. Journal of Economic Behavior & Organization 46 (2): 209–25.

Grout, Paul A. 1984. 'Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach – The Econometric Society'. Econometrica 52 (2): 449–60.

Güth, Werner, Radosveta Ivanova-Stenzel, Matthias Sutter, and Hannelore Weck-Hannemann. 2003. 'Investment and Bargaining in Joint Ventures: A Family Decision Making Experiment'. Journal of Institutional and Theoretical Economics (JITE) 159 (2): 323–23.

Hacket, Steven C. 1994. 'Is Relational Exchange Possible in the Absence of Reputations and Repeated Contact'. The Journal of Law, Economics and Organization 10 (2): 360–89.

Hackett, Steven C. 1993. 'Incomplete Contracting: A Laboratory Experimental Analysis'. Economic Inquiry 31 (2): 274–97.

Haruvy, Ernan, Elena Katok, Zhongwen Ma, and Suresh Sethi. 2019. 'Relationship-Specific Investment and Hold-up Problems in Supply Chains: Theory and Experiments'. Business Research 12 (1): 45–74.

Kass, Robert E., and Adrian E. Raftery. 1995. 'Bayes Factors'. Journal of the American Statistical Association 90 (430): 773–95.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Morita, Hodaka, and Maroš Servátka. 2013. 'Group Identity and Relation-Specific Investment: An Experimental Investigation'. European Economic Review 58 (C): 95–109.

–. 2018. 'Investment in Outside Options as Opportunistic Behavior: An Experimental Investigation'. Southern Economic Journal 85 (2): 457–84.

Oosterbeek, Hessel, Joep Sonnemans, and Susan van Velzen. 2003. 'The Need for Marriage Contracts: An Experimental Study'. Journal of Population Economics 16 (3): 431–53.

Plummer, Martyn. 2003. 'JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling'. 3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria 124 (April).

R Core Team. 2025. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Sapuan, Noraina Mazuin. 2016. 'An Evolution of Mudarabah Contract: A Viewpoint From Classical and Contemporary Islamic Scholars'. Procedia Economics and Finance, 7th International Economics & Business Management Conference (IEBMC 2015), 35 (January): 349–58.

Sloof, Randolph, Hessel Oosterbeek, and Joep Sonnemans. 2007. 'Does Making Specific Investments Unobservable Boost Investment Incentives?' Journal of Economics & Management Strategy 16 (4): 911–42.

Sloof, R., J. Sonnemans, and H. Oosterbeek. 2004. 'Specific Investments, Holdup, and the Outside Option Principle'. European Economic Review 48 (December).

Sonnemans, Joep, Hessel Oosterbeek, and Randolph Sloof. 2001. 'On the Relation Between Asset Ownership and Specific Investments*'. The Economic Journal 111 (474): 791–820.

Zhang, Sookie Xue, and Ralph-Christopher Bayer. 2024. 'Revenue Sharing Agreements and Hold-up Problems in Joint Projects: Theory and Experiments'. Journal of Economic Behavior & Organization 228 (December): 106774.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III.
Judges and Decision-Making

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

„Catch Me If You Can“
Die Flucht als Straftat im Auslieferungsverfahren

Markus Englerth

A. Einleitung

Literatur und Film haben seit jeher ein Herz für Straftäter auf der Flucht. Mögen diese schuldig sein wie der verurteilte Einbrecher Frank Morris in „Flucht aus Alcatraz“, gänzlich zu Unrecht inhaftiert wie Edmond Dantès in „Der Graf von Monte Cristo“, oder sich irgendwo zwischen diesen beiden Polen bewegen wie „Thelma & Louise“ oder Dimitri Fjodorowitsch in „Die Brüder Karamasow“ [1] – das Publikum sympathisiert meist unweigerlich mit dem ungebrochenen Freiheitsdrang des von der Justiz verfolgten Protagonisten.

Auch dem deutschen Strafrecht ist diese Empfindung nicht ganz fremd. Es pönalisiert weder die Selbstbefreiung eines Gefangenen noch sonstige (gewaltfreie) Formen der Flucht. Hingegen stellt in zahlreichen anderen Ländern schon der Versuch des Beschuldigten oder des bereits Verurteilten, sich der Strafjustiz zu entziehen, eine eigenständige Straftat dar.

Diese Diskrepanz in der Behandlung von Fluchthandlungen hat Konsequenzen für den Auslieferungsverkehr, die bislang nur wenig diskutiert werden. Das muss überraschen, steht doch ein solcher Fluchtversuch sehr häufig am Anfang eines Auslieferungsverfahrens und bildet oftmals sogar dessen *raison d'être* – nicht umsonst wird die gesuchte Person als „Verfolgter“ bezeichnet.

Der vorliegende Beitrag wirft zunächst einen Blick auf die unterschiedliche strafrechtliche Behandlung des Themas Flucht in Deutschland und in anderen Staaten (B), stellt danach die relevanten auslieferungsrechtlichen Prinzipien vor (C) und versucht sich schließlich an einer Kate-

1 Dimitri ist zwar unschuldig am Tod seines Vaters, für den er letztlich verurteilt wird, hat aber zumindest dessen Hausdiener Grigorij lebensgefährlich verletzt. Trotzdem heißt sogar der fromme Karamasow-Bruder Alexej die Dimitris Flucht in den letzten Kapiteln gut.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

gorisierung der hieraus folgenden Konsequenzen für die Überstellung von Personen, die sich der Strafjustiz entzogen haben (D). Er endet mit einigen Bemerkungen zu verwandten Konstellationen und offenen Fragen (E).

B. Nationale Divergenzen in der Behandlung der Flucht

I. Begriff der Flucht

Der Begriff der Flucht wird in diesem Beitrag – als Sammelbegriff für verschiedene Verhaltensweisen – denkbar weit verstanden. Gemeint ist nicht nur das Entweichen aus physischem Gewahrsam (wie z. B. bei einem Gefängnisausbruch), sondern jede Handlung mit dem Ziel, sich dem Zugriff des Staates zu entziehen, um der Strafverfolgung oder der Vollstreckung einer Strafe zu entgehen. Erfasst ist also auch der Fall, dass der Täter zur Vermeidung von Strafe untertaucht, sich ins Ausland absetzt oder von dort nicht zurückkehrt. Auch sonstige evasive Verhaltensweisen – wie die Missachtung von Ladungen, gerichtlichen Weisungen oder Kautionsauflagen – werden in die Betrachtung einbezogen.

II. Grundsätzliche Straffreiheit im deutschen Recht

Der Versuch, sich der eigenen Bestrafung durch Flucht zu entziehen, ist – für sich allein genommen – nach deutschem Recht straflos. Diese Wertung kommt besonders deutlich in § 120 StGB zum Ausdruck, der die (gewaltlose) *Selbst*befreiung eines Gefangenen aus der Haft von der grundsätzlichen Strafbarkeit der Gefangenenbefreiung ausnimmt.[2] Erst recht bleibt straflos, wer – auf freiem Fuß befindlich – untertaucht, sich ins Ausland absetzt oder anderweitig dem staatlichen Zugriff entzieht. Selbst ein Beschuldigter, gegen den bereits ein Haftbefehl wegen Fluchtgefahr erlassen und nur gegen anwesenheitssichernde Auflagen (§ 116 StPO) außer Vollzug gesetzt wurde, muss keine zusätzliche Bestrafung befürchten, wenn er unter Missachtung dieser Auflagen doch Reißaus nimmt.[3]

2 Näher dazu *Ostendorf* NStZ 2007, 313.

3 Etwas anders kann es sich nach Abschluss des Strafverfahrens verhalten. Verstößt ein bereits verurteilter Straftäter im Rahmen der Führungsaufsicht gegen die ihm nach § 68b StGB vom Gericht erteilte Weisung, seinen Wohn- oder Aufenthaltsort nicht zu

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die Gründe für die Privilegierung von Fluchthandlungen werden kontrovers diskutiert.[4] Teils wird die Straffreiheit mit dem strafverfahrensrechtlichen Nemo-tenetur-Grundsatz in Verbindung gebracht: Ebenso wie niemanden die Pflicht treffe, an der eigenen Überführung mitzuwirken, sei auch niemand verpflichtet, die Flucht zu unterlassen.[5] Teils wird der Kriminalisierungsverzicht auf Grundlage eines gemischtfunktionalen Schuldmodells mit dem „Zusammentreffen generalpräventiver Kompensationstechniken, sozialpsychologischer Betrachtungen und liberal-humaner Maximen“ erklärt.[6] Der Staat halte sich mit Fluchtsanktionen nicht nur zurück, weil er das Bestreben des Täters, der Strafe zu entgehen, als zumindest verständlich anerkenne, sondern auch und vor allem, weil er sich zutraue, diesem Bestreben in aller Regel effektiv entgegentreten zu können.[7] Wieder andere sehen den Verzicht auf eine Strafbarkeit der Flucht als Facette eines Prinzips der straflosen Selbstbegünstigung, „das sich auf das allgemeine Persönlichkeitsrecht, letztlich auf die Respektierung der Menschenwürde (Art. 2 I i.V.m. Art. 1 I GG) gründet: Das natürliche Bestreben des Menschen, Strafleid von sich abzuwehren, [soll] nicht zum Anknüpfungspunkt werden, neues Strafleid zu verhängen oder sonstige Nachteile damit zu verknüpfen“.[8]

Das Für und Wider dieser – teils durchaus miteinander kompatiblen – Erklärungsmodelle mag hier dahinstehen. Auch die höchstrichterliche Rechtsprechung erkennt jedenfalls an, dass „das Gesetz aus humanen Beweggründen dem Freiheitsdrang eines Menschen glaubte, Rücksicht schenken zu sollen“.[9] Nach Ansicht des BGH folgt aus dieser gesetzgeberischen Wertung sogar ein Verbot, die Flucht des Täters im Rahmen der Strafzumessung als Schärfungsgrund zu berücksichtigen, sofern da-

---

verlassen, so kommt insofern eine Strafbarkeit nach § 145a StGB in Betracht. Zu Recht weist allerdings *Schneider* darauf hin, dass auch in diesem scheinbaren Ausnahmefall nicht die Flucht an sich bestraft werde, sondern nur „die damit einhergehende Zweckvereitelung der Führungsaufsicht“, vgl. *Schneider*, Grund und Grenzen des strafrechtlichen Selbstbegünstigungsprinzips auf der Basis eines generalpräventiv-funktionalen Schuldmodells (1991), S. 344, Fn. 43. Kritisch zur Sinnhaftigkeit dieser Regelung und ihrer jüngsten Verschärfung SK-StGB/*Wolters*, 9. Aufl. (2019), § 145a Rn. 1.

4 Ausführlich hierzu *Schneider*, a. a. O. (Fn. 3), S. 187 ff. und 336 ff.

5 So etwa SK-StGB/*Wolters* § 120 Rn. 14.

6 So etwa *Schneider*, a. a. O. (Fn. 3), S. 386.

7 Vgl. *Schneider*, a. a. O. (Fn. 3), S. 385 ff.

8 *Ostendorf* NStZ 2007, 313 (316); ähnlich noch LK-StGB/*von Bubnoff*, 12. Aufl. (2019), § 120 Rn. 2a.

9 BGH NJW 1962, 2260 (2261) in Anschluss an RGSt. 3, 140, 141.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

durch nicht ausnahmsweise neues Unrecht verwirklicht wurde oder der Täter neue Ziele verfolgte, die ein schlechtes Licht auf ihn werfen.[10]

III. Strafbarkeit der Flucht in anderen Rechtsordnungen

Keineswegs alle Länder gehen im Hinblick auf die Straffreiheit fluchtbezogener Selbstbegünstigungshandlungen derart weit. Zu den eher raren Beispielen, die wie Deutschland ebenfalls nicht einmal das Entweichen aus staatlichem Gewahrsam unter Strafe gestellt haben, zählen z. B. Österreich[11], die Schweiz[12], die Niederlande[13], Schweden[14] und Kroatien[15]. Dahinter stehen dann oftmals ganz ähnliche Erwägungen wie hierzulande. So wird beispielsweise das Diktum des österreichischen OGH, wonach es „keine Pflicht des verdächtigen Flüchtenden, nicht zu flüchten"[16] gebe, in der Literatur als Rekurs auf das Nemo-tenetur-Prinzip verstanden.[17]

Indes ist oftmals eine Selbstbefreiung von Gefangenen selbst in denjenigen Ländern strafbar, die ansonsten keine eigenständige Pönalisierung

10 BGHR StGB § 46 II Nachtatverhalten 13, 17; NStZ 2011, 512 (513).

11 § 300 Abs. 2 des österreichischen StGB nimmt den Gefangenen selbst von der Strafbarkeit wegen Gefangenbefreiung nach Abs. 1 aus.

12 In der Schweiz ist die Gefangenenbefreiung in eigener Sache ebenfalls nicht strafbar, vgl. Art. 310 StGB.

13 Vgl. § 367 des niederländischen StGB (*Wetboek van Strafrecht*), der nur die Gefangenenbefreiung durch Justizbedienstete unter Strafe stellt. Zuletzt gab es in den Niederlanden allerdings Bestrebungen, auch die Flucht des Gefangenen selbst unter Strafe zu stellen, vgl. NL Times, „Escaping from prison is not a crime in the Netherlands, but that could soon change", 15.05.2024, >https://nltimes.nl/2024/05/15/escaping-prison-crime-netherlands-soon-change?utm_source=chatgpt.com< (zuletzt abgerufen: 10.09.2025).

14 Vgl. Kapitel 17, § 11 f. des schwedischen StGB (*Brottsbalken*), wonach Unterstützungshandlungen Dritter sanktioniert werden.

15 Art. 306 und 307 des kroatischen Strafgesetzbuchs stellen nur das Entweichen unter Einsatz eines qualifizierten Nötigungsmittels unter Strafe.

16 OGH 1 Ob 97 /15 v Zak 2015 /540, 299 = JBl 2015, 587 = ÖJZ EvBl 2015 /147, 1037 (Brenn/Nordmeyer).

17 *Pfeifer*, „Nemo tenetur"-Grundsatz: Recht auf selbstbestimmte Kooperation (2024), S. 27 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

von Fluchthandlungen vorsehen. Zu dieser Gruppe gehören etwa Dänemark[18], Japan[19], Italien[20], die Türkei[21] und Israel[22].

Viele Rechtsordnungen – vor allem, aber keineswegs nur, in der angelsächsischen Welt – wählen einen noch repressiveren Ansatz und behandeln sogar Bestrebungen des auf freiem Fuß befindlichen Beschuldigten oder Verurteilten, sich dem staatlichen Zugriff zu entziehen, als eigenständige Straftat.

Exemplarisch stellt 18 U.S.C. § 3146 es in den Vereinigten Staaten unter Strafe, wenn ein haftverschonter Angeklagter Gerichtstermine versäumt oder einer Ladung zum Haftantritt nicht nachkommt. Zuwiderhandlungen können – abhängig von der Ausgangsstraftat – mit einer *zusätzlichen* Freiheitsstrafe von bis zu 10 Jahren geahndet werden. Daneben droht 18 U.S.C. § 1073 Flüchtigen, die sich in einen anderen Bundesstaat oder ins Ausland absetzen, mit bis zu fünf Jahren Freiheitsentzug.

Auch Kanada stellt nicht nur den Ausbruch aus einer Haftanstalt, sondern auch typischerweise mit der Flucht verbundene Verhaltensweisen eines auf freiem Fuß befindlichen Beschuldigten – z. B. das Versäumen von Gerichtsterminen – unter Strafe (vgl. Section 145 Criminal Code, R.S.C. 1985, c. C-46).

Eine ähnliche Regelung trifft Section 67A des südafrikanischen Criminal Procedure Act 51 of 1977. Und im australischen Bundesstaat New South Wales droht Art. 79 des Bail Act 2013 einem Beschuldigten, der unter Missachtung seiner Kautionsauflagen unentschuldigt einen Gerichtstermin versäumt, bis zu drei Jahren zusätzlicher Haft an. Auch im Vereinigten Königreich macht sich strafbar, wer gegen Kaution aus der Haft entlassen wurde und es anschließend versäumt, einer gerichtlichen Ladung Folge zu leisten.[23]

Auch Section 174A des indischen Strafgesetzbuches stellt die Nichtbefolgung einer gerichtlichen Vorladung durch den in Freiheit befindlichen Angeklagten unter Strafe. Der Oberste Gerichtshof Indiens hat

18 Vgl. § 124 Abs. 1 des dänischen StGB (*Straffeloven*).

19 vgl. Art. 97 des japanischen StGB (Gesetz Nr. 45 vom 24. April 1907).

20 Vgl. Art. 385 des italienischen StGB (*Codice penale*).

21 Vgl. Art. 292 (1) des türkischen StGB (*Türk Ceza Kanunu)*.

22 Vgl. Art. 257 des Strafgesetzes 5737–1977.

23 Vgl. Section 6 des *Bail Act of 1976:* Auf dieser Rechtsgrundlage wurde etwa Julian Assange festgenommen, vgl. High Court of Justice (Queen's Bench Division, Administrative Court), Government of the United States of America v. Julian Paul Assange, Urt. v. 10.12.2021 – CO/150/2021, [2021] EWHC 3313 (Admin), Rn. 5.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

inzwischen sogar geurteilt, dass das Missachten einer solchen Vorladung auch dann eine eigenständige Straftat darstellt, wenn die Vorladung später für nichtig erklärt oder der Angeklagte von den ihr zugrundeliegenden Vorwürfen freigesprochen wird.[24]

Sogar in der Europäischen Union gibt es etliche Länder, die fluchttypische Selbstbegünstigungshandlungen selbst dann pönalisieren, wenn sich der Täter – noch oder wieder – auf freiem Fuß befindet. So stellt Art. 434–27 ff. des französischen Code Pénal es nicht nur unter Strafe, wenn sich ein Gefangener aus der Haft oder dem polizeilichen Gewahrsam befreit, sondern kriminalisiert z. B. auch die Flucht des Beschuldigten nach Bekanntgabe eines Untersuchungshaftbefehls oder des bereits Verurteilten während eines Hafturlaubs.[25] Ähnliche Regelungen finden sich u. a. in Polen[26] und Spanien[27].

Warum einige Länder dem Flüchtenden gegenüber großzügig sind, während andere ihm mit der vollen Härte des Strafrechts begegnen, kann hier nicht ansatzweise erschöpfend diskutiert werden.[28] Klar ist jedoch, dass sich aus diesen Divergenzen Probleme für den zwischenstaatlichen Auslieferungsverkehr ergeben können.

C. Der auslieferungsrechtliche Rahmen

Dies hat mit zwei elementaren Strukturprinzipien des Auslieferungsrechts zu tun – dem Grundsatz der beiderseitigen Strafbarkeit und dem Spezialitätsgrundsatz – deren Gehalt zunächst skizziert werden soll.

24 Supreme Court of India, Urt. v. 2. 1. 2025 – Criminal Appeal No. 4359 of 2024, 2025 INSC 21 (Daljit Singh v. State of Haryana & Anr.).

25 Vgl. Art. 434–28 Nr. 3 bzw. 434–29 Nr. 2 *Code Pénal*.

26 Art. 242 des polnischen StGB (*Kodeks karny*).

27 Art. 468 des spanischen StGB (*Código Penal*).

28 Dass sich dies – wie offenbar *Schneider*, a. a. O. (Fn. 3), S. 188, meint – primär damit erklären lasse, dass diese Strafrechtssysteme liberal-individualistisch orientiert seien, während „der Pönalisierung der Selbstbefreiung ein autoritär-etatistischer Zug kriminalpolitischen Denkens eigen" sei, muss allerdings stark bezweifelt werden. Ein Muster, wonach Länder, welche Fluchthandlungen kriminalisieren, dem Staat generell eine stärkere Position gegenüber dem Individuum zubilligen als Systeme, die dies nicht tun, ist nicht erkennbar.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I. Grundsatz der beiderseitigen Strafbarkeit

Gem. § 3 I IRG ist die Auslieferung nur zulässig, wenn sie wegen einer Tat begehrt wird, die – ggf. nach sinngemäßer Umstellung des Sachverhalts – auch in Deutschland den Tatbestand eines Strafgesetzes verwirklichen würde. Historisch ist dieser „Grundsatz der beiderseitigen Strafbarkeit“[29] eng verwandt mit dem völkerrechtlichen Reziprozitätsgedanken.[30] Inwieweit er völker- bzw. verfassungsrechtlich zwingend ist, wird sehr kontrovers diskutiert, mag aber für die Zwecke der vorliegenden Abhandlung dahinstehen.[31] Die deutschen Gerichte haben ihn jedenfalls *de lege lata* als Zulässigkeitsvoraussetzungen in den meisten Fällen schon von Amts wegen zu beachten.

Geltung beansprucht er nämlich nicht nur für den vertragslosen Bereich. Auch in allen bedeutsamen Auslieferungsabkommen hat er in ähnlicher Form Ausdruck gefunden.[32] Für den unionsinternen Auslieferungsverkehr aufgrund eines Europäischen Haftbefehls gilt er ebenfalls – allerdings mit den in § 81 Nr. 4 IRG bezeichneten und praktisch durchaus weitreichenden Einschränkungen.

Beiderseitige Strafbarkeit bedeutet nicht, dass die deutsche Rechtsordnung die Tat, derentwegen die Auslieferung begehrt wird, rechtlich identisch kategorisieren oder bewerten müsste wie der ersuchende Staat. Lässt sich der im Auslieferungsersuchen konkret bezeichnete Sachverhalt aber unter gar keinen Tatbestand des deutschen Strafrechts subsumieren

29 BT-Drs. 9/1338, 36; kritisch zum Begriff *Kubiciel* in Ambos/Rackow/König, Rechtshilferecht in Strafsachen, 2. Aufl. (2020), § 3 IRG Rn. 23, der mit guten Gründen vorschlägt, stattdessen vom „Grundsatz der hypothetischen Strafrechtswidrigkeit im Auslieferungsstaat“ zu sprechen.

30 *Kubiciel* in Ambos/Rackow/König § 3 IRG Rn. 23; *Oehler* ZStW 96 (1984), 557 (557); skeptisch *Vogel/Burchard* in: Grützner/Pötz, Internationaler Rechtshilfeverkehr in Strafsachen, 89. Lieferung, 5/2009, § 3 IRG Rn. 18 f.

31 Näher dazu *Vogel/Burchard* in: Grützner/Pötz § 3 IRG Rn. 8 ff.; anders *Schünemann* ZRP 2003,185 (188).

32 Vgl. etwa Art. 2 I EuAlÜbk; Art. 2 I 1 Auslieferungsvertrag zwischen der Bundesrepublik Deutschland und den Vereinigten Staaten von Amerika v. 20. 6. 1978 (BGBl. 1980 II S. 646, 1300) i.d.F. des Zusatzvertrags v. 21. 10. 1986 (BGBl. 1988 II S. 1087; 1993 II S. 846); Art. 2 I Vertrag zwischen der Bundesrepublik Deutschland und Australien über die Auslieferung vom 14. April 1987 (BGBl. 1990 II S. 111); Art. 2 Abs. 1 Satz 1 Vertrag zwischen der Bundesrepublik Deutschland und der Republik Indien über die Auslieferung v. 27. 6. 2001 (BGBl. 2003 II S. 1635).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

(oder wäre die Tat nach hiesigem Verständnis gerechtfertigt), so ist die Auslieferung unzulässig.[33]

Entsprechen von mehreren prozessualen Taten in einem Auslieferungsersuchen lediglich einige dem Erfordernis der gegenseitigen Strafbarkeit, so wird das Gericht die Auslieferung hinsichtlich der übrigen Taten für unzulässig erklären müssen.[34] An eine solche Teilzulässigkeitserklärung ist der ersuchende Staat gebunden.[35] Er muss seine Strafverfolgung entsprechend beschränken[36] (bzw. bei Auslieferungen zur Strafvollstreckung die zu verbüßende Strafe neu bestimmen[37]).

II. Grundsatz der Spezialität

An dieser Stelle wird der inhaltliche Nexus zwischen dem Grundsatz der beiderseitigen Strafbarkeit und dem Spezialitätsgrundsatz deutlich. Denn es ist gerade letzterer, der – zumindest mittelbar – den Zielstaat zur Beachtung der Beschränkungen verpflichtet, die sich aus der Zulässigkeits- und Bewilligungserklärung des Auslieferungsstaates ergeben.[38]

Allgemein gesprochen soll der Grundsatz der Spezialität gewährleisten, dass der Verfolgte nach seiner Auslieferung nur wegen derjenigen Sachverhalte verfolgt, bestraft oder in seiner Freiheit beschränkt wird, derentwegen er ausgeliefert wurde.[39] Er zählt zu den allgemeinen Regeln des Völkerrechts i. S. v. Art. 25 GG.[40] Das IRG konkretisiert ihn in § 11 und – für den Bereich des Europäischen Haftbefehls – in § 83h.[41] Zudem ist er in

33 *Böhm* in Ahlbrecht/Böhm/Esser/Eckelmanns, Internationales Strafrecht, 2. Aufl. (2018), Rn. 790.

34 Zu beachten ist in diesem Zusammenhang jedoch § 4 IRG.

35 *Kubiciel* in Ambos/Rackow/König, § 3 IRG Rn. 30.

36 *Vogel/Burchard* in Grützner/Pötz, § 3 IRG Rn. 31.

37 BGHSt 27, 168 (173).

38 Genauer hierzu *Kubiciel* in Ambos/Rackow/König, § 11 IRG Rn. 127 f.; *Schierholt* in Schomburg/Lagodny, Internationale Rechtshilfe in Strafsachen, 6. Aufl. (2020), § 11 Rn. 5 ff.

39 Erfolgt die Auslieferung aufgrund eines Europäischen Haftbefehls, soll der Spezialitätsgrundsatz – anders als bei Auslieferungen auf völkervertraglicher Grundlage – kein Verfahrenshindernis, sondern nur noch ein Vollstreckungshindernis begründen (EuGH Slg. 2008, I-8993 – Leymann und Pustovarov).

40 BT-Drs. 9/1338, 46; BVerfG StV 2001, 42 (43); NJOZ 2016, 1064 (1068).

41 § 83h gilt unmittelbar nur für die Einlieferung aufgrund eines Europäischen Haftbefehls. Über Art. 27 und 28 RB-EUHb gelten identische Bedingungen aber für die Auslieferung.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

diversen bi- und multilateralen Auslieferungsabkommen verankert, etwa aus Art. 14 EuAlÜbk bzw. Art. 22 AuslV D-USA.

Die Spezialitätsbindung, die sich direkt aus diesen Auslieferungsverträgen oder durch eine individuelle zwischenstaatliche Vereinbarung aufgrund von § 11 IRG ergibt,[42] soll verhindern, dass sich der ersuchende Staat über Bedingungen hinwegsetzt, die der ersuchte Staat zulässigerweise an die Auslieferung geknüpft hat.[43] Dies ist natürlich besonders dort relevant, wo diese Bedingungen – wie beim Grundsatz der beiderseitigen Strafbarkeit – aus Besonderheiten des nationalen Rechts folgen.[44] Der Grundsatz schützt indes nach vorherrschender Anschauung nicht allein die Souveränität des ersuchten Staates, sondern zugleich auch die Rechte des Verfolgten.[45] Aufgrund seines Stellenwerts sind die nationalen Gerichte von Amts wegen zur Prüfung verpflichtet, ob seine Anerkennung im ersuchenden Staat gewährleistet ist – und zwar nicht nur auf dem Papier, sondern auch in der gelebten Praxis.[46]

D. Auslieferungsrechtliche Konsequenzen

Wie wirken sich diese Prinzipien nun in Fluchtfällen angesichts deren unterschiedlicher Behandlung durch die nationalen Strafrechtsordnungen aus? Dies soll im Folgenden entlang verschiedener Fallgruppen erörtert werden.

I. Drohende Strafverfolgung der Flucht im ersuchenden Staat

Hat der ersuchende Staat die Flucht des Beschuldigten unter Strafe gestellt und begehrt er die Auslieferung explizit (auch) zur Verfolgung dieses Vorwurfs, muss das zur Entscheidung berufene deutsche Oberlandesgericht

42 Näher *Schierholt* in Schomburg/Lagodny, § 11 Rn. 5 ff.

43 BVerfG, Beschl. v. 24.03.2016 - 2 BvR 175/16, Rn. 59; *Vogel/Burchard* in: Grützner/Pötz, § 11 IRG, Rn. 3.

44 Vgl. *Vogel/Burchard* in Grützner/Pötz, § 11 Rn. 3.

45 EuGH, Urt. v. 1.12.2008, C-388/08 PPU, EU:C:2008:669, Rn. 43 f. (Artur Leymann und Aleksei Pustovarov); BVerfG, Beschl. v. 15.02.2023 - 2 BvR 2009/22, Rn. 26; *Kubiciel* in Ambos/Rackow/König, § 11 IRG Rn. 126 m.w.N..

46 BVerfG NJW 1995, 1667 (1667).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

die Auslieferung zumindest insoweit für (teil)unzulässig erklären. Denn in Ermangelung eines entsprechenden Tatbestandes im deutschen Recht würde es ansonsten an der Auslieferungsvoraussetzung der beiderseitigen Strafbarkeit fehlen. Die Spezialitätsbindung sorgt dann im Anschluss dafür, dass der Verfolgte nach seiner Überstellung vor einer Bestrafung wegen der Flucht geschützt ist.

Dies hat exemplarisch das OLG Hamm im Fall eines verurteilten Drogendealers beachtet, dessen Auslieferung die spanischen Behörden begehrten, nachdem er sich während eines Hafturlaubs unerlaubt nach Deutschland abgesetzt hatte.[47] Zu Recht vertrat das OLG die Ansicht, dass eine Auslieferung zwar zur weiteren Vollstreckung der wegen BtM-Einfuhr verhängten Strafe in Betracht komme. Soweit Spanien aber auch die Auslieferung zur Strafverfolgung wegen der Flucht begehrte, stehe der Grundsatz der beiderseitigen Strafbarkeit dem entgegen.[48] Denn die eigennützige Vereitelung der Strafvollstreckung (*Quebrantamiento de condena*) sei zwar nach Art. 468 des spanischen Código Penal, nicht aber nach dem deutschen – vom Selbstbegünstigungsprinzip geprägten – Recht strafbar.

Doch auch, wenn sich das Auslieferungsersuchen nicht explizit auf den ausländischen Flucht-Tatbestand erstreckt, wird das Gericht sich mit dessen Existenz vor einer Entscheidung über die Zulässigkeit auseinandersetzen müssen. Dies gilt jedenfalls dann, wenn die reale Gefahr besteht, dass der ersuchende Staat den Spezialitätsgrundsatz missachten und seine Strafverfolgung nachträglich um den Vorwurf der Flucht erweitern werde. Um zu beurteilen, ob ein solches Szenario droht, wird das Gericht die behördliche Praxis im ersuchenden Staat in den Blick nehmen müssen. Nicht selten lässt sich diese leicht aus frei zugänglichen Verlautbarungen der zuständigen Behörden erkennen – z. B. wenn sich Strafverfolgungsbehörden öffentlich damit brüsten, für eine harte Bestrafung ausgelieferter Flüchtiger zu sorgen.[49]

47 Vgl. OLG Hamm, Beschl. v. 08.09.2009 - (2) 4 Ausl. A 176/09.

48 Delikte gegen die Rechtspflege wie z. B. Fluchthandlungen zählen nicht zu den Katalogtaten im Sinne des Art. 2 Abs. 2 des Rahmenbeschlusses 2002/584/JI des Rates vom 13. Juni 2002 über den Europäischen Haftbefehl und die Übergabeverfahren zwischen den Mitgliedstaaten (ABl. L 190 vom 18.7.2002, S. 1), sodass die Prüfung der beiderseitigen Strafbarkeit nicht nach § 81 Nr. 4 IRG unterbleiben kann.

49 Ein schönes Beispiel liefert diese Pressemitteilung des US-amerikanischen Justizministeriums: *Failure to Surrender to Serve Sentence Results in New Federal Conviction*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Besteht Unsicherheit, ob der ersuchende Staat die Spezialitätsbindung respektieren wird, ist damit die Unzulässigkeit der Auslieferung allerdings noch keineswegs besiegelt. Vielmehr wird das deutsche Gericht in diesem Fall zunächst die Abgabe einer völkerrechtlich verbindlichen Zusicherung erbitten, wonach die ausländischen Behörden von der Verfolgung und Sanktionierung des Flucht-Vorwurfs definitiv absehen wollen. Eine solche Zusage soll nach Auffassung der Rechtsprechung in der Regel genügen, um die Besorgnis einer Missachtung des Spezialitätsgrundsatzes mit hinreichender Sicherheit auszuräumen.[50]

Ein Zusatzproblem ergibt sich allerdings, wenn der ersuchende Staat sich zwar völkervertraglich an den Spezialitätsgrundsatz gebunden fühlt, dessen individualschützende Dimension aber leugnet. Dies wird seit einer aufsehenerregenden Entscheidung des United States Court of Appeals for the Second Circuit im Hinblick auf die gesamten Vereinigten Staaten befürchtet.[51] Das New Yorker Bundesgericht vertrat in jenem Fall die Auffassung, dass Individualrechte, die dem Verfolgten nach den seiner Auslieferung zugrundeliegenden Verträgen zustehen, nur von dem ausliefernden Staat abgeleitet werden könnten. Daher könne der Verfolgte eine Verletzung des Spezialitätsgrundsatzes auch nur dann rügen, wenn die Regierung des ersuchten Staates zuvor offiziell bei der US-Regierung interveniert habe.

Zu Recht hat das BVerfG daraus geschlossen, dass einem an die USA auszuliefernden Verfolgten damit „die Geltendmachung einer Verletzung des Grundsatzes der Spezialität nur noch unter erschwerten, verfassungsrechtlich unzulässigen Bedingungen möglich ist“ und dass „diesem Umstand bereits im Rahmen der Zulassungsentscheidung Rechnung getragen werden“ müsse.[52] Konkret bedeutet dies, dass das zuständige Oberlandesgericht nicht nur auf die verbindliche Zusicherung hinwirken muss, dass der Verfolgte nur wegen des Sachverhalts verfolgt und bestraft werden wird, aufgrund dessen seine Auslieferung für zulässig erklärt und bewilligt wurde. Es muss ergänzend eine Klarstellung der US-Behörden erbitten, ob

---

and Sentence, justice.gov, 11.12. 2020, https://www.justice.gov/usao-ndia/pr/failure-surrender-serve-sentence-results-new-federalconviction-and-sentence (zuletzt aufgerufen: 10.09.2025).

50 Zweifelhaft kann allerdings bisweilen die Belastbarkeit einer solchen Zusicherung sein; vgl. dazu grundlegend EGMR, Urt. v. 17.01.2012, Nr. 8139/09, Rn. 188ff („Othman“).

51 *United States v. Suarez, No. 14–2378 (2d Cir. 2015).*

52 BVerfG, Beschl. v. 24.03.2016 - 2 BvR 175/16, Rn. 48.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

sich der Verfolgte auf den Grundsatz der Spezialität selbst berufen oder in welcher Weise dessen Beachtung anderweitig verfassungskonform sichergestellt werden kann.[53]

II. Drohende Strafschärfung wegen der Flucht

Unstreitig verbietet der Grundsatz der Spezialität die Verfolgung der ausgelieferten Person wegen Taten, die im Rahmen der Zulässigkeitsentscheidung nicht geprüft wurden und entsprechend auch nicht Bestandteil der Bewilligungsentscheidung waren.[54] Nach heute ganz h.M. erschöpft sich sein Gehalt darin aber nicht. Unzulässig soll es vielmehr auch sein, von der Auslieferungsbewilligung umfasste Taten zum Nachteil des Verfolgten bei der Strafzumessung zu berücksichtigen.[55]

Exemplarisch hat das OLG Hamburg dies im Falle eines Verfolgten anerkannt, dessen Auslieferung die Vereinigten Staaten wegen verschiedener, nach dortigem Recht *in toto* strafbarer Embargoverstöße begehrten, welche jedoch – nach deutscher Rechtslage – allenfalls teilweise als strafbares Verhalten zu werten waren:

„Über den unmittelbaren Wortlaut hinaus ist es hiernach auch untersagt, den Ausgelieferten wegen einer anderen Straftat zu bestrafen. Dass nicht nur die Aburteilung, sondern auch eine Bestrafung nicht zugelassener Taten durch den Spezialitätsgrundsatz untersagt ist, versteht sich von selbst und gehört zum anerkannten Kernbestand des Spezialitätsgrundsatzes [...] Zudem ist – entgegen der Ansicht der Vertreter des ersuchenden Staates – davon auszugehen, dass auch die strafschärfende Berücksichtigung nicht zugelassener Taten bei der Aburteilung von Auslieferungstaten einen Verstoß gegen das Bestrafungsverbot und damit gegen den Spezialitätsgrundsatz des Art. 22 des AuslV D-USA darstellt [...] Ausgangspunkt für ein Verbot strafschärfender Berücksichtigung ist dabei die sich aufdrängende Überlegung, dass das eine Tat betreffende Bestrafungsverbot unschwer umgangen werden könnte, wenn die Tat zu einer Straferhöhung für eine andere Tat führen könnte. Eine solche Erhö-

53 Vgl. OLG Karlsruhe, Beschl. v. 12.11.2019 - Ausl 301 AR 82/19 - Rn. 63 f.

54 BVerfG, Beschl. v. 24.03.2016 - 2 BvR 175/16, Rn. 59.

55 BGHSt 22, 318; BGH, Urt. v. 12. 01.2012 - 4 StR 499/11 Rn. 19; BGH, Beschl. v. 28.03.2012 - 5 StR 49/12 Rn. 8; *Hackner* in: Schomburg/Lagodny, Internationale Rechtshilfe in Strafsachen, 6. Aufl. (2020), § 72 IRG Rn. 20.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

hung würde – für alle Parteien des Auslieferungsvertrages ohne weiteres erkennbar – dem Sinngehalt des Spezialitätsgrundsatzes widersprechen […]. Das gilt erst recht, wenn – wie im Falle der USA – erhebliche Strafschärfungen sogar für Taten zu befürchten wären, die nach dem Recht des ersuchten Staates straflos bleiben sollen“.[56]

Diese Erwägungen lassen sich unschwer auf die hier behandelten Fluchtfälle übertragen. Da das gewaltfreie Entweichen aus staatlichem Gewahrsam und erst recht die Flucht eines auf freiem Fuß befindlichen Beschuldigten hierzulande keine Straftatbestände verwirklichen, ist die Auslieferung insgesamt unzulässig, wenn dem Verfolgten im ersuchenden Staat eine strafschärfende Berücksichtigung entsprechender Handlungen droht. Ob dies der Fall ist, hat das zur Entscheidung berufene deutsche Gericht – von Amts wegen – aufzuklären, indem es Erkundigungen zur Rechtslage und -praxis im ersuchenden Staat anstellt.

Im Falle der Vereinigten Staaten muss es dabei fast zwangsläufig auf § 3C1.1 der Federal Sentencing Guidelines stoßen, die eine Strafschärfung für Handlungen vorsehen, die auf eine Behinderung der Strafverfolgung abzielen. Der offizielle Kommentar zu diesen Richtlinien benennt ausdrücklich das Versäumen von Gerichtsterminen und den Versuch, sich dem Verfahren durch Flucht zu entziehen, als Beispiele für solche Verhaltensweisen.[57] Damit berücksichtigt das US-amerikanische Strafrecht Fluchthandlungen, die in Deutschland nicht strafbar und damit auch nicht auslieferungsfähig sind, durch die Hintertür doch zu Lasten des Angeklagten, auch wenn sie nicht als eigenständige Straftaten verfolgt werden – und dies im Regelfall sogar zwingend.

Gerade im Zusammenspiel mit der abgeschwächten Geltung des Spezialitätsgrundsatzes kann sich hieraus ein Auslieferungshindernis ergeben, wie jüngst – und soweit ersichtlich erstmals – das OLG Frankfurt a. M. entschieden hat.[58] Der Verfolgte in jenem Fall war kurz vor Antritt einer Haftstrafe u. a. wegen Betruges aus den USA geflohen und in Deutschland festgenommen worden. Die US-Behörden begehrten seine

56 OLG Hamburg, Beschl. v. 17.08.2023 - Ausl 63/22, Rn. 21 ff.

57 U.S. Sentencing Commission, Federal Sentencing Guidelines Manual § 3C1.1, Commentary 4 (e): “*escaping or attempting to escape from custody before trial or sentencing; or willfully failing to appear, as ordered, for a judicial proceeding*“.

58 OLG Frankfurt, Beschl. v. 19.05.2025 – 1OAusl 112/24 (2 AuslA 184/24); bislang nicht veröffentlicht. Der Verfasser dieses Beitrags war als Beistand des Verfolgten an dem Verfahren beteiligt.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Auslieferung zur Strafvollstreckung sowie zur Verfolgung weiterer Betrugsvorwürfe. Mit Blick auf das Problem der fehlenden beiderseitigen Strafbarkeit erklärte die zuständige Bezirksstaatsanwaltschaft recht bald, auf eine selbständige Ahndung des Fluchtgeschehens verzichten zu wollen.

Der zuständige OLG-Senat gab sich damit aber nicht zufrieden und ersuchte die US-Behörden darüber hinaus um die völkerrechtlich verbindliche Zusicherung, dass die Flucht des Verfolgten ihm im Falle einer erneuten Verurteilung auch im Rahmen der Strafzumessung nicht zum Nachteil gereichen werde. Nachdem die US-Behörden hierauf nur vage reagiert und sinngemäß erklärt hatten, der Flucht werde als Strafzumessungsfaktor jedenfalls keine herausragende Bedeutung zukommen, erklärte das OLG Frankfurt – der oben dargestellten Argumentation folgend – die Auslieferung insgesamt für unzulässig und erlaubte dem Verfolgten die Weiterreise in sein Herkunftsland.

Unklar blieb damit allerdings, wie eine Zusicherung der USA hätte aussehen müssen, um dieses Ergebnis abzuwenden. Denn während sich eine Staatsanwaltschaft verpflichten kann, bestimmte Vorwürfe nicht zu verfolgen oder – im US-Kontext immer wieder relevant – die Todesstrafe nicht zu fordern, ginge es dabei um einen diffizileren Eingriff in die gerichtliche Strafzumessungspraxis und damit letztlich um den Kernbereich richterlicher Unabhängigkeit. Die Vereinigten Staaten hätten möglicherweise zusichern müssen, ihr eigenes Recht missachten zu wollen.

Das Beispiel macht jedenfalls deutlich, dass im Extremfall das Zusammenspiel von auslieferungsrechtlichen Strukturprinzipien und Divergenzen zwischen nationalen Straf- und Strafverfahrensrechten, über die der ersuchte und der ersuchende Staat auch noch ungeschickt kommunizieren, zu dem paradoxen Resultat führen kann, dass ein Straftäter gerade deshalb nicht ausgeliefert werden kann, *weil* er zuvor geflüchtet war.

III. Drohende Rücknahme gewährter Vergünstigungen wegen der Flucht

1. Prozessuale und vollzugsrechtliche Konsequenzen

Dass die missglückte Flucht auch abseits einer strafrechtlichen Ahndung unangenehme Konsequenzen haben kann, steht außer Frage. Das ist auch nach deutschem Recht der Fall: Setzt sich ein Beschuldigter, gegen den bislang keine freiheitsbeschränkenden Maßnahmen vorlagen, ins Ausland

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ab, um fortan für die Ermittlungsbehörden unerreichbar zu sein, wird der Erlass eines Haftbefehls (§ 112 StPO) die naheliegende Reaktion darstellen. Erst recht muss ein Beschuldigter, der nur unter Auflagen aus der Untersuchungshaft entlassen wurde und diese Gelegenheit nun zum Untertauchen missbraucht, nach erneuter Ergreifung mit einer längeren Inhaftierung rechnen, weil weniger einschneidende Maßnahmen (§ 116 StPO) sich als offenkundig untaugliches Mittel zur Einhegung der Fluchtgefahr erwiesen haben. Und schließlich hat, wer sich aus staatlichem Gewahrsam befreit, nach seiner Wiederergreifung strengere Beschränkungen (vgl. § 119 StPO) zu gewärtigen.

Entsprechend steht auch die Aussicht auf derartige strafprozessuale und vollzugsrechtliche Konsequenzen einer Überstellung in den ersuchenden Staat nicht entgegen. Der Grundsatz der beiderseitigen Strafbarkeit ist hier ohnehin meist nicht tangiert, weil es gar nicht um eine Bestrafung der Flucht, sondern lediglich um die Sicherung des Strafverfahrens bzw. -vollzugs im Hinblick auf den ursprünglichen Sachverhalt geht.[59]

Ähnliches dürfte für eine disziplinarische Ahndung der Flucht eines Verurteilten durch die Justizvollzugsanstalt gelten. Selbst diejenigen, die dem Selbstbegünstigungsprinzip im deutschen Strafrecht Verfassungsrang beimessen, halten es – auch mit Blick auf § 14 StVollzG – für zulässig, „dass die Nichtrückkehr vom Freigang und Ausgang sowie aus dem Urlaub disziplinarrechtlich geahndet werden kann, [denn die] Öffnung des Vollzuges bedeutet einen Vertrauensvorschuss und gleichzeitig eine besondere Inpflichtnahme, dieses Vertrauen nicht zu missbrauchen“.[60] Ob dies – gleichsam erst recht – auch für die gewaltlose Selbstbefreiung eines Strafgefangenen gilt, ist umstritten, dürfte aber zu bejahen sein.[61] Auch im

59 Diese Schlussfolgerung ist weniger eindeutig, als es den Anschein haben mag. So legt der EGMR – jedenfalls im Kontext von Art. 7 EMRK – den Begriff der Strafe völlig anders aus, als es die deutsche Rspr. tut; vgl. EGMR, Urt. v. 9. 2. 1995, Series A307-A, Rn. 28 („Welch“). Der Gerichtshof nimmt vor allem die faktischen Wirkungen einer Maßnahme in den Blick, während der Unterscheidung von präventiver und repressiver Zielsetzung keine zentrale Bedeutung zukommt. Entscheidungen im Strafvollzug, wie etwa die nachträgliche Verschärfung des Vollzugs, betreffen aber auch nach dem Verständnis des EGMR typischerweise nicht die „Strafe“, vgl. EGMR, Urt. v. 10. 7. 2003 – 43522/98 Rn. 51 („Grava“). Siehe aber auch EGMR NJOZ 2014, 1587 (1588), wo der Gerichtshof eine Art Gesamtbetrachtung vornimmt.

60 *Ostendorf* NStZ 2007, 313 (314).

61 OLG Hamm StV 1989, 441; a.A. *Ostendorf* NStZ 2007, 313 ff. m.w.N. zu beiden Ansichten.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Auslieferungskontext kann ein Konflikt mit dem Grundsatz der beiderseitigen Strafbarkeit bzw. der ihn absichernden Spezialitätsbindung folglich nicht auftreten, wenn dem Verfolgten nach seiner Auslieferung lediglich disziplinarische Maßnahmen im Vollzug drohen.[62]

2. Aufkündigung verfahrensbeendigender Absprachen

Fraglich ist jedoch, wie mit Konsequenzen umzugehen ist, die über die Rücknahme von Vollzugslockerungen oder eine disziplinarische Ahndung im Vollzug hinausgehen, zugleich aber hinter einer unmittelbaren Sanktionierung der Flucht zurückbleiben. Zu denken ist dabei vor allem an die Aufkündigung verfahrensbeendender Absprachen. In den Vereinigten Staaten etwa endet die überwältigende Mehrheit aller Strafverfahren mit einem sog. „Plea Agreement".

Als Gegenleistung für ein Schuldeingeständnis des Angeklagten lässt sich die Staatsanwaltschaft auf eine für den Angeklagten mildere Bewertung bestimmter Taten ein oder verzichtet sogar ganz auf die Verfolgung bestimmter Sachverhalte.[63] Flieht der Angeklagte jedoch im Anschluss an seine Verurteilung, anstatt die „vereinbarte" Strafe anzutreten, kann dies dazu führen, dass auch die Staatsanwaltschaft sich an ihre Zusagen nicht mehr gebunden fühlt und eine Nachtragsanklage („superseding indictment") erhebt, womit der Verfolgte im Falle seiner Auslieferung plötzlich eine deutlich schwerere Strafe zu erwarten hat – und dies als direkte Folge seines Fluchtversuchs. Auch hieraus dürfte nach hier vertretener Ansicht jedoch meist kein Auslieferungshindernis erwachsen.

Dass die Justiz des ersuchenden Staates die dem Ersuchen zugrundeliegende prozessuale Tat nach erfolgter Auslieferung *rechtlich* neu bewertet (z. B. als Mord statt als Totschlag), ist ohnehin unproblematisch – und zwar sowohl vor dem Hintergrund der beiderseitigen Strafbarkeit als auch der Spezialität, weil weiterhin derselbe Sachverhalt betroffen ist, der

62 Dies gilt jedenfalls solange, wie die Maßnahmen keinen strafähnlichen Charakter haben. Je nach Art der Disziplinierung kann der Auslieferung aber der Ordre-Public-Vorbehalt des § 73 S. 1 IRG entgegenstehen.

63 American Bar Association „2023 Plea Bargain Task Force Report urges fairer, more transparent justice system", 22.02.2023, abrufbar unter >thickhttps://www.americanbar.org/news/abanews/aba-news-archives/2023/02/plea-bargain-task-force/?utm_source=chatgpt.com< (zuletzt abgerufen: 10.09.2025). Danach beruhen mittlerweile 98 % aller Schuldsprüche auf eine solchen Absprache.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Gegenstand der deutschen Zulässigkeits- und Bewilligungsentscheidung war.[64]

Aber auch die Wiedereinbeziehung von *Sachverhalten*, die im Rahmen des *Plea Agreements* von der Verfolgung ausgenommen worden waren, ist grundsätzlich zulässig. Auch ein solches Vorgehen ist nicht vergleichbar mit dem oben diskutieren Fall einer unzulässigen strafschärfenden Berücksichtigung der Flucht im Rahmen der Strafzumessung. Mit der Aufkündigung der Plea Agreements zugesagten Beschränkung der Strafverfolgung wird nicht die Flucht sanktioniert. Vielmehr wird ein Zugeständnis, welches der Staat dem Verfolgten im Gegenzug für dessen Kooperation gemacht hatte, wieder kassiert, weil der Verfolgte sich seinerseits absprachewidrig verhalten hat.[65]

Ein Problem dürfte sich lediglich in drei Sonderfällen ergeben: (1) Zum einen, wenn sich das Auslieferungsersuchen nicht von vornherein (auch) auf die Überstellung zum Zwecke der Verfolgung der wieder einbezogenen Taten richtete und entsprechend die Auslieferung auch nur im Hinblick auf die Altsachverhalte bewilligt wurde. Eine nachträgliche Erweiterung ohne deutsche Zustimmung wäre dann mit dem Spezialitätsgrundsatz nicht vereinbar. (2) Zum anderen müssen mit Blick auf das Erfordernis der beiderseitigen Strafbarkeit natürlich auch die wieder einbezogenen Sachverhalte nach deutschem Recht tatbestandsmäßig und rechtswidrig sein. Daran kann es im Einzelfall fehlen. (3) Und schließlich wird man in extrem gelagerten Fällen – etwa, wenn sich infolge der Aufkündigung des Plea Agreements die Straferwartung dramatisch erhöht und weit über das hinausschießt, was in vergleichbaren Fällen sonst üblich wäre – darüber nachdenken müssen, ob die Behörden des ersuchenden Staates nicht doch eine klandestine Sanktionierung der Flucht anstreben. All diese Szenarien dürften aber eher selten vorkommen.

64 *Kubiciel* in Ambos/Rackow/König, § 11 Rn. 130. Dies gilt unstreitig jedenfalls dann, wenn es sich auch nach der Neubewertung weiterhin um eine auslieferungsfähige Tat handelt.

65 Das deutsche Strafverfahrensrecht kennt durchaus ähnliche Situationen. Endet eine Hauptverhandlung etwa nach Zustimmung aller Beteiligten mit einer Verfahrenseinstellung gegen Auflagen (§ 153a StPO) und erfüllt der Angeklagte diese Auflagen nicht innerhalb der gesetzten Frist – z. B. weil er sich inzwischen ins Ausland abgesetzt hat – ist das Gericht an den ursprünglichen „Deal“ nicht mehr gebunden. Es kann den Angeklagten nach erneuter Hauptverhandlung zu einer „richtigen“ Strafe verurteilen, die regelmäßig härter ausfallen wird, als es die nichterfüllte Auflage gewesen wäre. Ein ähnlich konditionaler Mechanismus findet sich auch in § 257c Abs. 4 S. 2 StPO.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

E. Schlussbemerkungen

Das in diesem Beitrag behandelte komplexe Zusammenspiel von Spezialitätsgrundsatz, Erfordernis der beiderseitigen Strafbarkeit und vielschichtigen Unterschieden in der Kriminalisierung von Fluchthandlungen in den nationalen Rechtsordnungen spielt oft gerade in solchen Auslieferungsfällen eine bedeutsame Rolle, die auf den ersten Blick völlig unproblematisch erscheinen. Dies mag ein Grund dafür sein, dass es häufig übersehen wird – und zwar sowohl von den Beiständen des Verfolgten als auch von den Gerichten, die sich eigentlich von Amts wegen damit auseinandersetzen müssten. Zu oft unterbleibt schon eine sorgfältige Analyse der Strafbarkeitsrisiken, welche dem Verfolgten gerade infolge seiner Flucht bei einer Überstellung im ersuchenden Staat drohen.

Dabei leuchtet der vorliegende Beitrag nur einen Ausschnitt eines deutlich größeren rechtlichen Mosaiks aus. Denn was hier am Beispiel der Flucht ausgeführt wird, gilt letztendlich in ähnlicher Weise für sämtliche Formen selbstbegünstigenden Handelns. Mit anderen Worten: Nicht nur bezüglich der Kriminalisierung von Fluchthandlungen bestehen massive nationale Unterschiede. Auch im Hinblick auf die Manipulation von Beweismitteln, die Irreführung von Ermittlungspersonen oder das Verleiten von Zeugen zur Lüge divergieren die Rechtsordnungen weltweit ganz erheblich. Eine ergänzende Einbeziehung dieses Phänomenbereichs hätte freilich den Rahmen dieses Beitrags gesprengt. Dasselbe gilt für eine rechtspolitische Diskussion, ob sich der Grundsatz der beiderseitigen Strafbarkeit nicht in einigen Bereichen überlebt hat und daher (weiter) eingeschränkt werden sollte.[66] Für die Zukunft erscheint jedoch beides als durchaus lohnendes Unterfangen.

66 Vgl. dazu ausführlich *Vogel/Burchard* in: Grützner/Pötz, § 3 IRG Rn. 7 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Are Trump Judges Different? Evidence from Immigration Cases

Daniel M. Klerman[1]

Judges appointed by President Trump are more likely to vote in favor of the government in cases challenging the second Trump administration's immigration policies. While Trump's Supreme Court nominees behave like other Republican nominees on the Court, Trump's lower court nominees are twice as likely to vote in favor of the government as nominees of other Republican presidents; in contrast, other Republican nominees to the lower courts are statistically indistinguishable from Democratic nominees. The difference between Trump nominees and other judges is driven almost entirely by judges 55 years old or younger, who may be influenced by the prospect of promotion.

A. Introduction

Many Presidents have been disappointed by their judicial appointments. For example, Eisenhower "frequently remarked that his biggest mistake was the appointment of 'that dumb son of a bitch Earl Warren'" (Ambrose 1984, 2:190; Cray 1997, 10). Does President Trump have reason to be similarly upset about his judicial nominees?

This chapter investigates that question by examining the decisions of Trump's judicial nominees in cases challenging the second Trump administration's immigration policies. It finds that, overall, Trump nominees are little different from other judges appointed by Republican presidents. On the other hand, when one focuses on lower court judges, and especially on judges under the age of 55, Trump appointees are startlingly

1 The author thanks Julia Sills for excellent research assistance. The author also thanks Scott Altman, Jordan Barry, Christina Boyd, Rebecca Brown, Ryan Bubb, Lee Epstein, William Eskridge, Tamar Kricheli-Katz, Eric Posner, Mark Ramseyer, Cass Sunstein, and Angela Zhang for helpful comments and suggestions.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

different.[2] District and appellate court judges appointed by Trump are twice as likely to uphold Trump administration immigration policies as judges appointed by other Republican presidents. Moreover, lower court judges appointed by Trump who are 55 years old or younger are more than three times as likely to vote to uphold Trump policies, while those over 55 vote in ways indistinguishable from other judges appointed by Republican presidents; in fact, Trump-appointed judges over 55 are statistically indistinguishable from judges appointed by Democratic presidents in immigration cases. One reason judges 55 or under may be so different is that they may be "auditioning" for the Court of Appeals or Supreme Court. That is, they may be hoping that President Trump will promote them and so they may be tailoring their decisions to meet his preferences (Brown et al. 2025).

Section B of this chapter reviews the literature. Section C describes the data, and Section D analyzes the data.

B. Literature

For several decades, if not longer, it has been well understood that judicial decisions reflect the ideology and policy preferences of the deciding judge, and that a good proxy for judges' ideology or preferences is the political party of the president who appointed them (Segal & Spaeth 2002; Sunstein et al. 2004). The political party of the appointing president (Democratic or Republican) rather than the identity of the president (e.g. Obama or G.W. Bush) is a good predictor because presidents of the same party are relatively similar, especially with respect to judicial appointments. Is Trump different? He has embraced policies – such as tariffs and a conciliatory policy toward Russia – that deviate from more than half a century of Republican orthodoxy. Perhaps his judicial nominees also reflect a significant deviation from nominees by Presidents George Bush, H.W. Bush,

2 One might have thought that the overall figures would be similar to the figures for district and appellate court judges because there are so many more lower court judges than Supreme Court justices. On the other hand, because, as discussed below, the unit of analysis in this article is the judicial vote not the case, each Supreme Court case results in nine observations (nine votes), while each district court decision results in just one observation and each court of appeals decision results in only three observations. As a result, inclusion or exclusion of the Supreme Court can have a large effect, as discussed in more detail in section 3.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

and Ronald Reagan. On the other hand, given that during his first term, Trump reportedly outsourced judicial appointments to the Federalist Society (Toobin 2017) – a group whose ideology reflects decades of mainstream conservative ideology (e.g. originalism and opposition to *Roe* v. *Wade*) – one might have expected that Trump's judicial nominees would be like the nominees of other recent Republican presidents.

A number of studies have analyzed the behavior of judges appointed by Trump. Most have found that they are significantly different from those appointed by other Republican presidents. Rebecca Brown, Lee Epstein, and Mitu Gulati (2025) studied federal district court and circuit court judges' votes in cases relating to gun rights. They found that Trump appointees rendered decisions favoring gun rights about 40 % of the time, while judges appointed by other Republican presidents supported gun rights 28 % of the time. Republican appointees who were 55 or under, whom the authors called "auditioners" (potential appointees to the Supreme Court), were more than twice as likely to vote in favor of gun rights as older judges (60 % v. less than 30 %) in cases after the Supreme Court's landmark *Bruen* decision, while Democratic appointees' behavior did not differ by age.

Stephen Choi, Mitu Gulati, and Eric Posner (2025) examined judicial voting patterns in court of appeals cases involving the free exercise of religion. They found that Trump-appointed judges were more likely to vote for the plaintiff in cases brought by Christians or Jews than judges appointed by Democratic presidents (54 % v 31 %), but less likely to vote for the plaintiff in brought by Muslims (17 % v 44 %). Judges appointed by other Republican presidents had voting records in between those of Trump appointees and Democratic appointees. The difference between Trump-appointed judges and Democratic-appointed judges is statistically significant for both types of cases, but the difference between Trump appointees and other Republican appointees is only significant for cases brought by Christians and Jews. The authors did not find any significant difference between Trump appointees who were 55 or younger and those who were older.

Zalman Rothschild (2022) examined judicial voting patterns in free exercise cases. Unlike Choi et al. (2025), he looked at district court as well as appellate court decisions. His results are even more stark. Trump appointed judges upheld free exercise claims 77 % of the time, judges appointed by other Republican Presidents upheld free exercise claims 44 % of the time, and Democratic appointees upheld such claims only 7 % of

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the time. Rothschild did not check for differences between judges by age.

C. Data

The data for this study are all federal cases challenging the legality of Trump administration actions relating to immigration and citizenship during his second term in office. Cases analyzed include those challenging the use of the Alien Enemies Act, those challenging Trump's Executive Order on Birthright Citizenship, and those challenging DHS's revocation of Temporary Protective Status (TPS). Cases decided by immigration judges and immigration cases that disputed the application of settled law – e.g. whether an individual is subject to removal or deportation under provisions of the Immigration and Naturalization Act that were not affected by new orders or policies of the second Trump administration – were not included. The cases challenging Trump administration policies were assembled by Just Security as part of its Litigation Tracker. According to its website, "Just Security is an editorially independent, non-partisan, daily digital law and policy journal … based at the Reiss Center on Law and Security at New York University School of Law." The database was last checked on August 24 so the cases span roughly seven months, from Trump's inauguration (January 20, 2025) through August 24, 2025.

For most district court cases, the judge makes dozens of decisions, most of a procedural nature. A district court decision was only analyzed in this study if it granted or denied a temporary restraining order, a preliminary injunction, or a permanent injunction. If such a decision was followed by a decision clarifying or amending the prior decision, only the later decision was analyzed, unless the later decision followed an appeal. A decision was coded as favoring the individual (and therefore not in favor of the government) if it granted a temporary restraining order, preliminary injunction, or permanent injunction against some aspect of Trump's immigration policy, even if it did not grant all the relief that the plaintiff(s) requested.

Appellate court decisions – whether by circuit courts or the Supreme Court – were analyzed if they involved the appeal of a coded district court decision. The unit of analysis was the judge's vote (not the case), so a circuit court opinion would generally involve three observations – the votes of the three participating judges. If the appellate decision was unanimous,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the votes were coded as favoring the government if the decision granted a stay or vacated a temporary restraining order, preliminary injunction, or permanent injunction, but not if the appellate court narrowed an order or injunction issued by the district court but left part of it in force. If an appellate decision was not unanimous, the votes of the judges were coded based on their relative position. That is, if some judges would uphold a preliminary injunction in its entirety, but others on the same panel would narrow it, the latter would be coded as pro-government, whereas the former would not.

Decisions were coded by a second-year law school research assistant under the supervision of the author.

Information about which president appointed each judge and the judge's birth year are from the Federal Judicial Center Biographical Directory of Article III judges, while a judge's gender, when not clear from his or her name, is from other sources, such as Wikipedia. Judges were coded as over 55 if they were born in 1969 or earlier.

D. Analysis

Figure 1 below analyzes all judicial votes[3] on cases challenging President Trump's immigration policies during the first months of the Second Trump administration. It includes votes by judges of all levels of court – district court, court of appeals, and Supreme Court – and judges of all ages.

Republican-appointed judges, whether appointed by Trump or not, are more than twice as likely to vote for the government's position on immigration matters than judges appointed by Democrats (51 % or 62 % v. 20 %). Although Trump-appointed judges are slightly more likely to vote pro-government than judges appointed by other Republican presidents (62 % v 51 %), the difference is not statistically significant.

Figure 2 separates judges by whether they are on the Supreme Court or whether they are on a lower court (court of appeals or district court).

3 District court judges do not technically vote because they ordinarily make decisions alone. Nevertheless, for simplicity, district court decisions are classified as "votes" for the purpose of this analysis.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

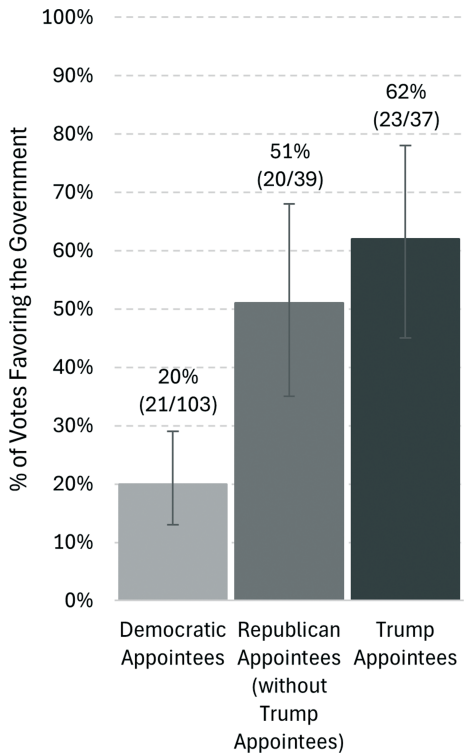


Figure 1. Voting on Immigration Matters, All Judges (2025)
Note. Error bars indicate the exact Clopper-Pearson binomial 95 % confidence intervals.

Figure 2 shows that, while Trump nominees on the Supreme Court are similar to other Republican nominees, Trump nominees to the lower courts are very different.

Republican nominees to the Supreme Court (whether appointed by Trump or not) are about four times as likely to uphold Trump immigration policies as Democratic appointees (78 % or 67 % v 17 %). The voting of Trump appointees on the Court is actually very similar to that of other Republican nominees. Although the data suggest that Trump appointees are slightly *less* likely to uphold Trump immigration policies (67 % v 78 %), the difference is not statistically significant. The main reason Trump's Supreme Court nominees look similar to other Republican appointees is

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

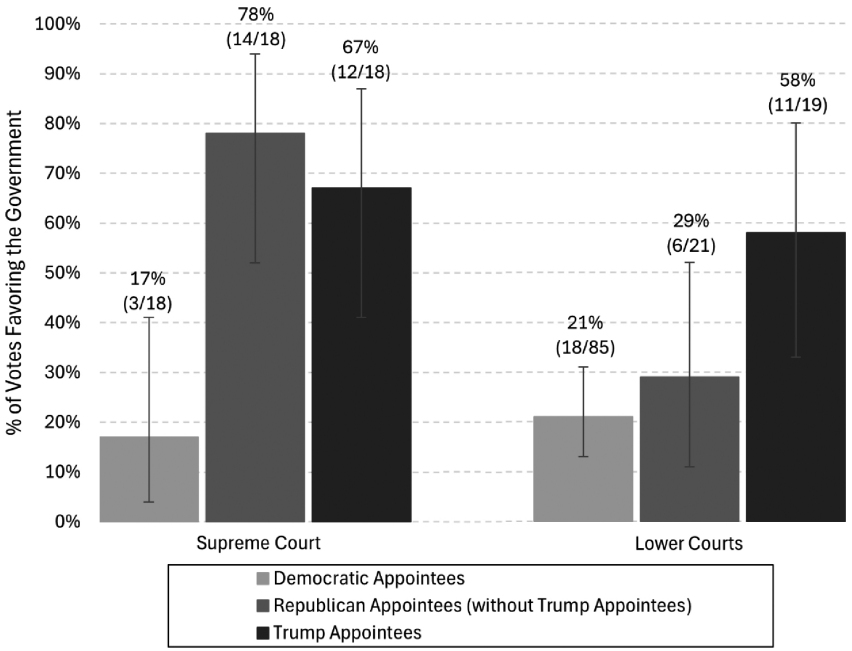


Figure 2. Voting on Immigration Matters by Court Level (2025)
See Note to Figure 1 for information on the error bars.

that the two most conservative Justices – Alito and Thomas – were appointed by Presidents George and H.W. Bush. In comparison to those two justices, Trump's nominees – Barrett, Kavanaugh, and even Gorsuch – although still conservative, are much more moderate.[4]

When one looks at the lower courts (the right half of Figure 2), a very different picture emerges. Trump-appointed judges are twice as likely to uphold government immigration policy as other Republic appointees (58 % v. 29 %). The difference is statistically significant at the p=0.06 level, very close to the 5 % cutoff often used. The absence of more clearly robust statistical significance reflects the small number of votes available for analysis – 19 votes by Trump appointees and 21 votes by other Republican

4 According to Epstein et al. 2025, p. 10, Alito and Thomas receive Martin-Quinn scores of 2.5 and 3.1, whereas Barrett, Kavanaugh, and Gorsuch receive scores of 0.4, 0.6, and 1.1 respectively.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

appointees. Given that the votes analyzed come from only a seven-month period, the number of votes is remarkably large, but it does not provide the power necessary to yield results that are statistically significant at the 5% level even for the large differences found here.

In contrast, lower court judges appointed by Republican presidents other than Trump vote in ways similar to Democratic appointees, upholding government actions about one quarter of the time (29% for other Republican appointees, 21% for Democratic appointees). The difference between Democratic appointees and non-Trump Republican appointees is not statistically significant by any standard (p=0.47), whereas the difference between Trump appointees and Democratic appointees is clearly significant (p=0.001).

This analysis suggests that, while Trump's appointments to the Supreme Court have garnered the most attention and, in fact, have moved the Court significantly to the right, it is his lower court nominees who are a more significant departure from historical norms.

It is also notable that even Trump-appointed lower court judges vote to invalidate Trump administration policies 42% of the time. One could interpret that as a testament either to their independence or to the radicalism of the immigration policies that the Trump administration has implemented.

The difference between Trump nominees and other judges in immigration cases is larger than what Brown et al. (2025) found with respect to gun rights. They found a 17-percentage-point difference between Trump lower court nominees and other Republican nominees (45% v 28% post *Bruen*), whereas the analysis here suggests a 29-percentage-point difference (58% v 29%). The large differences found here with respect to immigration are more similar in magnitude to the differences relating to religion found by Choi et al. (2025) and Rothschild (2022).

Figure 3 further breaks down the voting of lower court judges by age: whether the judge is 55 or under, or older than 55. The break point of 55 was chosen because it corresponds to the age at which many scholars think judges become too old to be promoted, because Presidents prefer to maximize their influence by nominating judges who will serve for several decades (Brown et al. 2025). The 55 cutoff also conveniently divides the data on Trump-nominee lower courts almost perfectly in half (10 votes by judges 55 or younger, 9 votes by judges over 55). Of course, by slicing the data this finely, the number of votes analyzed becomes relatively small, so results must be viewed with caution.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

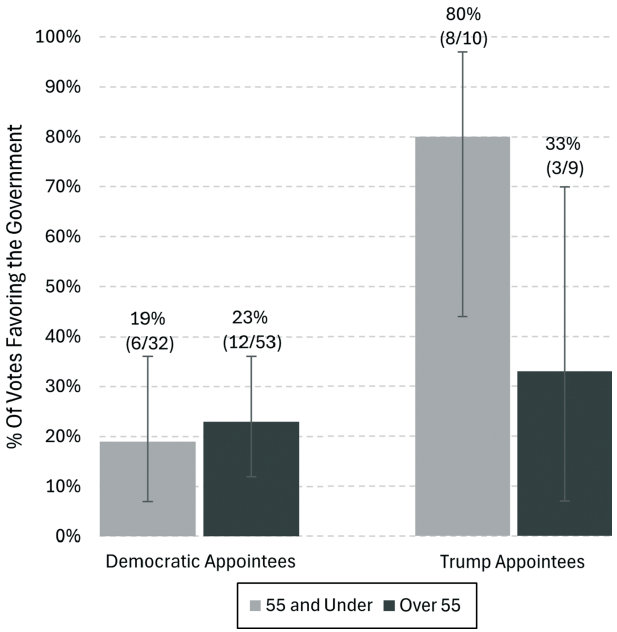


Figure 3. Voting on Immigration Matters by Age (Lower Courts, 2025)
See Note to Figure 1 for information on the error bars.

The difference between younger Trump nominees and those over 55 is stark. Younger Trump nominees vote to uphold government immigration policies 80 % of the time, whereas those under over 55 do so only 33 % of the time. Even with the small number of observations, this difference is statistically significant (p=0.04). In fact, Trump judges over 55 are barely different from other Republican appointees, and the difference between them and Democratic appointees is not statistically significant.[5] Although it is possible that older judges just have different judicial ideologies or preferences, the most plausible interpretation of the difference between younger and older Trump appointees is that the prospect of promotion induces younger Trump appointees to vote more in line with Trump administration preferences.

5 As can be seen in Figure 2, other Republican nominees vote for the government 29 % of the time, and Democratic nominees vote for the government 21 % of the time. The differences between Trump nominees under 55 and Democratic and other Republican nominees is not statistically significant.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In contrast to Republican nominees, there is almost no difference between Democratic appointees by age. As can be seen in the left side of Figure 3, only four percentage points separate Democratic nominees over 55 and those who are younger. Democratic nominees have no chance of being promoted by Trump, so they get no benefit from tilting their votes toward the administration. While one could think that Democratic appointees might hope to be promoted by a future president, that does not seem to induce them to be more (or less) pro-immigrant, perhaps because it is hard to predict whether a future Democratic president will be moderate or more pro-immigrant on immigration issues. Also, by the time a Democratic president is in power, many if not most of the Democratic appointees now 55 or under are likely to be over the age at which promotion is plausible.

It is not possible to analyze the voting of judges nominated by other Republican presidents by age because all of them involved in disputes in the dataset are over 55, so none of them are plausible "auditioners" for promotion.

It is notable that the gap between auditioners and non-auditioners is larger in immigration cases than in other subject matters. For example, in gun cases, Brown et al. (2025) found a 31-percentage-point gap (60 % v 29 %), whereas the analysis here finds a 47-percentage-point gap (80 % v 33 %). Choi et al. (2025) did not find any difference in religion case between those 55 or under and those who were older. Again, the large difference in immigration likely reflects the importance of immigration to Trump and his supporters and the attention immigration cases receive in the media.

E. Conclusion

Trump's nominees have provided strong support for his immigration policies. His nominees to the Supreme Court, like other Republican nominees, vote to uphold his policies about seventy percent of the time, which is more than four times as often as Democratic nominees. In the lower courts, Trump nominees are more pro-government on immigration matters than both Democratic nominees and judges appointed by other Republican presidents. The magnitude of the differences is large, with Trump appointees voting to uphold Trump immigration policies almost sixty percent of the time, while other judges do so less than thirty percent

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of the time. The difference between Trump nominees and other judges is particularly stark for nominees 55 or younger, who may be influenced by the possibility that Trump (or a future Republican president) might promote them.

While these results could be interpreted to cast doubt on judicial independence and impartiality, there are three ways in which one might interpret the results in a more positive way:

1) Even Trump nominees ruled against the president in a substantial percentage of immigration cases. Trump nominees to the Supreme Court voted against him one third of the time, and nominees to the lower courts voted against him forty-two percent of the time.
2) While Trump nominees to the lower courts seem ideological in their voting, the fact that there is hardly any difference between Democratic nominees and other Republican nominees in immigration matters might suggest that most judges are relatively constrained by the law.
3) While many of Trump's appointees are now 55 or under, most will relatively soon be beyond the age at which promotion is plausible. The theory that younger judges are influenced by the prospect of promotion would suggest that Trump nominees are likely to become more moderate as they get older. Similarly, if a Democrat or more moderate Republican is elected in 2028, even Trump appointees under 55 will lose the promotion incentive to be pro-government on immigration, which may moderate their decision-making.

References

Ambrose, S.E. (1984). *Eisenhower: The President*. Simon & Schuster.

Brown, R.L., Epstein, L., & Gulati, M. (2025). Guns, Judges, and Trump. *Duke Law Journal Online*, *74*(Spring 2025), 81–109.

Choi, S.J., Gulati, M., & Posner, E.A. (2025). Trump's Law Court Judges and Religion: An Initial Appraisal. *Journal of Legal Studies, 54*(1), 1–41.

Cray, E. (1997). *Chief Justice: A Biography of Earl Warren*. Simon & Schuster.

Epstein, L., Martin, A.D., & Nelson, M.J. (2025). *Provisional Data Report on the 2024 Term*. Unpublished manuscript. Retrieved from https://static1.squarespace.com/static/60188505fb790b33c3d33a61/t/685f50748be9977089829a96/1751076981684/2024termdatareport.pdf

Federal Judicial Center. (2025, August). Biographical Database of Article III Federal Judges, 1789-present. https://www.fjc.gov/history/judges

Just Security. (2025, August). About Us. https://www.justsecurity.org/about-us/

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Just Security. (2025, August). *Litigation Tracker: Legal Challenges to Trump Administration Actions*. https://www.justsecurity.org/107087/tracker-litigation-legal-challenges-trump-administration/

Rothschild, Z. (2022). Free Exercise Partisanship. *Cornell Law Review, 107*(4), 1067–1135.

Segal, J.A., & Spaeth, H.J. (2002). *The Supreme Court and the Attitudinal Model Revisited*. Cambridge University Press.

Sunstein, C.R., Schkade, D., & Ellman, L.M. (2004). Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation. *Virgina Law Review, 90*(1). 301–354.

Toobin, J. (2017, April 10). A Conservative Pipeline to the Supreme Court. *The New Yorker*. https://perma.cc/5QKE-BUKX

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Judging Theories
A Survey in Honor of Christoph Engel

Keren Weinshall

A. A Personal Introduction and Dedication

I first met Christoph Engel in 2014 at a workshop in Jerusalem, where we engaged with each other's papers. I must have done something right – shortly afterward, he invited me to one of the vibrant Journal of Institutional and Theoretical Economics (JITE) symposia. There, he became particularly intrigued by a dataset I had constructed for the symposium paper. While I initially conceived of it as a resource for studying class actions, Christoph – true to his creative and original thinking – reimagined it entirely. In his hands, it became a dataset for exploring how legal innovations spread, both theoretically and empirically. If you are curious to see how this transformation unfolded, I invite you to read the two articles we co-authored on the subject (Engel et al. 2018, Engel and Weinshall 2022).

We enjoyed the collaboration so much that we went on to write another paper together – this time on a completely different topic: the effects of judicial workload, using a natural experiment approach.

From these experiences, I've drawn two key insights about working with Christoph. First, you always know where a project begins, but never where it will lead. The breadth of Christoph's interests, his intellectual curiosity, and his deep knowledge make the possibilities virtually limitless. Second, the journey itself is always a joy. Working with Christoph is a deeply rewarding intellectual and personal experience.

It is with these sentiments in mind that I dedicate this chapter on judicial decision-making to Christoph, in deep appreciation of his scholarship, mentorship, and – above all – his friendship.

The broader literature on judicial behavior seeks to illuminate why and how judges make the choices they do, and what consequences those choices have for society (Epstein et al. 2023). Within this field, studies of judicial decision-making focus on the former question: what motivates judges? What and how do they maximize? In this chapter, I review the current state of the literature and highlight Christoph Engel's contributions to it.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The chapter is structured as follows. The next section introduces the scholarly field of judicial decision-making and maps seven prominent theoretical approaches to its study. The following section elaborates on each approach – its origins and development – demonstrating how Christoph Engel's intellectual versatility has enabled him to make meaningful contributions across all seven. The final section concludes with reflections on the significance of Engel's scholarship for the comparative study of judicial behavior and the inspiration it offers for future research.

B. Mapping the Field: Seven Approaches to Judicial Decision-Making

Why do judges decide cases the way they do? What motivates their decisions, and what considerations guide them when resolving legal disputes? For nearly a century, scholars from law, political science, economics, sociology, and psychology have advanced distinct theoretical paradigms and approaches to answer this question. These approaches reflect different assumptions about what the law is, institutional constraints, and methodological priorities. Some approaches view judges as neutral appliers of legal rules, others see them as political actors, strategic players, or socially embedded individuals.

This survey maps seven prominent paradigms in the study of judicial decision-making, as summarized in Table 1: the Legal Model (often referred to as "old legalism"); the Attitudinal Model; Identity-Based Approaches; Strategic Judging; the Labor Market Model; the Neo-Institutional Approach (also known as "new legalism"); and Cognitive Theories (sometimes labeled "thinking fast judging"). While originally developed in distinct disciplinary contexts, these approaches are increasingly applied across legal systems, court types, and national contexts.

The mapping offered here draws on and extends earlier efforts to classify the field (Epstein, Šadl & Weinshall 2021, 2023; Epstein et al. 2024). My aim is to present the core assumptions, disciplinary foundations, and distinctive contributions of each paradigm to our understanding of judicial decision-making – and, in doing so, to provide a framework for situating Christoph Engel's rich and wide-ranging contributions across these theoretical perspectives.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 1. Approaches to Judicial Decision-Making

Approach	Legal (old)	Attitudinal	Identity	Strategic	Labor Market	New- institutional	Cognitive
Discipline (main)	Law	Political science, psychology	Sociology, psychology	Economics, political science	Economics	Law & political science	Psychology, behavioral economics
Foundational Years	–	1960s	1970s	1980s	1980s	Late 1990s	Late 2000s
Scientific Approach	Legal Positivism	Behavioralism	Behavioralism	Rational Choice Theory	Rational Choice Theory	New Institutionalism	Cognitive Psychology
Common Methodology	Legal Analysis	Quantitative – Observational	Quantitative, Qualitative, Network & Text Analysis	Econometric Modeling	Integrated	Integrated	Experimental
Judges' Choices Are Affected By	Legal Rules	Ideological Attitudes Toward Case Facts	Personal Characteristics	Anticipated Actions of Other Actors	Personal Motivations	Legal Norms	Heuristics and Biases

https://doi.org/10.2139/ssrn.4847257 https://www.hilltop.com/dehqb - Open Access - [redacted]

C. The Legal Model

The legal model evolved from the legal formalism and positivism of the early 19th century, which viewed legal logic as internally coherent and self-contained in its perfection. This tradition held that judges could apply a quasi-mathematical reasoning process to the facts of a case in order to arrive at the "correct" legal outcome (Cross 1997). The legal model thus asserts that judges decide cases by neutrally applying legal rules. According to this model, judges are apolitical decision-makers whose rulings are determined by "the law" – that is, statutes, precedents, and legislative intent – as applied to the facts of the case. This conception portrays judges primarily as rule-appliers, not as policymakers or moral agents.

Importantly, the legal model functioned for a long time as an unquestioned baseline: it was not even framed as a "model" until the rise of alternative accounts, particularly the attitudinal model. Until then, the notion that judges (in a democratic system) simply follow the law was seen as self-evident, rather than as a theory to be interrogated.

Although traditional legalism has lost ground among empirically oriented scholars, it remains central to judicial self-understanding, legal education, and constitutional discourse. Its influence endures particularly in jurisdictions shaped by the German legal tradition, where it continues to dominate both legal education and judicial reasoning (Morell 2023). As will be discussed later, neo-institutionalist scholarship has revived legalism in a more nuanced and dynamic form – emphasizing law not as a static set of rules but as a constraining institutional framework that shapes and channels judicial behavior.

This chapter is dedicated to Christoph Engel, who was educated in German universities and has spent his career working in – and often leading – research institutions embedded in jurisdictions shaped by the German legal tradition. For Engel, too, this legalist perspective was a natural point of departure. As the chapter will show, his scholarship has developed in innovative ways that both engage with and transcend this foundational orientation.

D. The Attitudinal Model

The attitudinal model emerged in the 1960s, shaped by the rise of behavioralism in social sciences, realism in legal theory, and early developments

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

in social psychology (e.g., Schubert 1965). Behavioral scholars – particularly in the United States – began systematically analyzing judicial voting patterns and found a strong correlation between U.S. Supreme Court justices' decisions and the political party of the appointing president. This early insight evolved into increasingly sophisticated methods for measuring judicial ideology, partisanship, and voting behavior. In essence, the attitudinal model posits that judges decide cases primarily based on their personal ideological attitudes rather than legal rules or neutral principles. According to this view, the factual patterns in legal disputes serve as heuristic "cues" that trigger judges' preexisting preferences. Judicial decisions are thus best explained by the alignment between case facts and the judge's ideological or partisan orientation (Segal & Spaeth 2002). Within this framework, law is seen not as a constraining or guiding force, but as a flexible rhetorical tool – used to justify a wide range of outcomes and to legitimize decisions that are, at their core, ideologically driven.

The attitudinal model gained prominence through studies of the U.S. Supreme Court, where institutional features such as political appointment prosses, life tenure, control over the docket, and limited external accountability create favorable conditions for value-driven judging. Among the various approaches outlined in Table 1, it remains the dominant paradigm in U.S. scholarship. However, unlike many other theories of law and courts, the attitudinal model has traveled slowly beyond the American context. Although it originated in the 1960s, it is only in the past two decades that systematic efforts have been made to test its assumptions in apex courts of other democracies. These comparative studies are particularly important because they help to isolate the institutional conditions under which attitudinal voting is more or less likely to emerge (Weinshall, Sommer, and Ritov 2018).

A recent paper by Christoph Engel makes a significant contribution to this emerging comparative literature (Engel 2024). the study provides a rigorous empirical examination of whether attitudinal assumptions hold in the context of the German Federal Constitutional Court. Despite its strong powers and the openly political nature of judicial appointments, the German Court is widely perceived as ideologically neutral. Engel investigates this perception using the full set of publicly available data and exploits two independent sources of quasi-random variation to identify causal effects. His findings offer no clear evidence of partisan influence: while some specifications suggest that justices nominated by the FDP and SPD exhibit greater judicial activism, this tendency does not align with the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ideological agendas of the nominating parties. The study thus underscores both the empirical limits of the attitudinal model outside the U.S. context and the potential pitfalls of directly transplanting theoretical frameworks across distinct institutional environments.

E. Identity-Based Approaches

Building on – and at times critiquing – the attitudinal model, identity-based approaches examine how judges' social backgrounds, life experiences, education, and personal characteristics shape their judicial behavior. The central insight is that judges are not disembodied legal reasoners or politicians-in-robes, but individuals whose decisions are influenced by their social identities.

Early research in this tradition focused primarily on professional biographies, asking, for instance, whether former corporate lawyers were more likely to favor powerful economic actors (Tate & Sittiwong 1989). As judicial appointments in many countries have become more socially diverse, the range of identity characteristics studied has expanded to include gender (Weinshall 2021), religion (Koev 2019), race (Sen 2015), ethnicity (Shayo & Zussman 2011), and nationality (Voeten 2008). In some studies, scholars also examine how the identities of litigants or lawyers interact with those of judges to influence outcomes (e.g. Shayo & Zussman 2011).

Here too, Engel's work has advanced the field. In a 2013 paper, Engel and Glöckner employ an experimental design to investigate the mechanisms underlying the effects of professional background. They show that merely assuming the role of a defense attorney or prosecutor produces lasting, role-induced inclinations to acquit or convict. Remarkably, the effect persists even when participants are later asked to predict how a court would rule – after having exited the assigned role – and even when accuracy is incentivized through financial rewards. Engel and Glöckner attribute this effect to "coherence-based reasoning": a tendency to selectively process information in favor of the outcome aligned with one's role, primarily by downplaying contradictory evidence. Although their experiments involved lay participants rather than judges, the findings have clear implications for judicial decision-making and for policy debates on the value of professional diversity in judicial appointments.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

F. Strategic Approach

The strategic approach emerged in the early 1980s, drawing heavily on economics and game theory. It builds on rational choice theory and portrays judges as strategic actors who make calculated decisions to advance their goals and maximize expected outcomes. According to this approach, judges do not make decisions in a vacuum, but rather attend to the preferences and likely actions of other actors, including their judicial colleagues, superiors, politicians, and the public (Epstein & Weinshall 2021). Strategic judges consider the reactions of these actors when making decisions, often modifying their votes or opinions to secure desired outcomes or preserve court legitimacy.

Strategic accounts of judging typically rest on three core assumptions. First, judges are goal-oriented actors who seek to maximize their preferences. These goals may include advancing their policy preferences (as emphasized by the attitudinal model), protecting institutional interests such as the court's public legitimacy, or pursuing personal objectives – some of which are further explored in the labor market model. Second, judges are strategic or interdependent decision-makers: they understand that achieving their goals requires anticipating and responding to the preferences and likely actions of other relevant actors. Third, institutional structures shape and constrain judges' interactions with these actors, influencing both the opportunities and constrains of strategic behaviour.

As a scholar grounded in law and economics, Engel has made numerous contributions to the strategic approach. One example is his work on panel effects – the idea that judges anticipate the preferences of their fellow panel members and adjust their behavior accordingly. In a study of the German Federal Constitutional Court, Engel (2022A) extends the theory of panel effects beyond final votes on the merits. He finds that familiarity among justices significantly influences decisions not only on substantive outcomes but also on procedural matters and on requests for preliminary rulings.

At the same time, as a leading scholar in the application of behavioral economics to legal decision-making, Engel is acutely aware of the limits of purely strategic accounts of judging. As discussed further in Section 9 on the cognitive approach, Engel and Güth (2018) challenge the assumption that judges consistently act in ways that maximize their goals. Instead, they propose a more plausible model of the "satisficing judge" – a decision-maker who aims not for optimal outcomes, but for

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

decisions that meet a threshold of confidence in avoiding significant error.

G. The Labor Market Account

As shown in Table 1, labor market accounts can be understood as a subtheory within the broader strategic approach. Emerging at roughly the same time, this model draws from economic theory to suggest that judges are utility-maximizing actors. However, the labor market account emphasizes personal utility, proposing that judges, like other professionals, are motivated and constrained by a range of primarily nonpecuniary costs and benefits – such as effort, esteem, criticism, and influence.

While most theories of judicial decision-making incorporate preferences to some degree, they differ in the types of motivations they emphasize. Attitudinal and strategic models focus on ideology and partisanship, whereas the legal and neo-institutional models highlight legal norms and institutional constraints. The labor market approach complements these perspectives by foregrounding personal motivations – those rooted in judges' individual goals and self-perceptions – without dismissing legal or political influences.

According to these accounts, judicial behavior is shaped by judges' efforts to maximize utility across five broad personal dimensions:

a. Internal job satisfaction – the internal fulfillment of doing meaningful legal work and maintaining positive relationships with colleagues, clerks, and court staff.
b. External satisfactions – such as prestige, reputation, influence, power, and public recognition (Garoupa & Ginsburg 2015).
c. Leisure – the desire to manage workload and avoid excessive effort.
d. Salary and income – all else equal, judges – like most professionals – prefer higher income and greater material comfort.
e. Promotion – the aspiration to be appointed or elected to a more prestigious judicial or public office.

In two influential papers, Engel has demonstrated the significant impact of internal job satisfaction – particularly the aspiration to perform meaningful legal work – on judicial decision-making. In a quasi-natural experiment, Engel and Weinshall (2020) studied a policy reform that reduced judges' caseloads by 12 to 16 percent. They found that judges used the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

additional time to improve the quality of their work: they wrote longer and more detailed judgments, heard more witnesses, and employed more labor-intensive methods of evidence collection. Similar insights emerge from a laboratory experiment by Engel and Zhurakhovska (2017), in which participants assigned the role of a judge exhibited a stronger inclination to decide objectively. The study highlights what the authors term the "office motive" – a commitment to fulfilling the normative expectations associated with an assigned task.

H. The New Institutional Approach

By the late 20th century, scholars in the neo-institutionalist branch of political science began advocating a shift in focus from individual judicial preferences to the institutional context in which judges operate. This "new institutional" approach emphasizes how the structure, norms, and dynamics of courts as institutions shape judicial behavior. Rather than viewing judges solely as policy/personal-driven actors, it asks how institutional features – such as life tenure, organizational norms, and the judiciary's relationship to other branches of government – condition both the preferences and the choices of judges (Gillman and Clayton 1999).

A key insight of the neo-institutionalist approach is that institutional norms influence not only the strategic calculations of judges but also their identity and sense of role. In contrast to earlier theories, which generally accept institutional constraints as background conditions, neo-institutionalism foregrounds the institution itself as a driver of behavior. It holds that judges internalize institutional logics – including legal norms – which then shape their motivations, not merely limit them.

Moreover, the neo-institutionalist approach emphasizes an informal and expansive conceptualization of institutions, viewing not only formal organizations but also normative frameworks as institutional structures. Within this view, the law itself is understood as an institution – not merely as a set of rules derived from plain meaning, legislative intent, or precedent, but as a shared mindset, language, and reasoning process among jurists. In contrast to the legal model, which treats law as a determinate guide, neo-institutionalism – drawing on legal realism – acknowledges the indeterminacy of law while maintaining that legal norms still meaningfully constrain and shape judicial decision-making.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

These insights resonate with Engel's empirical findings discussed above. In both a laboratory experiment (Engel and Zhurakhovska 2017) and a quasi-natural experiment (Engel and Weinshall 2020), Engel shows that judges – and even lay participants assigned the role of judge – are strongly motivated by the perceived expectations of their office. In particular, they exhibit a pronounced inclination to decide cases objectively and conscientiously, consistent with internalized legal norms. These studies also offer behavioral evidence for the neo-institutionalist claim that judges are not merely constrained by institutional frameworks, but actively shaped by them.

I. Cognitive Theories (Thinking-Fast Judging)

A wave of research in social psychology and behavioral economics gained prominence in the 2000s, challenging the rationalist assumptions underpinning both strategic and legalist models of judicial behavior. Drawing on dual-process theories of cognition (Kahneman 2011), this scholarship emphasizes that human decision-making often relies on mental shortcuts – heuristics, intuitions, and emotions – that facilitate fast and effortless judgments. When applied to judicial behavior, this perspective suggests that even highly trained judges, regardless of their preferences or goals, are not immune to the cognitive limitations. Cognitive approaches thus offer an alternative account: judicial decisions may be shaped less by deliberate legal reasoning or strategic calculation and more by heuristics, intuitions, and biases, particularly under conditions of time pressure or cognitive load.

Experimental studies have demonstrated that judges, like laypeople, are subject to systematic cognitive biases, including anchoring effects, framing effects, and racial bias (Spamann and Klöhn 2016, Guthrie et al. 2000). These findings underscore the psychological dimensions of judicial decision-making, challenging the ideal of the purely rational or neutral judge.

As noted above, Engel is a pioneer in the study of behavioral legal decision-making, particularly through his role as one of the founding figures in applying experimental methods to legal scholarship. In several papers, he demonstrates how judges are vulnerable to cognitive biases (see the overview in his lecture on the topic: Engel 2022). What I find most compelling, however, is his research on how institutional safeguards,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

procedural rules, and judicial norms or training can help mitigate these biases. A key example is his paper on the psychological case for requiring judges to provide reasons for their decisions, published in a volume he co-edited titled *The Impact of Court Procedure on the Psychology of Judicial Decision Making* (Engel 2007).

J. Conclusion: Reflections on a Versatile Scholar and a Dynamic Field

What motivates judges? The answer depends on whom you ask – and when. Scholars working within the legal model argue that judicial decisions are grounded in the law, understood as a combination of plain meaning, legislative intent, and binding precedent. By contrast, proponents of the attitudinal model contend that the law often serves as a veneer, concealing the judge's personal policy preferences and ideological leanings as the true drivers of decisions. Researchers emphasizing identity-based approaches highlight the influence of a judge's background. Strategic models may incorporate attitudinal or identity-based motivations but add a further layer: judges are also strategic actors who anticipate how their decisions will be received by other judges, institutions, or the public, and adjust their decisions to maximize the likelihood that their preferences will be implemented.

The labor market model shifts attention to judges' personal and career-related motivations. Neo-institutionalist theories incorporate many of the above motivations but emphasize the role of institutional legal norms and structures. In this account, "the law" is not a static set of rules but a dynamic interpretive framework. Finally, cognitive theories remind us that judges are not only legal, political, and strategic actors – but also human beings. As such, they are subject to the same cognitive biases and limitations as all individuals.

As demonstrated, Christoph Engel's contributions span the full range of approaches to judicial decision-making. In particular, his work brings a vital comparative perspective to the field and has advanced both strategic and cognitive accounts in innovative and empirically grounded ways. Yet Engel's scholarship does not stop there – he is also leading the field into its future. All the approaches discussed in this chapter assume that judges are human. But as Engel has begun to explore in his recent work, the future may involve hybrid systems of human–machine legal decision-making, and perhaps even fully automated "robot-judges." In a new study (Engel

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

et al. 2025), he empirically examines algorithmic decision-making in the criminal context, revealing systematic biases – such as a tendency to favor detention over release – and arguing that normative judgments must not be hidden in the code. This is precisely the kind of work we have come to expect from Christoph Engel: creative, forward-looking, and always grounded in rigorous empirical inquiry. I can only hope – for the field's sake more than his – that he continues to lead us into the next generation of research.

References

Cross, F. (1997). Political science and the new legal realism: A case of unfortunate interdisciplinary ignorance. *Northwestern University Law Review, 92*, 251–326.

Engel, C. (2007). The psychological case for obliging judges to write reasons. In C. Engel & F. Strack (Eds.), *The impact of court procedure on the psychology of judicial decision making* (pp. 71–110). Nomos Verlagsgesellschaft.

Engel, C. (2022a). Lucky you: Your case is heard by a seasoned panel – Panel effects in the German Constitutional Court. *Journal of Empirical Legal Studies, 19*(4), 1179–1221.

Engel, C. (2022b). *Judicial decision-making: A survey of the experimental evidence* (Discussion Paper No. 2022/6). Max Planck Institute for Research on Collective Goods.

Engel, C. (2024). *The German Constitutional Court – Activist, but not partisan?* (Discussion Paper No. 2024/4). Max Planck Institute for Research on Collective Goods.

Engel, C., & Glöckner, A. (2013). Role-induced bias in court: An experimental analysis. *Journal of Behavioral Decision Making, 26*(3), 272–284.

Engel, C., & Güth, W. (2018). Modeling a satisficing judge. *Rationality and Society, 30*(2), 220–246.

Engel, C., & Weinshall, K. (2020). Manna from heaven for judges: Judges' reaction to a quasi-random reduction in caseload. *Journal of Empirical Legal Studies, 17*(4), 722–751.

Engel, C., & Weinshall, K. (2022). Diffusion of legal innovations. *Annual Review of Law and Social Science, 18*(1), 139–153.

Engel, C., & Zhurakhovska, L. (2017). You are in charge: Experimentally testing the motivating power of holding a judicial office. *Journal of Legal Studies, 46*(1), 1–50.

Engel, C., Klement, A., & Weinshall, K. (2018). Diffusion of legal innovations: The case of Israeli class actions. *Journal of Empirical Legal Studies, 15*(4), 708–731.

Engel, C., Linhardt, L., & Schubert, M. (2025). Code is law: How COMPAS affects the way the judiciary handles the risk of recidivism. *Artificial Intelligence and Law, 33*(2), 383–404.

Epstein, L., & Weinshall, K. (2021). *The strategic analysis of judicial behavior: A comparative perspective*. Cambridge University Press.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Epstein, L., Grendstad, G., Šadl, U., & Weinshall, K. (2023). Introduction to the study of comparative judicial behaviour. In L. Epstein, G. Grendstad, U. Šadl, & K. Weinshall (Eds.), *The Oxford handbook of comparative judicial behaviour*. Oxford University Press.

Epstein, L., Šadl, U., & Weinshall, K. (2021). The role of comparative law in the analysis of judicial behavior. *The American Journal of Comparative Law, 69*(4), 689–719.

Epstein, L., Šadl, U., & Weinshall, K. (2024). Judiciary and judicial behaviour. In T. Ginsburg & K. Whittington (Eds.), *Handbook of comparative political institutions* (pp. 347–361). Edward Elgar Publishing.

Garoupa, N., & Ginsburg, T. (2015). *Judicial reputation: A comparative theory*. University of Chicago Press.

Gillman, H., & Clayton, C. W. (1999). Institutional approaches to Supreme Court decision-making. In C. W. Clayton & H. Gillman (Eds.), *Supreme Court decision-making: New-institutionalist approaches* (pp. 1–14). University of Chicago Press.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2000). Inside the judicial mind. *Cornell Law Review, 86*, 777–830.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Koev, D. (2019). Not taking it on faith: State and religious influences on European Court of Human Rights judges in freedom of religion cases. *Journal of Human Rights, 18*(2), 184–200.

Morell, A. (2023). Legalism and professional norms. In L. Epstein et al. (Eds.), *The Oxford handbook of comparative judicial behaviour*. Oxford University Press.

Segal, J. A., & Spaeth, H. J. (2002). *The Supreme Court and the attitudinal model revisited*. Cambridge University Press.

Sen, M. (2015). Is justice really blind? Race and appellate review in U.S. courts. *Journal of Legal Studies, 44*(1), 187–229.

Shayo, M., & Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics, 126*(3), 1447–1484.

Spamann, H., & Klöhn, L. (2016). Justice is less blind, and less legalistic, than we thought: Evidence from an experiment with real judges. *Journal of Legal Studies, 45*(2), 255–280.

Tate, C. N., & Sittiwong, P. (1989). Decision making in the Canadian Supreme Court: Extending the personal attributes model across nations. *Journal of Politics, 51*(4), 900–916.

Voeten, E. (2008). The impartiality of international judges: Evidence from the European Court of Human Rights. *American Political Science Review, 102*(4), 417–433.

Weinshall, K. (2021). Courts and diversity. *The Law & Ethics of Human Rights, 15*(2), 187–220.

Weinshall, K., Sommer, U., & Ritov, Y. (2018). Ideological influences on governance and regulation. *Regulation & Governance, 12*(3), 334–352.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Familiarity Breeds Contempt or Deference? Homage to Christoph Engel*

Patrick C. Wu, Chang-Ching Lin & Yun-chien Chang

Christoph Engel is a pioneer in assessing the impact of familiarities among judges on case outcomes. In particular, Engel finds that judges on the German Constitutional Court working together for more cases are more likely to declare statutes unconstitutional. Inspired by this work, we examine the behavioral pattern in a different context: whether judges in Taiwan's ordinary courts are more likely to reverse lower court decisions when they have sat in the same panel for more cases. Based on the random assignment of cases in the Taiwanese court and a rigid panel-formation rule, we estimate that a 1% increase in mean, median, minimum, or maximum co-working experience (about 1 to 3 cases sitting together) is associated with an increase in the probability of reversal by 0.02 to 0.05 percentage points. Personal-knowledge familiarity shapes judicial behavior.

A. Introduction

The common adage "familiarity breeds contempt" is often accepted as truth. However, in the context of the judiciary, does this maxim truly apply? Recent research in judicial behavior suggests otherwise, indicating that familiarity among judges can lead to deference, sometimes veering into favoritism (Lin and Chang 2025), and can also foster cooperation (Swalve 2022). This study, utilizing an extensive dataset spanning 16 years from a civil-law system with career judges, explores the complexities of judicial familiarity. Our goal is to unravel the paradox of familiarity, highlighting both its benefits, such as the acquisition of tacit knowledge, and its pitfalls, including the potential for slacking and cognitive biases. In

* An earlier version of this chapter is part of another longer paper with a similar title. All of our gratitude expressed there applies here. This chapter has also been delivered as part of Chang's keynote speech at the 2025 Asian Law & Economics Association Annual Meeting. We thank Ben Chen, Lauren Lin, John Liu, and Alan Miller for helpful comments.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

doing so, we aim to provide a better understanding of familiarity's multifaceted role within the judiciary.

Currently, the study of judicial behavior is informed by six primary, and at times intersecting, approaches: the attitudinal model, legalism, the "thinking fast" judging, identity accounts, the labor market model, and strategic accounts (Epstein and Knight 2024). This research specifically contributes to the discourse on familiarity within the ambit of the labor market model. A pioneer in the study of familiarity in court is none other than the man of the hour, Christoph Engel (2022).

Familiarity, while closely linked to a more well-known concept "collegiality," can, in specific contexts such as ours, present a tension with it. Collegiality entails efforts to foster positive relationships among coworkers, notably among judges within the same court. The rich body of literature on collegiality (e.g., Cross and Tiller 1998; Epstein and Knight 2017: 327) demonstrates its significant effects, such as changing votes (Anwar, Bayer, and Hjalmarsson 2019) and reducing rates of dissent (e.g., Epstein, Landes, and Posner 2011). Unlike collegiality, which is fostered through deliberate or inadvertent efforts to enhance the work environment, familiarity refers to a scenario where key agents have a deep acquaintance with each other. Swalve (2022: 223) argues that familiarity is antecedent to collegiality, but Epstein and Knight (2024) highlight that judges often cannot choose their panel members or office locations – key factors for developing familiarity. Yet, they retain the autonomy to foster or withdraw from collegial efforts (see also Nash 2022: 1576–1582). Collegiality and familiarity, therefore, are distinct in specific scenarios.

Familiarity, as discussed in prior studies on judicial behavior, embodies two distinct interpretations leading to divergent normative consequences. These interpretations have been, at times, ambiguously amalgamated under the single concept of familiarity. The notion we designate as "social-identity familiarity" stems from social identification (Budziak 2016; Wu, Lin, and Chang 2026). On the other hand, "personal-knowledge familiarity" emerges from agents' implicit understanding of their collaborators (Littlepage, Robison, and Reddington 1997; Sanbonmatsu et al. 2012) and business matters (Ang, de Jong, and van der Poel 2014; Wang and Yin 2018), often nurtured through trust, diminished conflict, and enhanced communication during collaboration (Jones and George 1998; Swalve 2022: 225–226). Personal-knowledge familiarity arises from individuals' ongoing Bayesian updates of their perceptions of co-workers, leading to superior outcomes. These two facets

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of familiarity can coexist. Our sister work (Wu, Lin, and Chang 2026) highlights the role of social-identity familiarity. This chapter instead explores whether personal-knowledge familiarity shapes judicial behavior.

This study analyzes over 80,000 substantive appellate civil decisions from second-instance courts in Taiwan to explore the impact of personal-knowledge familiarity on judicial outcomes. The random assignment of cases in Taiwan's court system provides a unique opportunity for causal inference. Our findings reveal that personal-knowledge familiarity significantly impact judicial decisions in the Taiwanese appellate courts, as panels comprising judges from different administrative sections reversed decisions less frequently than those from the same administrative section. Despite being part of the same court, judges from different administrative sections rarely work together, implying that their hesitancy to overturn lower court rulings stems from unfamiliarity rather than collegiality, highlighting the influence of personal-knowledge familiarity.

The rest of this chapter is structured as follows: Part II frames our hypothesis within existing literature. Part III details our regression model specifications. Part IV outlines the four datasets employed. Part V interprets our results. Part VI offers concluding remarks.

B. Research Question and Literature Review

Our research question delves into the dynamics of judicial collaboration within panels, specifically examining whether judges who collaborate less frequently are less likely to reverse decisions compared to those who have a history of frequent collaboration and are, therefore, more familiar with one another. Insights from studies on other judicial systems, such as the German Constitutional Court, suggest that familiarity among judges can significantly impact their decision-making processes. For instance, Swalve (2022: 240) observes that justices who have served together longer on the German Constitutional Court tend to produce more detailed (and presumably higher quality) opinions, likely because these justices feel more comfortable sharing candid feedback and are less inclined to engage in conflicts. Similarly, Engel (2022) finds that the longer justices have served together on the same panel, the more likely they are to make bold decisions, such as declaring a law unconstitutional, indicating that familiarity fosters a conducive environment for making challenging rulings.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Translating these insights to the Taiwanese judicial context, where the culture among judges similarly disfavors reversals due to the stigma of being reversed,[1] we hypothesize that a reversal – analogous to a declaration of unconstitutionality on a smaller scale – is more probable among judges who are more familiar with each other. Through more regular interaction, judges familiar with one another have developed a level of trust that might embolden them to candidly express their concerns about defects in the lower court decisions. Personal-knowledge familiarity helps overcome the inertia of maintaining the lower court decisions – as writing an opinion that dismisses appeals is easier than writing an opinion that identifying errors in judicial decisions. Personal-knowledge familiarity thus pushes judges to undertake the more contentious act of reversing a decision – as judges are in the same career network with other judges, and no judges like being reversed. Hence, we expect to find the following:

Hypothesis: The reversal rate among cases handled by panels with higher co-working experience is higher than those with lower co-working experience.

C. Empirical Strategy

This part elaborates our empirical method. Random assignment of judicial cases in Taiwan constitutes the core of our empirical strategy. We do not repeat the detail here due to word limitation – Lin and Chang (2025) has demonstrated that the data set we used contain randomly assigned cases. Below we lay out the specifications of our linear probability models. Only necessary institutional background regarding the Taiwanese legal system is contained here. For more explanation, see Wu, Lin, and Chang (2026).

We ran a linear probability model with robust standard errors clustered by assigned judges. Our regression models take the following form:

1 See the discussion of reversal aversion in the American context in Epstein, Landes, and Posner (2013: 49) and Posner (2008: 70–71).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

$$
\begin{aligned}
D_{ihdt} = & \delta\, CoWork_{ihdt} + \sum_{j=\{A,P,S\}} (\beta_{1j} TP^{j}_{ihdt} + \beta_{2j}\, Home^{j}_{ihdt} \\
& + \beta_{3j}\, TP^{j}_{ihdt} \times Home^{j}_{ihdt} + \mu^{j}_{h}) \\
& + X_{ihdt}\gamma + \mu_d + \eta_t + \varepsilon_{ihdt}
\end{aligned} \tag{1}
$$

where the dependent variable D_{ihdt} is whether the case i appealed from District Court d at time t was reversed by a High Court (with judge h as the assigned, presiding, or side judges), with a complete or partial reversal coded as 1 and a dismissal of appeals coded as 0.[2]

Our treatment variable is a continuous measure of the co-working experience among the three panel judges. More specifically, we calculated the mean, median, maximum, and minimum number of cases in which each of the three pairs of judges had sat together. If a case in our dataset is the first time a pair has sat together, their co-working experience is coded as 1, and so forth. Each of the four measures is used separately in a regression model. We examine the effects of all four measures, rather than committing to a single one, because it is unclear which best reflects judicial familiarity. Perhaps the minimum co-working experience matters most because if even one pair of judges is unfamiliar with each other, they may prefer to withhold their true opinions. Perhaps the maximum co-working experience matters most because if one pair of judges has reached a high threshold of familiarity, they may be able to dominate panel dynamics. Perhaps the mean co-working experience is the most appropriate because it incorporates all three pairs of co-working experiences. Or perhaps the median co-working experience is preferable because it is less sensitive to outliers.

Our research design aims to causally identify the effect of judicial co-working experience on reversal rates. We rely on the random assignment of cases to judges and the pre-determined panel-formation rule for causal identification. Whether the three judges in a panel are familiar with one another should be independent of the probability of reversal. When one of the three judges determined by the case-assignment algorithm and pre-determined rule leaves the panel – due to recusal, illness, vacation, maternity leave, or similar circumstances – the judge is replaced by another

2 The Civil Division in High Courts review cases *de novo*; that is, High Courts can reverse lower court decisions on the basis of either questions of law or questions of fact.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

judge, again pursuant to a pre-determined order. These substitutions are highly unlikely to be driven by a judge's predisposition toward reversal. Indeed, if a judge wished to influence the outcome, the most effective strategy would be to remain on the panel, rather than introducing a replacement. Our understanding is that there is also a strong norm against altering panel composition. Granted, we cannot rule out the possibility of strategic maneuvering entirely, but we note that such concerns equally affect all real-world random assignment of judicial cases. Moreover, even if a judge deliberately withdrew to bring in an unfamiliar replacement (for example, to maintain the lower court's decision), such a maneuver would not affect all four co-working measures simultaneously. For instance, by definition, a change in the minimum co-working experience would not alter the maximum co-working experience, and vice versa. We are therefore confident in the validity of our causal identification. Based on our hypothesis that greater familiarity among judges increases the likelihood of reversal, we expect all four measures to have positive and statistically significant coefficients.

We include several variables from our sister work (Wu, Lin, and Chang 2026). On the one hand, we explore whether co-working experience still matters after other factors related to judicial behavior are taken into account. On the other hand, we check whether taking into account the effect of familiarity among panel judges undercuts the transitorily promoted ("TP") judge effect and home court effect discussed in our sister work (Wu, Lin, and Chang 2026). More specifically, these variables include $TP^{j}_{ihdt}, j = \{A, P, S\}$, which are dummy variables, representing whether assigned judges (A), presiding judges (P), and side judges (S), respectively, are TP judges. That is, for instance, TP^{A}_{ihdt} is a dummy variable that equals 1 if the assigned judge in the case was at the time a transitorily promoted judge, and it equals 0 if the assigned judge was a permanent judge at High Court at the time of the case. In addition, $Home^{j}_{ihdt}, j = \{A, P, S\}$ are also all dummy variables representing whether the appealed decision comes from the (former) home court of assigned judges (A), presiding judges (P), and side judges (S), respectively. That is, for instance, $Home^{A}_{ihdt}$ equals 1 if the assigned judge served in district court d before being promoted to an appellate court, and the appealed case in question was rendered by district court d. Another set of variables, $TP^{j}_{ihdt} \times Home^{j}_{ihdt}$, is interaction terms of the previous two sets. X_{ihdt} represents other control variables and μ^{j}_{h}, μ_{d}, and η_{t} represent judge, district court, and year fixed effects, respectively.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

See Wu, Lin, and Chang (2026) for a more detailed explanation of our control variables.

D. Data

This article, like our sister work (Wu, Lin, and Chang 2026), uses all the substantive,[3] appellate[4] decisions rendered by the civil sections in the five High Courts in Taiwan between July 1, 2003, and June 30, 2019. Our data start in 2003 because the TP program started in 2003 and the administrative data in and after 2003 are more reliable. Our data end at mid-2019 because we only have full texts of all decisions before June 30, 2019. In total, our data set includes 84,335 court decisions, and here we use a subset of them (74,237 court decisions), which contains complete information.

The information required by our research design comes from multiple official sources. First, in 2019, the Judicial Yuan, the administrative organ of the judicial system in Taiwan, provided the full texts of all decisions rendered by the civil sections before June 30, 2019. Anyone may manually check and download cases from the official website of the Judicial Yuan.[5]

Second, the Judicial Yuan coded a plethora of variables regarding each dispute filed in courts. Its administrative data set can be freely downloaded from its website.[6] We take advantage of this data set to compute the number of times a pair of judges have sit in the same panel, and use this statistics in Table 1 and Figure 1. Included in the tallying of this co-working experience are appellate decisions (including procedural dismissals and substantive judgments but excluding settlements and plaintiff withdrawals) rendered by the civil sections in the High Courts in Taiwan between July 1, 2003, and the date each of the cases in our dataset was decided.

---

3 Appeals that were procedurally dismissed or transferred for jurisdictional reasons are excluded. We also excluded cases that were remanded by the Supreme Court, as the appellate court does not have the full discretion regarding whether to reverse the district court decisions.

4 In a small set of cases, the High Courts are legally required to be the court of first instance. These cases are excluded.

5 *See* https://judgment.judicial.gov.tw/FJUD/default.aspx.

6 https://opendata.judicial.gov.tw/.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 1 LPM Regression Results for Continuous Treatment Variables
Dependent variable: Reverse District Court Decision =1; Dismiss Appeal =0.

	(1)	(2)	(3)	(4)
	Mean	Median	Max	Min
Mean number of cases sitting in panels (ln)	**0.050*****			
	(0.004)			
Median number of cases sitting in panels (ln)		**0.023*****		
		(0.002)		
Max number of cases sitting in panels (ln)			**0.049*****	
			(0.004)	
Min number of cases sitting in panels (ln)				**0.020*****
				(0.002)
Other Treatment Variables	Yes	Yes	Yes	Yes
Fixed Effects and Other Controls	Yes	Yes	Yes	Yes
Observations	74237	74237	74237	74237
Reversal Rate	38.3	38.3	38.3	38.3
R^2	0.053	0.052	0.053	0.051
Adjusted R^2	0.040	0.039	0.040	0.039

Notes: Robust standard errors, clustered by assigned judges, are in parentheses. This table reports results from linear probability models (LPMs). Other treatment variables include the nine variables on TP status and home court used in our sister paper. Fixed effects for Case Types, Case Years, District Courts, Assigned Judges, Presiding Judges, and Side Judges are included. The reported reversal rate is the mean reversal rate for all the observations used in the regression. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Third, to identify all the appellate judges and the district courts they have served in before moving to the appellate court, we cross-checked the court decision data with another source of information: the meeting minutes of the Personnel Review Committee at the Judicial Yuan,[7] which informed us of which judges were transferred or promoted to which courts at what time.

7 The meeting minutes are available on-line at https://www.judicial.gov.tw/tw/lp-1915-1.html and https://jirs.judicial.gov.tw/GNNWS/Personnel.asp?page=10. These links may only be available when viewers have a Taiwan IP address. Minutes for meetings held before 2002 are available in print only.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

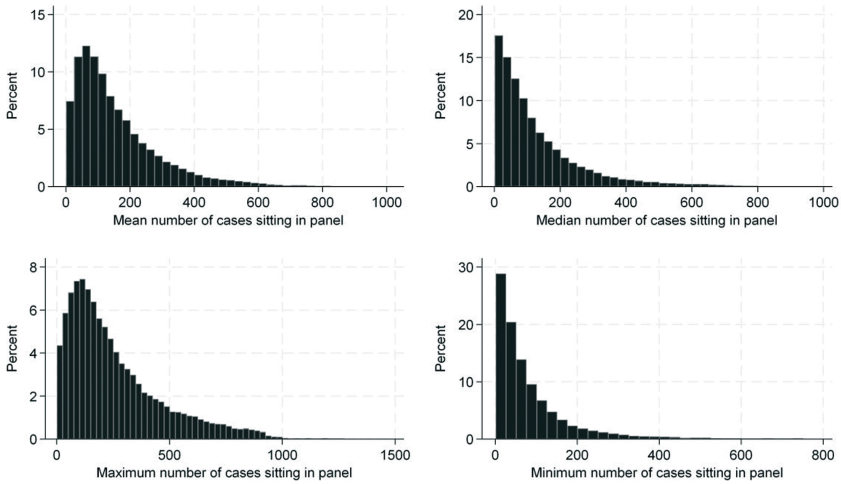


Figure 1 Frequency of Collaborations among Judges (Continuous Variable)
Notes: *N*=84,335 in each sub-plot. The unit of observations is the mean, median, maximum, or minimum number of cases of the three pair of judges sitting in a High Court panel. The number equals 1 if Judge A and Judge B sit in a panel for the first time.

Fourth, most High Courts are composed of multiple civil sections and generally section members form a three-judge panel to handle appealed cases. Each year, each High Court announced on its website the section affiliation of all judges, with whom each judge will form a panel, and the order of judges who will fill in for an absent judge in a panel. Such information from previous years has been removed from the websites of High Courts. We asked each High Court for such information and acquired it for most studied years.

There are three levels of courts in Taiwan that handle civil cases: 22 District Courts, 6 High Courts, and the Supreme Court. In this chapter, we refer to the former two as the district courts and the appellate courts or District Courts and High Courts. 20 District Courts and 5 High Courts are included in this study. Only the two small district courts not on the Island of Taiwan and their High Court are excluded because that High Court has only three judges, so it is not ideal for our research design.

We ran four linear probability models with almost the same specification. The main differences across the models are the continuous mea-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

surement of judicial co-working experience. The observations used in the regression include the 69,422 "clean" observations used in Column (1) of each reported regression tables in our sister article, plus 4,815 cases not included there. These 4,815 cases were excluded there because they were handled by "mixed panels" which consist of judges from different administrative sections. By contrast, the 69,422 cases used there were handled by "pure panels" which consist of judges from the same administrative sections. As shown in Figure 2, assigned and side judges in mixed panels tend to be less familiar with each other than those in pure panels. Mixed panels are an important, but not exclusive, source of the unfamiliarity.

E. Findings and Discussion

This Part discusses the effect of judicial co-working experience. As Table 1 and Figure 3 show, no matter how co-working experience among the three judges in a panel is measured (the mean, median, maximum, or minimum number of cases sitting together), the variables are statistically significant at the 0.1 % level and are positive. As the coefficients are between 0.02 and

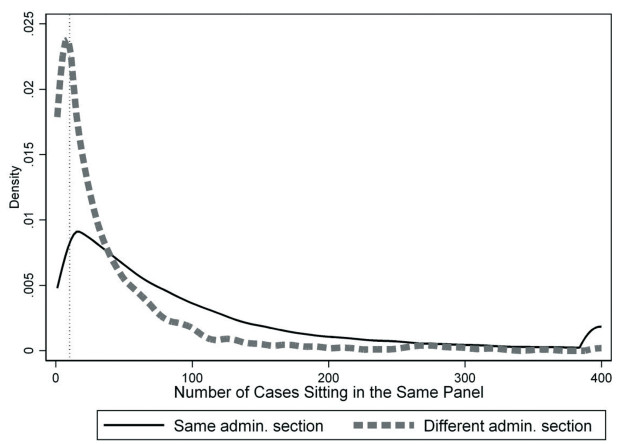


Figure 2 Frequency of Collaborations among Judges (Pure versus Mixed Panels)
Notes: This kernel density plot shows the distribution of the number of cases the assigned judge in a case has sat with the first side judge in that case. There are 81,182 unique pairs of the judge combinations. 75,848 pairs of judges work in the same administrative sections at the time of the panel deliberations, whereas 5,334 pairs of judges work in different administrative sections. The vertical dotted line marks 10 cases.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

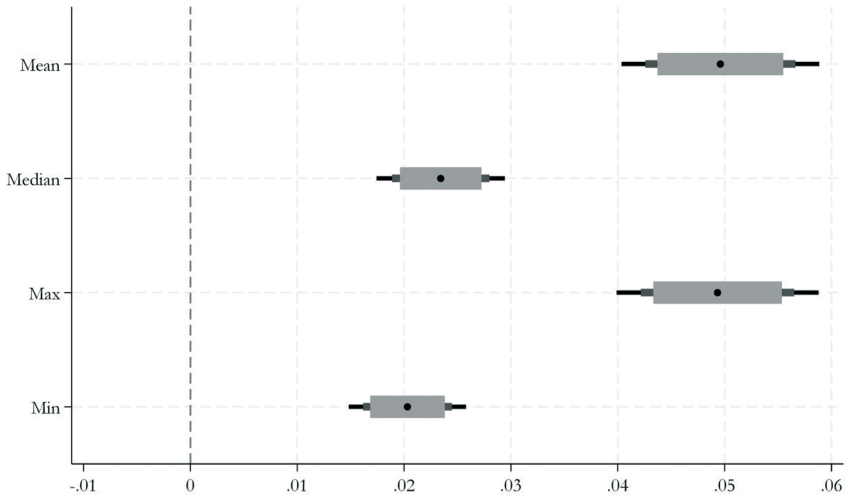


Figure 3 Coefficient Plots for Number of Cases Sitting Together in Panels
Notes: The thickest, shortest lines represent the 90 % confidence interval; the thinner, longer lines represent the 95 % confidence interval; the thinnest, longest lines represent the 99 % confidence interval. Horizontal coefficient confidence intervals that cross the vertical, dashed line at 0 mean that the coefficient is not statistically significant at that level.

0.05, this means that a 1 % increase in co-working experience (about 1 to 3 cases sitting together) is associated with an increase in the probability of reversal by 0.02 to 0.05 percentage points.

Familiarity does appear to shape judicial behavior. Judicial panels in which judges have worked together less often reverse less frequently, consistent with the theory of *personal-knowledge familiarity* – judges working together often know better one another's working habit, presumably trusting one another more, and thus would be able to overcome the status quo bias by reversing lower court decisions. Our earlier work (Lin and Chang 2025) shows that the reversal rate of appealed decisions penned by a trial judge who has served with the reviewing appellate judges in the High Courts is lower than that authored by a trial judge who has not served with the reviewing appellate judges in the High Courts, despite having served in the High Court. *Personal-knowledge familiarity* again appears to be the most promising explanation. In this chapter, the personal-knowledge familiarity arises from a group of three judges knowing one another's working habits, viewpoints, etc. and the decision to reverse or not does not directly affect these three reviewing judges. By contrast,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the personal-knowledge familiarity in our earlier work arises from the (net)working relationship between the reviewed judges and the reviewing judges. Among them, at least the reputation of the reviewed judges is at stake. The benevolent interpretation is that reviewing judges defer to good reviewed judges the former knows personally, while a darker interpretation is that the former gives the latter a favor, at the expense of the rule of law.

By contrast, our sister work demonstrates that *social-identity familiarity* shapes judicial behavior in Taiwan. Taken together, our three articles show that what has often been grouped under the umbrella term "familiarity" actually consists of two distinct forces: personal knowledge and social identity. Both have shaped judicial behavior in Taiwan. Moreover, familiarity can and should also be distinguished from "collegiality," a well-studied topic. As this Chapter shows, collegiality – typically understood as the motivation to maintain an amicable working environment – does not fully explain why judges with more co-working experience are more likely to reverse lower-court decisions. If collegiality were the main driver, then judges who had worked together frequently might reverse more, reverse less, or reverse at the same rate.

In the context of our paper, personal-knowledge familiarity may or may not be normatively problematic. The key issue is the baseline: what is the appropriate level of deference? One view is that judges on a panel who have worked together for a long time exhibit the proper degree of deference to trial courts, whereas a panel of judges with little co-working experience show excessive deference and thereby wrong some appellants. We are inclined toward this view. If so, the solution would be to minimize the number of mixed panels or panels in which all three judges have rarely worked together. One way to achieve this is to expand court sections. Our sister paper (Table 1) shows that 12 % of administrative sections have only three judges. If any one of them is unavailable, an out-of-section judge must be added to the panel. Increasing the size of an administrative section to at least four judges could mitigate this problem. Another strategy is to prioritize placing new judges into administrative sections with more senior judges, rather than grouping them together or pooling them with predominantly junior judges, in order to minimize reductions in the levels of paired working experience.

By contrast, if the proper baseline is that judges unfamiliar with one another demonstrate the correct level of deference, then the implication is that a panel composed entirely of long-term colleagues would reverse too

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

readily. Under this view, the court system might consider abandoning the administrative section hierarchy and placing all judges within a High Court into a single pool for panel formation, so that each judge would sit with all others – but none too frequently. Nonetheless, if there is a certain threshold of co-working experience beyond which panels take reversals too lightly, this design might have the undesirable effect of raising the overall reversal rate. Moreover, eliminating administrative sections would make it impossible to establish specialized panels that handle cases requiring expertise developed through consistent exposure to specific subject areas.

Compared with the personal-knowledge familiarity documented in our prior work (Lin and Chang 2025) – where a cynical observer might easily disparage the favoritism shown when reviewing judges favor trial judges they know – the most pessimistic interpretation of the phenomenon examined here, under the second baseline, is that judges familiar with one another become overly aggressive in reversing. This outcome is difficult both to evaluate normatively (since reversed cases may still be correctly decided upon remand) and to explain mechanistically. We currently see no clear way to address this potential problem head-on, to the extent it exists. Although we are not inclined to adopt the second baseline, it remains an open question requiring further empirical investigation.

In any event, the three collaborative works we have produced demonstrate that familiarity shapes judicial behavior in Taiwan, and this phenomenon warrants greater attention from both academics and policymakers.

F. Conclusion

Drawing on all substantive High Court decisions in Taiwan between July 2003 and June 2019, randomly assigned to judges, we provide new evidence on familiarity and judicial behavior. Panels composed of judges with limited co-working experience reverse less often than panels with substantial co-working experience. Consistent with our recent work (Lin and Chang 2025), we find that personal-knowledge familiarity influences judicial decision-making.

Familiarity brings contempt or deference? Our study shows that unfamiliarity among panel judges lead to lower reversal rates of lower court decisions. This somewhat surprising phenomenon is first reported by Christoph Engel, a towering figure in many fields. With his long and dec-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

orated career, he did not visit Asia until 2018, when we invited him to Taiwan. His first stop is Tainan, the ancient capital of Taiwan, and the hometown of one of us (Lin). Christoph was put in a unique house that encircles a tree. This was not by design. The owner of the B & B, after knowing Christoph's stature, upgraded him to the best property for free. Christoph, an opera fan, finds it amusing to stay in a house like Hunding's (see *Die Walküre* by Richard Wagner). Needless to say, Christoph's classy dressing style and academic insights left the Taiwanese audience with a great impression. Does familiarity breed contempt or deference? In Christoph's case, familiarity breeds affection.

References

Ang, James, Abe de Jong, and Marieke van der Poel. 2014. Does Familiarity with Business Segments Affect CEOs' Divestment Decisions? *Journal of Corporate Finance* 29:58–74.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2019. Politics in the Courtroom: Political Ideology and Jury Decision Making. *Journal of the European Economic Association* 17 (3):834–875.

Budziak, Jeffrey. 2016. Promotion, Social Identity, and Decision Making in the United States Courts of Appeals. *Journal of Law and Courts* 4 (2):267–290.

Cross, Frank B., and Emerson H. Tiller. 1998. Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals. *The Yale Law Journal* 107 (7):2155-2176.

Engel, Christoph. 2022. Lucky You: Your Case Is Heard by a Seasoned Panel – Panel Effects in the German Constitutional Court. *Journal of Empirical Legal Studies* 19 (4):1179–1221.

Epstein, Lee, and Jack Knight. 2017. The Economic Analysis of Judicial Behavior. In *The Oxford Handbook of U.S. Judicial Behaviour* 320–339, edited by Lee Epstein and Stefanie A. Lindquist. Oxford: Oxford University Press.

– – – . 2024. How Personal Motivations Affect Judges' Decisions. In *The Oxford Handbook of Comparative Judicial Behaviour* 199–222, edited by Lee Epstein, Gunnar Grendstad, Urška Šadl and Keren Weinshall. Oxford: Oxford University Press.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2011. Why (and When) Judges Dissent: A Theoretical and Empirical Analysis. *Journal of Legal Analysis* 3 (1):101–137.

– – – . 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, MA: Harvard University Press.

Jones, Gareth R., and Jennifer M. George. 1998. The Experience and Evolution of Trust: Implications for Cooperation and Teamwork. *Academy of Management Review* 23 (3):531–546.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Lin, Chang-ching, and Yun-chien Chang. 2025. Learning or Networking? The Causal Effect of Judges Sitting by Designation. *Journal of Legal Studies* 54 (1):43–81.

Littlepage, Glenn, William Robison, and Kelly Reddington. 1997. Effects of Task Experience and Group Experience on Group Performance, Member Ability, and Recognition of Expertise. *Organizational Behavior and Human Decision Processes* 69 (2):133–147.

Nash, Jonathan Remy. 2022. Measuring Judicial Collegiality Through Dissent. *Buffalo Law Review* 70 (4):1561–1635.

Posner, Richard A. 2008. *How Judges Think*. Cambridge, MA: Harvard University Press.

Sanbonmatsu, David M., Bert N. Uchino, Kevin K. Wong, and Joon Yong Seo. 2012. Getting Along Better: The Role of Attitude Familiarity in Relationship Functioning. *Social Cognition* 30 (3):350–361.

Swalve, Tilko. 2022. Does Group Familiarity Improve Deliberations in Judicial Teams? Evidence from the German Federal Court of Justice. *Journal of Empirical Legal Studies* 19 (1):223–249.

Wang, Ye, and Sirui Yin. 2018. CEO Educational Background and Acquisition Targets Selection. *Journal of Corporate Finance* 52:238–259.

Wu, Patrick C., Chang-Ching Lin, and Yun-chien Chang. 2026 forthcoming. How Judges' Social Identity Impacts Reversal Decisions? Evidence from the Transitory Promotion of Judges. Journal of Law, Economics, and Organization.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die Gleichheitsrechtsprechung des Bundesverfassungsgerichts
Eine empirische Analyse

*Niels Petersen**

Der Jubilar hat sich in besonderer Weise um die empirische Rechtsforschung in Deutschland und Europa verdient gemacht. Er hat nicht nur selbst zahlreiche wichtige Beiträge zur empirischen Forschung geleistet.[1] Vielmehr hat er durch die Ausbildung zahlreicher Schüler auch zur Verbreitung der empirischen Rechtsforschung im deutschsprachigen Raum erheblich beigetragen. Daher liegt es nahe, zur Feier des Jubilars ebenfalls einen empirischen Beitrag beizusteuern. In diesem Aufsatz stelle ich die Ergebnisse einer empirischen Analyse der Gleichheitsrechtsprechung des Bundesverfassungsgerichts vor.[2] Sie beruht auf einer umfassenden Auswertung aller Urteile des Gerichts zum Gleichheitssatz des Grundgesetzes und war Teil des rechtsvergleichenden Forschungsprojekts „Correcting Inequality through Law".[3]

* Der Beitrag entstammt einem Projekt, das mit Forschungsgeldern des Europäischen Forschungsrats (ERC) unter dem Horizon 2020 Forschungs- und Innovationsprogramm der EU gefördert wurde (grant agreement No. 817652).

1 Zur empirischen Forschung zum Verfassungsrecht s. etwa *Christoph Engel*, Das legitime Ziel in der Praxis des Bundesverfassungsgerichts: Eine quantitative Analyse der Entscheidungen des Jahres 2011, in: Matthias Jestaedt/Oliver Lepsius (Hg.), Verhältnismäßigkeit, 2015, 97; *ders.*, Does Efficiency Trump Legality? The Case of the German Constitutional Court, in: Yun-Chien Chang (Hg.), Selection and Decision in Judicial Process Around the World, 2019, 261; *ders.*, Lucky you: Your case is heard by a seasoned panel, Journal of Empirical Legal Studies 19 (2022), 1179; *ders.*, Grundrechtskonjunkturen, JZ 2022, 593; *ders./Johannes Kruse*, Kommentar ohne Autor: Können Sprachmodelle das Kommentieren übernehmen?, JZ 2024, 997; *Christoph Engel*, The German Constitutional Court – Activist, but not Partisan? (MPI Preprint 2024); *ders.*, Gewichtsformel–Wörtlich genommen: Ein empirischer Test mit der Hilfe eines Sprachmodells (MPI Preprint 2024).

2 Auch der Jubilar hat sich mit dem BVerfG in seiner empirischen Forschung intensiv beschäftigt. S. nur die Beiträge in Fn. 1.

3 Die Ergebnisse dieses Forschungsprojekts sind publiziert bei *Niels Petersen*, Equality's Guardians: How Courts Conceptualize Equal Protection and Non-Discrimination Guarantees, 2025.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Eine spezifische Analyse der Gleichheitsrechtsprechung des BVerfG bietet sich aus verschiedenen Gründen an. Zunächst weicht die Dogmatik der Gleichheitsrechte erheblich von der Dogmatik der Freiheitsrechte ab, auch wenn sich in jüngerer Zeit Annäherungen beobachten lassen. Die Prüfung der Freiheitsrechte folgt dem klassischen Dreischritt von Schutzbereich-Eingriff-Rechtfertigung, wobei die Verhältnismäßigkeitsprüfung regelmäßig den Kern der Prüfung bildet.[4] Demgegenüber hat sich die Gleichheitsdogmatik seit den Anfängen in der Nachkriegszeit erheblich entwickelt – von der Willkürkontrolle[5] über die neue[6] und neueste Formel[7] bis hin zur Stufenlosformel,[8] die derzeit vom Gericht angewandt wird.[9] Zudem hat Art. 3 GG für die Rechtsprechung des Gerichts eine enorme Bedeutung.[10] Wenn man lediglich die Fälle betrachtet, in denen ein Gesetz auf Basis eines Grundrechts für verfassungswidrig erklärt wurde, dann gehen ca. 40 % der erfolgreichen Entscheidungen auf eine Verletzung von Art. 3 GG zurück.[11] Im Folgenden soll zunächst das Forschungsdesign der Studie erklärt werden. Anschließend erfolgt eine eingehende empirische Analyse der Gleichheitsrechtsprechung, die verschiedene Aspekte näher untersucht. Der letzte Abschnitt zieht einige Schlussfolgerungen aus dem Befund.

4 S. bspw. *Fritz Ossenbühl*, Abwägung im Verfassungsrecht, DVBl. 1995, 904 (906); *Josef Isensee*, Bundesverfassungsgericht – quo vadis?, JZ 1996, 1085 (1090); *Franz Josef Lindner*, Theorie der Grundrechtsdogmatik, 2005, S. 217; *Matthias Jestaedt*, Phänomen Bundesverfassungsgericht: Was das Gericht zu dem macht, was es ist, in: ders. u. a. (Hg.), Das entgrenzte Gericht, 2011, 77 (146).

5 BVerfGE 1, 14 (52).

6 BVerfGE 55, 72 (88).

7 BVerfGE 88, 87 (96).

8 BVerfGE 129, 49 (69).

9 S. dazu *Christoph Brüning*, Gleichheitsrechtliche Verhältnismäßigkeit, JZ 2001, 669 (669); *Simon Kempny/Philipp Reimer*, Die Gleichheitssätze – Versuch einer übergreifenden dogmatischen Beschreibung ihres Tatbestands und ihrer Rechtsfolgen, 2012, S. 105 ff.; *Lerke Osterloh*, Der Gleichheitssatz zwischen Willkürverbot und Grundsatz der Verhältnismäßigkeit, in: FS Michael Kloepfer, 2013, 139 (141); *Jost Pietzcker*, Der allgemeine Gleichheitssatz, in: Detlef Merten/Hans-Jürgen Papier (Hg.), Hdb GR V, 2013, § 125, Rn. 40 ff.; *Gabriele Britz*, Der allgemeine Gleichheitssatz in der Rechtsprechung des BVerfG, NJW 2014, 346; *Joachim Englisch*, in: Klaus Stern/Florian Becker (Hg.), Grundrechte-Kommentar, 2018, Art. 3, Rn. 11 f.; *Thorsten Kingreen*, in: Wolfgang Kahl/Christian Waldhoff/Christian Walter (Hg.), Bonner Kommentar zum Grundgesetz, 2020, Art. 3, Rn. 321 ff.; *Sigrid Boysen*, in: Ingo von Münch/Philip Kunig (Hg.), Grundgesetz-Kommentar, 2021, Art. 3, Rn. 104 ff.

10 Zur Wichtigkeit von Art. 3 GG s. auch *Engel*, Grundrechtskonjunkturen (Fn. 1), 595.

11 *Niels Petersen*, Germany, in: Equality's Guardians (Fn. 3), 95 (95).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

A. Das Forschungsdesign

Die Analyse, die diesem Beitrag zugrunde liegt, umfasst alle Fälle des Bundesverfassungsgerichts im Zeitraum zwischen 1951 und Ende 2021, in denen das Gericht eine materielle Entscheidung auf Grundlage von Art. 3 oder Art. 6 Abs. 5 GG getroffen hat. Die Stichprobe wurde durch eine JURIS-Recherche zusammengestellt, bei der nach beiden Normen gesucht wurde. Anschließend wurde die Stichprobe manuell um Entscheidungen bereinigt, in denen das Gericht keine materielle Diskussion von Art. 3 oder 6 Abs. 5 GG vornahm – etwa aufgrund fehlender Zulässigkeit. Insgesamt umfasste die Stichprobe 1608 Fälle. Etwas mehr als die Hälfte der analysierten Entscheidungen waren Senatsentscheidungen (52,99 %).

Nach Zusammenstellen der Stichprobe wurden die Fälle analysiert und auf Grundlage eines vorgegebenen Schemas auf bestimmte Variablen hin kodiert. Dabei wurde insbesondere festgehalten, ob eine Verletzung von Art. 3 oder 6 Abs. 5 GG festgestellt wurde, welche konkrete Norm herangezogen wurde, ob die Unterscheidung auf einem problematischen Merkmal und ggf. auf welchem Merkmal sie beruhte und welchem Sachgebiet der Fall zuzuordnen war. Eine erste Auswertung der Fälle wurde durch Studentische Hilfskräfte vorgenommen, die schematisierte einseitige Zusammenfassungen der Fälle erstellt und bereits einige Variablen, wie etwa das Vorliegen einer Grundrechtsverletzung, vorkodiert haben. Anschließend habe ich die endgültige Kodierung aller Fälle vorgenommen.

Wenn man sich auf Basis dieser Daten die Erfolgswahrscheinlichkeiten von Fällen ansieht, wie ich dies im Folgenden tun werde, ist zu beachten, dass die Aussagekraft der Erfolgswahrscheinlichkeiten beschränkt ist. Wie der Jubilar in seinen Studien herausgestellt hat, wird nur ein kleiner Bruchteil der Entscheidungen des Gerichts überhaupt veröffentlicht.[12] Unveröffentlichte Entscheidungen können jedoch in der Analyse nicht berücksichtigt werden. Dies kann zu einer erheblichen Verzerrung führen. Die nicht veröffentlichten Entscheidungen sind ganz überwiegend Entscheidungen, in denen das Gericht gem. § 24 S. 2 BVerfGG von seiner Befugnis Gebrauch gemacht hat, unzulässige oder offensichtlich unbegründete Anträge ohne Gründe zu verwerfen.[13] Erfolgreiche Entscheidungen werden also fast ausnahmslos veröffentlicht, während die große Mehrzahl

12 *Engel*, Grundrechtskonjunkturen (Fn. 1), 494.

13 Ebd., 494 (unter Berufung auf Richterin am BVerfG Susanne Baer).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

der nicht erfolgreichen Entscheidungen nicht veröffentlicht wird. Die errechneten Erfolgswahrscheinlichkeiten werden also deutlich höher sein als die tatsächlichen Erfolgswahrscheinlichkeiten.

So betrafen fast 10 % der Entscheidungen in unserer Stichprobe die Gewährung von Prozesskostenhilfe. Das BVerfG leitet aus Art. 3 Abs. 1 GG den Grundsatz der Rechtsschutzgleichheit vor Gericht ab, demzufolge eine unbemittelte Partei in die Lage versetzt werden muss, „ihre Belange in einer dem Gleichheitsgebot gemäßen Weise im Rechtsstreit geltend zu machen."[14] Dies nimmt das Gericht zum Anlass, zivilgerichtliche Entscheidungen über die Gewährung von Prozesskostenhilfe daraufhin zu überprüfen, ob diese einen zu strengen Maßstab hinsichtlich der erforderlichen Erfolgsaussichten angelegt haben.[15] In unserer Stichprobe hatten diese Fälle auf Überprüfung von Prozesskostenhilfeentscheidungen eine Erfolgswahrscheinlichkeit von 82,67 % (124 aus 150). Eine derart hohe Erfolgswahrscheinlichkeit ist nicht plausibel und lässt sich mutmaßlich damit erklären, dass die große Mehrheit der Entscheidungen, in denen ein Antrag auf Überprüfung einer Prozesskostenhilfeentscheidung als unzulässig oder offensichtlich unbegründet verworfen wird, so dass sie nach § 24 S. 2 BVerfGG nicht veröffentlicht werden.

Um diesem Problem zu begegnen, fokussiere ich mich in der nachfolgenden Analyse vor allem auf Fälle, in denen das BVerfG die Verfassungsmäßigkeit eines Gesetzes überprüft hat. Auch hier besteht keine Garantie, dass nicht erfolgreiche Fälle veröffentlicht werden. Man kann aber in der Regel davon ausgehen, dass das Gericht diese Fälle ernster nimmt, sich ausführlicher mit ihnen auseinandersetzt und daher auch eine negative Entscheidung mit größerer Wahrscheinlichkeit veröffentlicht, so dass die Verzerrung geringer ausfällt.[16] Trotzdem sollte man sich davor hüten, allein aus der Erfolgswahrscheinlichkeit zu weitgehende Schlussfolgerungen zu ziehen.

14 BVerfGE 78, 104 (117 f.).

15 S. etwa BVerfG, Beschl. v. 12. Jan. 1993, 2 BvR 1584/92; Beschl. v. 2. Feb. 1993, FamRZ 1993, 664; Beschl. v. 20. Feb. 2002, NJW-RR 2002, 1069.

16 Zu Faktoren, die indizieren, wie ernst das BVerfG einen Fall nimmt, s. *Engel*, Efficiency (Fn. 1), 267 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

B. Die empirische Analyse der Gleichheitsrechtsprechung

In meiner vergleichenden Forschung zum verfassungsrechtlichen Gleichheitsschutz habe ich insgesamt vier verschiedene Modelle identifiziert, die Verfassungsgerichte in ihrer Gleichheitsrechtsprechung zugrunde legen – das Modell von Gleichheit als Diskriminierungsverbot, das Modell der Gleichheit als Konsistenzgebot, ein Modell der positiven Gleichheit und ein Modell, bei dem Gerichte, die Ausgestaltung von Gleichheit überwiegend der Einschätzungsprärogative des Gesetzgebers überlassen.[17]

Das vorherrschende Gleichheitsmodell in der internationalen Diskussion ist das Modell von Gleichheit als Diskriminierungsverbot.[18] Nach diesem Modell legen Gerichte einen strengeren Prüfungsmaßstab an, wenn eine Unterscheidung auf problematischen Merkmalen, wie Geschlecht, Hautfarbe, Religion, sexuelle Orientierung o. ä., beruht.[19] Bei Unterscheidungen, die nicht auf einem solchen problematischen Merkmal beruhen, wird allenfalls eine Willkürkontrolle mit geringer Kontrolldichte vorgenommen.

Betrachtet man die dogmatischen Formeln, die das BVerfG im Laufe der Zeit entwickelt hat, dann scheint das Modell der Gleichheit als Diskriminierungsverbot auf den ersten Blick auch die Gleichheitsrechtsprechung des Bundesverfassungsgerichts zu beschreiben. Diese Hypothese soll im ersten Abschnitt getestet werden. Die Untersuchung zeigt, dass ein anderes Gleichheitsmodell, das der Gleichheit als Konsistenzgebot, die Rechtsprechung des BVerfG besser beschreibt. Dieses Modell zeichnet sich dadurch aus, dass auch Unterscheidungen, die nicht auf einem problematischen Merkmal beruhen, grundsätzlich nach einem strikteren Prüfungsmaßstab kontrolliert werden, der über eine zurückgenommene Willkürkontrolle hinausgeht.[20]

Nach dieser generellen Einordnung der Gleichheitsrechtsprechung des Bundesverfassungsgerichts sollen noch zwei spezifische Aspekte untersucht werden. Zum einen analysiere ich den Einfluss des ehemaligen Verfassungsrichters Paul Kirchhof auf die Gleichheitsrechtsprechung des Gerichts. Kirchhof hat als Richter insbesondere in Bezug auf die Kon-

17 *Niels Petersen*, Mapping Equality Case Law Around the World, in: Equality's Guardians (Fn. 3), 1 (6 ff.).

18 Ebd., 6.

19 Ebd., 8.

20 Ebd., 7 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

trolle von Steuergesetzgebung neue dogmatische Entwicklungen angestoßen.[21] Da Steuergesetze überwiegend am Maßstab von Art. 3 Abs. 1 GG überprüft werden, bietet es sich an, diese Rolle auch empirisch näher zu analysieren. Der zweite Aspekt betrifft die Lenkungswirkung von dogmatischen Formeln. Wie in der Einleitung bereits angesprochen wurde, hat die Dogmatik des Gleichheitssatzes mehrere Entwicklungsstufen durchlaufen.[22] Daher soll untersucht werden, ob sich die unterschiedliche Ausprägung der dogmatischen Formeln auf die Erfolgswahrscheinlichkeit bei Gleichheitsfällen ausgewirkt hat.

I. Gleichheit als Konsistenzgebot

Betrachten wir zunächst die empirischen Daten für die Fälle, in denen das BVerfG die Verfassungsmäßigkeit eines Gesetzes am Maßstab eines Gleichheitssatzes überprüft hat. Wenn die Unterscheidung nicht auf einem problematischen Merkmal beruhte, lag die Erfolgswahrscheinlichkeit bei 25,07 % (N = 754). Wenn die Unterscheidung direkt oder indirekt auf einem problematischen Merkmal beruhte, lag sie etwas höher, bei 39,26 % (N = 135). Unterscheidet man bei den problematischen Merkmalen zwischen direkten und indirekten Diskriminierungen, beobachtet man – wenig überraschend – bei ersteren eine höhere Erfolgswahrscheinlichkeit. Die Erfolgswahrscheinlichkeit bei direkten Diskriminierungen beträgt 42,11 % (N = 92) und bei indirekten Diskriminierungen 32,5 % (N = 42). Dabei differieren die Erfolgsaussichten jedoch erheblich, je nach Differenzierungsmerkmal.

Wenn man sich die direkten Diskriminierungen ansieht, dann ist, wie Tabelle 1 zeigt, das Geschlecht das wichtigste Differenzierungsmerkmal. Hier waren 61,29 % aller Fälle erfolgreich. Die nicht erfolgreichen Fälle stammen aus der frühen Rechtsprechung des Gerichts. So sah dieses bspw. Unterscheidungen zwischen Männern und Frauen bei der Zuerkennung einer Witwenrente noch als zulässige Typisierung an.[23] In jüngerer Zeit hat das BVerfG Differenzierungen aufgrund des Geschlechts nur dann

21 *Olaf Köppe*, Bundesverfassungsgericht und Steuergesetzgebung, in: Robert Christian van Ooyen / Martin H.W. Möllers (Hg.), Handbuch Bundesverfassungsgericht im politischen System, 2015, 785 (787).

22 S. oben, Fn. 9 und begleitenden Text.

23 S. etwa BVerfGE 17, 1; 17, 62; 39, 169; 39, 196.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Tabelle 1. Erfolgswahrscheinlichkeit bei unmittelbar auf einem problematischen Merkmal beruhenden Unterscheidungen.

Unterscheidungskriterium	Anzahl der Fälle	Erfolgswahrscheinlichkeit
Geschlecht	31	61,29 %
Staatsangehörigkeit	11	36,36 %
Wohnsitz	13	15,38 %
Außereheliche Geburt	13	46,15 %
Behinderung	3	66,67 %
Alter	5	20 %
Sexuelle Orientierung	1	0 %
Ehestatus	15	40 %
Gesamt	92	42,11 %

akzeptiert, wenn durch die Regelungen Männer benachteiligt wurden.[24] Mit Ausnahme der Diskriminierung aufgrund einer Behinderung – einer Fallgruppe mit lediglich drei Fällen – haben alle anderen problematischen Merkmale eine Erfolgswahrscheinlichkeit von weniger als 50 %. Dies gilt für Differenzierungen aufgrund einer außerehelichen Geburt (46,15 %), aufgrund des Ehestatus (40 %), aufgrund der Staatsangehörigkeit (36,36 %) und des Wohnsitzes (15,38 %).

Tabelle 2. Erfolgswahrscheinlichkeit bei mittelbar auf einem problematischen Merkmal beruhenden Unterscheidungen.

Unterscheidungskriterium	Anzahl der Fälle	Erfolgswahrscheinlichkeit
Geschlecht	11	72,72 %
Staatsangehörigkeit	5	0 %
Religion	3	0 %
Außereheliche Geburt	6	0 %
Behinderung	6	33,33 %
Sexuelle Orientierung	10	30 %
Sozialer Status	1	0 %
Gesamt	42	32,5 %

24 BVerfGE 56, 363; 74, 163; 87, 1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Wie Tabelle 2 verdeutlicht, ist die Divergenz bei den indirekten Diskriminierungen noch stärker ausgeprägt. Auch hier dominiert das Geschlecht als wichtigstes Unterscheidungsmerkmal mit einer Erfolgswahrscheinlichkeit von 72,72 %. Eine hervorgehobene Rolle spielt auch das Merkmal der sexuellen Orientierung, selbst wenn dieses mit 30 % eine deutlich geringere Erfolgswahrscheinlichkeit hat. Hier lässt sich jedoch eine Entwicklung feststellen. Die meisten Fälle in dieser Kategorie betreffen die Unterscheidung zwischen Lebenspartnern und Eheleuten vor Einführung der Ehe für gleichgeschlechtliche Paare. In den frühen Entscheidungen sah das BVerfG in der Schlechterstellung von Lebenspartnern gegenüber Ehepartnern keine Diskriminierung.[25] Die Wende kam nach der *Maruko*-Entscheidung des EuGH, die in Differenzierungen zwischen Ehepartnern und Lebenspartnern eine Diskriminierung sah.[26] Nach dieser Entscheidung stufte das BVerfG Differenzierungen im Steuer- und Adoptionsrecht als Diskriminierung aufgrund der sexuellen Orientierung ein.[27]

Trotz der bedeutenden Rechtsprechungslinie zur Antidiskriminierung im Bereich des Geschlechts und der sexuellen Orientierung, ist die verfassungsgerichtliche Gleichheitsrechtsprechung allgemein jedoch dem Modell der Gleichheit als Konsistenzgebot zuzuordnen. Zwar sind die Erfolgsaussichten bei Diskriminierungen aufgrund eines problematischen Merkmals höher als bei Fehlen eines solchen Merkmals. Das ist allerdings nicht überraschend. Differenzierungen aufgrund eines problematischen Merkmals werden in den meisten Jurisdiktionen einem strengeren Prüfungsmaßstab unterworfen als Unterscheidungen, die nicht auf einem solchen Merkmal beruhen.

Der Hauptunterschied zu den Jurisdiktionen, die Gleichheit in erster Linie als Antidiskriminierungsrecht verstehen, ist das Anlegen eines strengeren Prüfungsmaßstabs für Fälle ohne problematisches Merkmal. Dies drückt sich in einer relativ hohen Erfolgswahrscheinlichkeit solcher Fälle sowie einer hohen absoluten Zahl an erfolgreichen Fällen aus.[28]

25 S. BVerfG, Beschl. v. 21. Mai 1999, NZA 1999, 878; Beschl. v. 28. Feb. 2005, NJW 2005, 1709; Beschl. v. 20. Sept. 2007, NJW 2008, 209; Beschl. v. 6. Mai 2008, NJW 2008, 2325.

26 EuGH, *Maruko*, C-267/06, Slg. 2008, I-757.

27 BVerfG, Beschl. v. 20. Sept. 2007, NJW 2008, 209; BVerfGE 126, 400; 133, 59.

28 *Niels Petersen*, The Diversity of Judicial Conceptions of Equality, in: Equality's Guardians (Fn. 3), 268 (272).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Zwar haben wir gesehen, dass die reine Erfolgswahrscheinlichkeit beim BVerfG aufgrund des dargestellten Selektionseffekts wahrscheinlich überschätzt wird. Aber auch in absoluten Zahlen gibt es häufig erfolgreiche Fälle, die kein problematisches Differenzierungsmerkmal aufweisen. Insgesamt gab es 189 erfolgreiche Fälle – deutlich mehr als erfolgreiche Antidiskriminierungsfälle. Die absolute Zahl der erfolgreichen Fälle hängt sicherlich auch mit der Art von Fällen zusammen, die vor Gericht gebracht werden. Dass diese Fälle aber überhaupt so weit kommen, zeigt, dass Prozessvertreter sich eine hinreichende Erfolgswahrscheinlichkeit ausrechnen.

Tabelle 3. Erfolgswahrscheinlichkeit nach Sachmaterie bei Fehlen eines problematischen
Merkmals.

Unterscheidungskriterium	Anzahl der Fälle	Erfolgswahrscheinlichkeit
Steuern & öffentl. Abgaben	156	30,12 %
Sozialleistungen	180	24,44 %
Familie	11	9,09 %
Beruf	79	21,52 %
Strafrecht	40	7,5 %
Migration	2	50 %
Politische Partizipation	48	45,83 %
Öffentliche Beschäftigung	79	30,38 %
Zivilprozess	24	20,83 %
Eigentum/Erbrecht	31	6,45 %
Universität	29	48,28 %
Sonstige	75	12 %
Gesamt	754	25,07 %

Wenn wir die Fälle ohne problematisches Merkmal einmal näher ansehen, können wir – wie Tabelle 3 zeigt – deutliche Unterschiede zwischen einzelnen Sachbereichen feststellen. Die größte Erfolgswahrscheinlichkeit haben Fälle in Bereichen der politischen Partizipation (45,83 %), im Bereich der Hochschulbildung (48,28 %) und im Beamtenwesen und der öffentlichen Beschäftigung (30,38 %). Blicken wir auf die absoluten Zahlen, sind das Steuerrecht und das Recht der öffentlichen Abgaben (47 erfolgreiche Fälle) und die Sozialleistungen (44 erfolgreiche Fälle) die wichtigsten Ka-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tegorien. In einigen dieser Gebiete hat das BVerfG gleichheitsrechtliche Sonderdogmatiken entwickelt, die die Bedeutung dieser Kategorien unterstreichen und wahrscheinlich zu der hohen Erfolgswahrscheinlichkeit beigetragen haben. Dies gilt im Bereich der politischen Partizipation etwa für die Prinzipien der Wahlrechtsgleichheit und der Chancengleichheit politischer Parteien[29] oder im Steuerrecht für die Grundsätze der steuerlichen Belastungsgleichheit[30] und der Folgerichtigkeit[31].

II. Gibt es einen Paul-Kirchhof-Effekt?

Ein Verfassungsrichter, der maßgeblichen Einfluss auf die Steuerrechtsprechung des BVerfG hatte, war Paul Kirchhof.[32] Angesichts der prominenten Stellung des Steuerrechts in der Gleichheitsrechtsprechung des BVerfG liegt es nahe, den Einfluss Kirchhofs auf diese Rechtsprechung näher zu untersuchen. Während Kirchhofs Amtszeit gab es insgesamt 10 Entscheidungen des zweiten Senats, die sich mit Steuergesetzen befassten und auf Grundlage von Art. 3 Abs. 1 GG entschieden wurden. Sechs dieser Entscheidungen waren erfolgreich, was eine Erfolgsquote von 60 % ergibt. Das ist fast doppelt so hoch wie die allgemeine Erfolgswahrscheinlichkeit im Steuerrecht von 30,12 %. Ein Chi-square-Test, der die Steuerrechtsfälle in der Kirchhof-Ära mit den Steuerrechtsfällen, an denen Kirchhof nicht beteiligt war, vergleicht, produziert einen p-Wert

29 S. dazu *Klaus Kröger*, Schematische Parteiengleichheit als Grundbedingung der Demokratie, in: FS Wilhelm G. Grewe, 1981, 507; *Heinz-Werner Kleffmann*, Chancengleichheit im Wahlkampf, NVwZ 1983, 532; *Hans Herbert von Arnim*, Der strenge und der formale Gleichheitssatz, DöV 1984, 85; *Uwe Volkmann*, Politische Parteien und öffentliche Leistungen, 1993, S. 189 ff.; *Niels Petersen*, Verfassungsgerichte als Wettbewerbshüter des politischen Prozesses, in: Dominik Elser u. a. (Hg.), Das letzte Wort, 2014, 59 (67 ff.); *Emanuel Towfigh*, Das Parteienparadox, 2015, S. 63 ff.

30 S. *Dieter Birk*, Das Leistungsfähigkeitsprinzip als Maßstab der Steuernormen, 1983; *Hans-Wolfgang Arndt*, Gleichheit im Steuerrecht, NVwZ (1988), 787.

31 S. dazu *Joachim Englisch*, Folgerichtiges Steuerrecht als Verfassungsgebot, in: FS Joachim Lang, 2010, 167; *Philipp Dann*, Verfassungsgerichtliche Kontrolle gesetzgeberischer Rationalität, Der Staat 49 (2010), 630; *Mehrdad Payandeh*, Das Gebot der Folgerichtigkeit: Rationalitätsgewinn oder Irrweg der Grundrechtsdogmatik?, AöR 136 (2011), 578; *Lerke Osterloh*, Folgerichtigkeit: Verfassungsgerichtliche Rationalitätsanforderungen in der Demokratie, in: FS Brun-Otto Bryde, 2013, 429; *Niels Petersen*, Gesetzgeberische Inkonsistenz als Beweiszeichen, AöR 138 (2013), 108; *Laurence O'Hara*, Konsistenz und Konsens, 2018.

32 *Köppe* (Fn. 21), 787.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

von 0,033. Folglich ist der Unterschied in der Erfolgswahrscheinlichkeit zwischen den Fällen, an denen Kirchhof unmittelbar beteiligt war und jenen, an denen er nicht beteiligt war, statistisch signifikant.

Der Einfluss Kirchhofs auf die Steuerrechtsprechung des Verfassungsgerichts wirkte auch nach seinem Ausscheiden noch nach. Vor Kirchhof lag die Erfolgswahrscheinlichkeit für Steuerrechtsfälle, die auf den Gleichheitsgrundsatz gestützt wurden, bei 18,84 %. Nach Kirchhofs Ausscheiden aus dem Gericht war die Erfolgswahrscheinlichkeit zwar nicht mehr so hoch wie während seines Wirkens. Sie war aber mit 36,36 % statistisch signifikant höher als vor seinem Eintritt ins Gericht ($p = 0{,}019$). Das legt nahe, dass Kirchhof die Rechtsprechung des zweiten Senats in Steuerrechtssachen entscheidend geformt hat.

Tabelle 4. Erfolgswahrscheinlichkeit zweiter Senat nach Jahrzehnt.

Dekade	Anzahl der Fälle	Erfolgswahrscheinlichkeit
1950	10	20 %
1960	38	34,21 %
1970	52	28,85 %
1980	38	28,95 %
1990	23	47,83 %
2000	47	21,28 %
2010	27	33,33 %

Demgegenüber ist der Einfluss auf die allgemeine Gleichheitsrechtsprechung deutlich geringer. Tabelle 4 zeigt die Entwicklung der Erfolgswahrscheinlichkeit von Fällen der Gesetzeskontrolle, in denen die Unterscheidung nicht auf einem problematischen Merkmal beruhte, über die verschiedenen Jahrzehnte hinweg. Man sieht, dass die Erfolgswahrscheinlichkeit in den 1990er Jahren, also der Dekade, in die der Großteil von Kirchhofs Wirken fällt, höher ist als in den anderen Dekaden. Ein Chi-square-Test bestätigt diese Intuition und stellt einen statistisch schwach signifikanten Unterschied fest ($p = 0{,}053$). Demgegenüber war Kirchhofs Einfluss auf die breitere Gleichheitsrechtsprechung des zweiten Senats nicht von Dauer. Die Periode vor 1990 unterscheidet sich nicht statistisch signifikant von der post-Kirchhof-Periode ab 2000 ($p = 0{,}534$).

Es lässt sich also festhalten: Die Bedeutung, die Paul Kirchhof als Richterpersönlichkeit für seinen Einfluss auf die Steuerrechtsrechtspre-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

chung des BVerfG nachgesagt wird, lässt sich auch statistisch bestätigen. Die Erfolgswahrscheinlichkeit bei der Überprüfung steuerrechtlicher Gesetze ist während Kirchhofs Amtszeit ungewöhnlich hoch. Das Wirken Kirchhofs hat dabei eine langfristige Wirkung auf die Steuerrechtsprechung des Gerichts auch nach seinem Ausscheiden. Demgegenüber ist der Einfluss auf die allgemeine Gleichheitsrechtsprechung deutlich geringer. Zwar lässt sich eine leicht erhöhte Erfolgswahrscheinlichkeit für die Hauptdekade seines Wirkens feststellen. Dieser Einfluss wirkt aber nach Kirchhofs Ausscheiden aus dem Gericht nicht nach.

III. Wie groß ist die Lenkungswirkung von dogmatischen Formeln?

Eine letzte Beobachtung betrifft die Lenkungswirkung von dogmatischen Formeln. Wir haben in der Einleitung gesehen, dass sich die Dogmatik des allgemeinen Gleichheitsgebotes seit dem Zweiten Weltkrieg erheblich entwickelt und die Rechtsprechung des BVerfG hier zur Konkretisierung des Gleichheitssatzes unterschiedliche dogmatische Formeln zugrunde gelegt hat. Spiegelt sich diese Entwicklung auch in der Erfolgswahrscheinlichkeit der unter dem allgemeinen Gleichheitssatz entschiedenen Gleichheitsfälle wider, die kein problematisches Merkmal beinhalten?

Tabelle 5 zeichnet die Erfolgswahrscheinlichkeiten unter den jeweiligen dogmatischen Formeln nach. In der Tabelle sind dabei alle Gleichheitsfälle ohne problematisches Merkmal in der jeweiligen Periode festgehalten, ohne dass im Einzelnen überprüft wurde, welche dogmatische Formel konkret verwendet wurde. Vielmehr wurden zur Abgrenzung der Perioden die Fälle herangezogen, die die Zäsur zwischen den Formeln markieren.[33] Dabei lassen sich jeweils nur geringe Unterschiede feststel-

Tabelle 5. Erfolgswahrscheinlichkeit nach dogmatischer Formel.

Dekade	Anzahl der Fälle	Erfolgswahrscheinlichkeit
Willkürformel	320	23,75 %
Neue Formel	161	26,71 %
Neueste Formel	201	24,37 %
Stufenlosformel	76	27,6 %

33 Dies sind BVerfGE 55, 72 (Willkürformel vs. neue Formel); BVerfGE 88, 87 (neue Formel vs. neueste Formel); BVerfGE 129, 49 (neueste Formel vs. Stufenlosformel).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

len. Zwar gibt es zwischen den einzelnen Zeiträumen geringe Variationen, diese sind aber statistisch nicht signifikant. So ergibt ein Chi-square-Test, dass zwischen der Periode der Willkürformel und der Periode der Stufenlosformel kein signifikanter Unterschied besteht ($p = 0{,}479$). Das legt nahe, dass sich zwar die dogmatischen Formeln jeweils weiterentwickelt haben, der Unterschied für die Rechtspraxis aber äußerst gering geblieben ist.[34]

Wenn man sich die Entwicklung der Dogmatik im Detail ansieht, ist dies weniger überraschend, als es auf den ersten Blick erscheinen mag. So hat das BVerfG unter der Willkürkontrolle oft strengere Maßstäbe angewandt.[35] So findet man etwa vielfach implizite Verhältnismäßigkeitserwägungen in den Begründungen.[36] Insofern ist die neue Formel wohl nur ein Ausdruck dessen, was das Gericht vorher bereits praktiziert hat.[37] Auch die Zäsur durch die Stufenlosformel ist geringer, als die Formulierung der Formel *prima facie* nahelegt. Vielmehr hat das BVerfG auch nach Einführung der Stufenlosformel oft schematisierte Stufen angewandt und zwischen einem „großzügigeren" und einem „strengeren" Prüfungsmaßstab unterschieden.[38] Insofern betreffen die verschiedenen Formeln zur Konkretisierung des Gleichheitssatzes wohl eher die dogmatische Verpackung als tatsächlich den Inhalt der Gleichheitsrechtsprechung.

34 Diese Hypothese steht unter der Annahme, dass das Verhältnis von beobachteten und unbeobachteten Fällen ungefähr gleichgeblieben ist. Die Zahl der unveröffentlichten Fälle des Bundesverfassungsgerichts hat über die Jahre zugenommen. Wie sich das auf den hier beobachteten Ausschnitt auswirkt, ist schwer zu sagen. Geht man davon aus, dass die Zahl der unbeobachteten Fälle im Verhältnis zu den beobachteten Fällen über die Jahre ebenfalls zugenommen hat, hätte das zur Folge, dass die Erfolgswahrscheinlichkeit von der Willkürformel zur Stufenlosformel graduell abgenommen hätte.

35 S. etwa *Reiner Schmidt*, Die Bindung des Gesetzgebers an den Gleichheitssatz (Art. 3 Abs. 1 GG) in der Rechtsprechung des Bundesverfassungsgerichts, 1963, S. 110 f.; *Alexander Somek*, The Deadweight of Formulae, UPenn J. Const. L. 1 (1998), 284 (300); *Kempny/Reimer* (Fn. 9), 107.

36 *Niels Petersen*, Gleichheitssatz und Einzelfallgerechtigkeit, Der Staat 57 (2018), 327 (340 f.).

37 Ebd., 343.

38 Ebd., 345 m.w.N.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. Schlussfolgerungen

Eine empirische Analyse der Gleichheitsrechtsprechung des Bundesverfassungsgerichts liefert schon auf beschreibender Ebene einige interessante Erkenntnisse. Zunächst haben wir gesehen, dass das BVerfG Gleichheit in erster Linie als Konsistenzgebot versteht. Zwar gibt es eine signifikante Antidiskriminierungsrechtsprechung, insbesondere in Bezug auf Differenzierungen aufgrund des Geschlechts und der sexuellen Orientierung. In absoluten Zahlen dominieren jedoch die erfolgreichen Fälle, in denen die Unterscheidung nicht auf einem problematischen Merkmal beruhte.

Weiterhin zeigt die Analyse, dass sich der Einfluss von Paul Kirchhof auf die Steuerrechtsprechung des Gerichts auch statistisch nachweisen lässt. Nach seiner Amtszeit ist die Erfolgswahrscheinlichkeit von Steuerrechtsfällen, die am Maßstab des Gleichheitssatzes entschieden werden, deutlich höher als vor seiner Amtszeit. Dieser Einfluss zeigt sich aber nur im Steuerrecht, nicht allgemein für den Gleichheitssatz. Schließlich haben wir gesehen, dass das dominante Narrativ der Dogmatik des Gleichheitssatzes als einer Dogmatik mit mehreren Entwicklungsstufen nur bedingt tragfähig ist. Zwar haben sich die dogmatischen Formeln über die Zeit verändert – Änderungen in der Erfolgswahrscheinlichkeit von Gleichheitsfällen ergaben sich jedoch nicht. Das legt nahe, dass die Prüfungsmaßstäbe trotz Änderung der dogmatischen Formeln weitgehend konstant geblieben sind. Der Wandel bezog sich also lediglich auf die dogmatische Verpackung, nicht auf den Inhalt des Gleichheitssatzes.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die „gewagte Formel“*
Plädoyer für eine statistisch informierte Gleichheitsdogmatik

*Emanuel V. Towfigh***

Die Dogmatik der Gleichheitsrechte ist Ursache für vielfältige Unzufriedenheit, die in der rechtswissenschaftlichen Literatur immer wieder artikuliert wird.[1] *Dies gilt für Art. 3 GG ebenso wie für die – ähnlich konstruierten – EU-Rechtssätze (Art. 20 und Art. 21 EU-GRCh) oder die (bisweilen begriff-*

* Im Gleichheitsrecht gab es zunächst die „Willkürformel“, die von der „neuen Formel“ und später von der „neuen Formel im neuen Gewand“ bzw. der „neuesten Formel“ abgelöst wurde. „Formel“ passt für den hier vorgeschlagenen, neuen dogmatischen Zugang weiterhin – vielleicht besonders gut, weil es tatsächlich auch um „Formeln“ im alltagssprachlichen Sinn des Begriffs geht. „Neu“ ist, auch mit Blick auf die Okkupation durch die *Neue Verwaltungsrechtswissenschaft* und die *NJW* keine Option. „Kritisch“ weckt Assoziationen zu den „*Critical Legal Studies*“ und zur *KJ*, die hier nicht passen. „Vermittelnd“ ist die Perspektive nicht, allenfalls „einheitlich“. „Gewagt“ erscheint sowohl mit Blick auf den Jubilar als auch angesichts des interdisziplinären Charakters des Vorschlags passend. Außerdem dürfte die Wortwahl erhöhte Aufmerksamkeit auf sich ziehen – ein Kollateralnutzen.

** Der Autor ist Inhaber des Lehrstuhls für Öffentliches Recht, Empirische Rechtsforschung und Rechtsökonomik an der EBS Universität für Wirtschaft und Recht, Distinguished Scholar in Residence an der Peking University School of Transnational Law sowie Max Planck Fellow und Leiter des Center for Diversity in Law am Max-Planck-Institut für ausländisches öffentliches Recht und Völkerrecht. Für wertvolle Anregungen zu diesem Beitrag dankt er Katharina Towfigh, Anna Katharina Mangold, Karsten Schneider, Iris Canor, Helena Kaim, Benedikt Surmann, Johannes Felsner und Julia Schönemann; für die Unterstützung bei der Recherche dankt er Jan Rosen.
Von 2007–2016 war er Senior Research Fellow in der Abteilung von Christoph Engel am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern und von 2016–2025 Research Affiliate ebenda. In die Zeit am Institut fallen neben der Erarbeitung der Habilitationsschrift prägende Auslandsaufenthalte; im Rahmen des Habilitationsverfahrens an der Universität Münster erstattete Herr Engel ein Amicus-Gutachten. Der vorliegende Text soll die tiefe Dankbarkeit des Autors gegenüber dem Jubilar ausdrücken – für die vielfältige, großartige Unterstützung und für die Möglichkeit, in einem interdisziplinären Umfeld mit herausragenden Wissenschaftler:innen, einer grandiosen Verwaltung und einer innovativen Bibliothek arbeiten und dabei mit vielen feinen Menschen enge Freundschaften knüpfen zu dürfen.

1 *F. Wollenschläger,* in: Huber/Voßkuhle (Hrsg.) GG, Art 3 Rn. 47.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lich anders geschnittenen, funktional aber ähnlich wirkenden) Gleichheitsgewährleistungen ausländischer Rechtsordnungen.[2] *Eine Möglichkeit, das Problem zu beschreiben, ist, dass der von einem formalen Gleichheitsverständnis ausgehende Allgemeine Gleichheitssatz (Art. 3 Abs. 1 GG) unterspezifiziert (dazu I.1) ist, da er keine Maßstäbe für die Wahl eines angemessenen, rechtlich relevanten Differenzierungsmerkmals bereithält, während der vom materialen Gleichheitsverständnis geprägte Besondere Gleichheitssatz (Art. 3 Abs. 2 und 3 GG) überspezifiziert (I.2) ist, da er bestimmte differentia specifica exklusiv normativ fixiert, essenzialisiert und damit eine kategoriale Engführung riskiert. In beiden Fällen drängt sich Kategorienkritik auf. Ausgehend von einer Intuition (II.1) entwickelt der Beitrag einen interdisziplinären Zugriff auf die Gleichheitsrechte und zeigt, wie sich mit Hilfe einer statistisch informierten Heuristik Art. 3 Abs. 1 und Abs. 3 GG re-interpretieren lassen (II.2), bevor er beispielhaft die Formalisierung dieser Heuristik skizziert (II.3). Er kann als Projektskizze und Auftakt eines umfangreicheren Forschungsprogramms verstanden werden.*

A. Problem

Aus der Gleichheitsgarantie des Art. 3 GG ergibt sich die Verpflichtung aller staatlicher Stellen, gleiche Fälle gleich (und ungleiche ungleich) zu behandeln.[3] Allein: Verlässliche *allgemeine* Maßstäbe, was rechtlich als „gleich" zu gelten hat und wann Ungleichbehandlungen folglich zu rechtfertigen sind, liefern weder der Text des Grundgesetzes noch die zu diesem entwickelte, überkommene Dogmatik.[4] Gesetzgebung, Verwaltung und Gerichte sind bei der Operationalisierung des Art. 3 GG auf sich gestellt. Die Maßstäbe der *besonderen* Gleichheitsgewährleistungen beziehen sich auf spezifische Individual- oder Gruppen-Merkmale, nach denen grundsätzlich nicht differenziert werden darf, die allerdings Differenzwahrnehmungen befördern und Identitätsfragen aufrufen. Schließlich klaffen zwischen allgemeinen und besonderen Gleichheitssätzen Lücken, es bleiben

2 *N. Petersen*, Equality's Guardians, S. 205 ff.

3 *U. Sacksofsky*, in: Herdegen, et al. (Hrsg.), HdB VerfR, § 19, Rn. 42; *S. Boysen*, in: v.Münch/Kunig (Hrsg.) GG, Art. 3, Rn. 62 ff.

4 *Boysen,* (Fn. 3), Rn. 208; *A. Nußberger*, in: Sachs (Hrsg.) GG, Art. 3, Rn. 4; die Notwendigkeit eines Maßstabs in Abrede stellen *S. Kempny/P. Reimer*, Die Gleichheitssätze, S. 44 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Ungleichbehandlungen, die jedenfalls von der überkommenen Dogmatik nur unbefriedigend verarbeitet werden.

I. Allgemeiner Gleichheitssatz und formales Gleichheitsverständnis

Der Allgemeine Gleichheitssatz, Art. 3 Abs. 1 GG, basiert auf einem abstrakt-generellen oder formalen Verständnis von Gleichheit und zielt auf den Schutz vor ungerechtfertigter oder gar willkürlicher Ungleichbehandlung durch den Staat: Menschen in wesentlich gleichen Lebenssituationen dürfen nicht unterschiedlich behandelt werden, sofern hierfür nicht ein sachlich gerechtfertigter Differenzierungsgrund vorliegt.[5] Ein Vergleich im Sinne des allgemeinen Gleichheitssatzes des Art. 3 Abs. 1 GG setzt voraus, dass die herangezogenen Gruppen in rechtlich relevanter Hinsicht „wesentlich gleich“ sind, also unter einen gemeinsamen Oberbegriff, ein *genus proximum,* gefasst werden können.

Bei dieser Gruppenbildung und -betrachtung verbleiben allerdings erhebliche Auslegungsspielräume, die gleichheitsrechtliche Fallkonstellationen häufig zum Gegenstand kontroverser Diskussionen und die Unzulänglichkeiten der geltenden Dogmatik in der Rechtsprechungspraxis sichtbar machen.[6] Ob sich die Öffnungszeiten einer Bahnhofsapotheke nach den Regeln für Bahnhofsgeschäfte oder für Apotheken richten,[7] ob (und wann) Beamt:innen, Angestellte und Arbeiter:innen unterschiedlich behandelt werden dürfen,[8] ob Altersgrenzen für Vertragsärzt:innen verfassungsmäßig sind[9] – all das hängt davon ab, welche *differentia specifica* (also Differenzierungskriterien) gewählt werden, eine Wahl, die regelmäßig das Ergebnis determiniert. Dabei lassen sich – wie auch die Auseinandersetzung des Bundesverfassungsgerichts mit konkurrierenden Vergleichsrahmen zeigt, in der Regel beliebig viele Vergleichsgruppen bilden.[10] Sinnvoll im Sinne einer verfassungsrechtlichen Prüfung ist dabei jede plausible *differentia specifica*, die zur Verfassungswidrigkeit führt.

5 Vgl. ausführlich *A. K. Mangold*, Demokratische Inklusion durch Recht, S. 187 ff. m.w.N.
6 *U. Kischel,* in: Epping/Hillgruber (Hrsg.), BeckOK GG, Art. 3, Rn. 17.
7 Vgl. BVerfGE 13, 225.
8 Vgl. BVerfGE 82, 126.
9 Vgl. BVerfG NJW, 1998, 1776.
10 Vgl. BVerfGE 155, 1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Dennoch bleibt die konkrete Konstruktion und Wahl der maßgeblichen *differentia specifica* eine weitgehend intuitive Entscheidung. Der Allgemeine Gleichheitssatz wird daher auch als Grundrecht ohne fest konturierten „Schutzbereich" bezeichnet, der Schutzbereich entsteht erst in der konkreten Rechtsanwendung etwa durch die Verwaltung oder im Rahmen der gerichtlichen Überprüfung staatlicher Entscheidungen.[11] Und obwohl die Wahl der *differentia specifica* für das Ergebnis so entscheidend ist, gibt es lediglich einen Plausibilitätsmaßstab: Leuchtet die Einordnung des Sachverhalts in die gebildeten Kategorien ein?

Eine Leerstelle des formalen Gleichheitsverständnisses findet sich ferner im Umgang mit faktischen sozialen, ökonomischen oder historisch bedingten Benachteiligungen, die nicht erfasst werden (können). Diskriminierung ist häufig kein offener, unmittelbarer oder bewusster Vorgang, sondern resultiert aus indirekter oder struktureller Benachteiligung – etwa durch ungleiche Bildungsvoraussetzungen, kulturell codierte Auswahlmechanismen oder systemisch verfestigte Exklusion.[12] Solche strukturellen Ungleichheiten bleiben, sofern sie nicht von einer besonderen Gleichheitsgewährleistung erfasst werden, rechtlich unerheblich und werden dadurch in der juristischen Praxis zumindest perpetuiert. Der formale Gleichheitsbegriff erfasst Kontexte nicht und beschränkt sich deshalb auf eine Gleichbehandlung, die Ungleichheiten oft nicht nur nicht beseitigt, sondern verfestigt und stößt damit gerade im Kontext der Diskriminierungstatbestände an harte Grenzen.[13]

II. Besonderer Gleichheitssatz und materiales Gleichheitsverständnis

Die Besonderen Gleichheitssätze des Art. 3 Abs. 2 und 3 GG verfügen einerseits ein Verbot, Benachteiligungen (negative Ungleichbehandlungen) an in der Menschheitsgeschichte und insbesondere im nationalsozialistischen Terrorregime missbrauchte Kriterien zu knüpfen. Umge-

11 *Wollenschläger,* (Fn. 1), Rn. 51.

12 *N. Markard,* Struktureller und institutioneller Rassismus, in: Froese/Thym (Hrsg.) Grundgesetz und Rassismus, S. 161 ff.; *S. Magen*, § 3 Verhaltenswissenschaftliche Aspekte des Antidiskriminierungsrechts, in: Mangold/Payandeh (Hrsg.), Handbuch Antidiskriminierungsrecht, Rn. 22 f; *I. Spiecker gen. Döhmann,* Kontexte der Demokratie: Parteien, Medien und Sozialstrukturen, VVDStRL 77 (2018), S. 9, 31.

13 *Mangold,* (Fn. 5), S. 8; *Sacksofsky* (Fn. 3), Rn. 14.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

kehrt lassen sich Privilegierungen (positive Ungleichbehandlungen, *affirmative action*) und differenzierende Maßnahmen wie Quotenregelungen, Förderprogramme oder Nachteilsausgleiche (Art. 3 Abs. 2 S. 2) legitimieren, die das Ziel verfolgen, faktische Gleichstellung zu erreichen.[14] Gleichheit und Ungleichheit werden an bestimmten, rechtlich fixierten Merkmalen festgemacht werden, gleichsam normativ gesetzte *differentia specifica*. Dafür kann gedanklich etwa an die *Vulnerabilität* von Individuen oder Gruppen mit spezifischen Merkmalen (z. B. phänotypische Zuschreibungen, Religion) oder die *Unverfügbarkeit* bestimmter persönlicher Merkmale (z. B. Alter, Geschlecht) angeknüpft werden.[15] Die besonderen Gleichheitssätze ergänzen so das Allgemeine Gleichheitsgebot, indem sie – einem materialen Gleichheitsverständnis folgend – gesellschaftliche Realitäten wahrnehmen, anerkennen und ausdrücklich in die rechtliche Betrachtung einbeziehen. Durch die Berücksichtigung auf struktureller und historischer Benachteiligung fußender unterschiedlicher Ausgangsbedingungen zielen sie darauf ab, tatsächliche Gleichheit zu schaffen.[16]

Allerdings hat auch das materiale Gleichheitsverständnis Schwächen. Zwar kann mit diesem Ansatz nicht (nur) Verhaltensunrecht, sondern auch Ergebnisunrecht verarbeitet werden, es macht eine – auch dem formalen Gleichheitsverständnis inhärente – grundlegende Ambivalenz sichtbar: Um Ungleichheit bekämpfen zu können, muss zunächst Unterschiedlichkeit festgestellt werden – etwa über „Gruppenzugehörigkeiten" wie Geschlecht, „Rasse", Herkunft oder Behinderung. Es müssen gerade die *differentia specifica* herausgestellt werden, an denen die ungleiche Behandlung ansetzt.[17] Wenn ein *Gender Pay Gap* nachgewiesen werden soll, müssen Gehaltszahlungen nach Geschlecht unterschieden werden. Damit läuft insbesondere eine auf ein materiales Gleichheitsverständnis ausgerichtete Dogmatik Gefahr, genau jene Kategorisierungen zu verfestigen,

14 *Boysen,* (Fn. 3), Rn. 134 ff.

15 *Boysen,* (Fn. 3), Rn. 115 ff., 125 ff.; BVerfGE 147, 1, 27; BVerfGE 102, 41, 46; BVerfGE 96, 288, 303.

16 *S. Grundmann*, Von formaler zu materialer Gleichheit, in: Grundmann / Thiessen (Hrsg.), S. 1, 15.

17 *E. V. Towfigh*, § 18 Der Umgang mit Empirie beim Nachweis von Diskriminierung, in: Mangold / Payandeh (Hrsg.), Handbuch Antidiskriminierungsrecht, Rn. 13.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

die eigentlich überwunden werden sollen. Dieses *Dilemma der Differenz* lässt sich mit einem materialen Gleichheitsverständnis nicht auflösen.[18]

Ein Weiteres kommt hinzu: Durch die Betonung des beanstandeten Merkmals kann es zu einer *Essentialisierung* kommen, das heißt der Blick auf eine Person oder Gruppe wird verengt, reduziert auf dieses *eine* Merkmal, das im Vergleich zu den vielen anderen – ebenfalls identitätsprägenden – Merkmalen überbetont wird.[19] Der Begriff der *Essentialisierung* lässt sich dabei an Aristoteles und dessen Essentialismus rückbinden, wonach sich das Wesen (*essentia*) jedes Dings durch die Kombination von *genus proximum* und *differentia specifica* bestimmen lässt.[20] Dasselbe begriffliche Muster liegt auch dem heutigen Gebrauch des Essentialismus-Begriffs, etwa im Antidiskriminierungsrecht, zugrunde: So wird kritisiert, dass bei Differenzierungen zwischen Individuen oder Gruppen (sei es durch diskriminierende Maßnahmen oder bei der Bekämpfung von Diskriminierung) oftmals ein einzelnes Merkmal so hervorgehoben werde, dass der Eindruck entstehen kann, dieses Merkmal beschriebe, prägte oder definierte das „Wesen" einer Person oder Gruppe. Diese Essentialisierung kann schnell hitzig debattierte und intensiv erlebte Identitätsfragen aufrufen, die am Innersten des Menschen rühren, und damit *existenzielle* Bedeutung erlangen. Polarisierung und gesellschaftliche Konflikte sind dann vorprogrammiert.

Eine weitere Gefahr der Essentialisierung liegt im *doppelten Andorra-Effekt*[21]: Menschen passen ihr Verhalten und ihre Persönlichkeit an die Erwartungen und Beurteilungen ihrer Umgebung an, die ihnen – aufgrund ihrer anhand des essentialisierenden Merkmals (z. B. sozial oder ethnisch) konstruierten Gruppenzugehörigkeit – zugeschrieben werden. Hinzu kommt, dass angeblich gruppentypisches Verhalten aufgrund entsprechender Erwartungen häufiger beobachtet wird, weil der Fokus eben auf solchem Verhalten liegt und jede entsprechende Beobachtung stereotype Erwartungen und damit die zugeschriebene Gruppenzugehörigkeit stabilisiert. Das verschärft gesellschaftliche Konflikte, weil die konstru-

18 *E. V. Towfigh*, „Dilemma der Differenz", in: Thym (Hrsg.), Deutschland als Einwanderungsland, S. 231, in: *D. Thym* (Hrsg.), Deutschland als Einwanderungsland„ S. 231, 235; *Towfigh,* (Fn. 17), Rn. 15.

19 *A. Phillips,* What's wrong with essentialism? Distinktion 11 (2010). S. 47, 55.

20 *Aristoteles*, Topiká VI, 6 (145a–145b).

21 benannt nach Max Frischs Drama *Andorra* (1961)*; R. K. Merton,* Die self-fulfilling prophecy, in: Meja/Stehr, Soziologische Theorie und soziale Struktur, S. 399.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ierte Realität zu einer selbsterfüllenden Prophezeiung führt und zum Anspruch, diese stereotypische Identität zu achten.

Mit Blick auf das materiale Gleichheitsverständnis ist daher festzuhalten, dass es *überspezifiziert* ist: Es verengt den Blick auf *ein* vermeintlich identitätsprägendes Merkmal und perpetuiert oder verstärkt so die Wahrnehmung und Bedeutung gerade jenes Merkmals, das Anknüpfungspunkt für Diskriminierung ist, die durch das materiale Gleichheitsverständnis eigentlich überwunden werden sollen.

B. Lösung

Aus zwei intuitiven Beobachtungen lässt sich für eine einheitliche Lösung der geschilderten Schwierigkeiten bei der Bewertung von Gleichheitsfragen durch Gesetzgebung, Verwaltung und Gerichte eine Heuristik ableiten, die sich wiederum in einem nächsten Schritt mit Blick auf spezielle Anwendungen präzisieren und formalisieren lässt.

I. Intuition

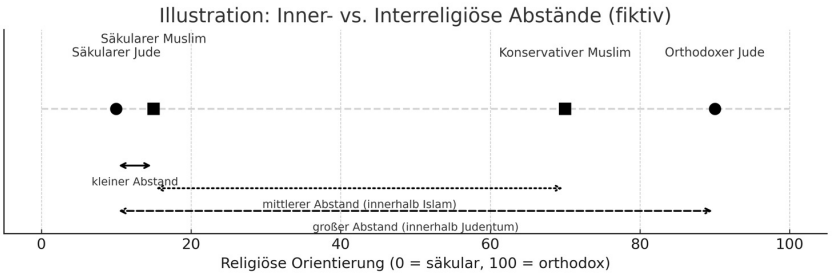
Den intuitiven Weg zu der hier vorgeschlagenen Problemlösung weisen zwei Beobachtungen.

1. Gruppenvarianzen

Die erste Beobachtung betrifft typische Gruppenkonstruktionen. Menschen lassen sich beispielsweise nach ihrer Religion in Gruppen einteilen – Zoroastrier, Juden, Christen, Muslime, Bahai usw.. Wenn man innerhalb der Bekenntnisse nach orthodoxer, konservativer, liberaler und säkularer Anhängerschaft differenziert, bekommt man den Eindruck, dass sich allerdings Glaube, Werte und Praxis der Gläubigen *innerhalb* einer Gemeinschaft stärker unterscheiden als *zwischen* Gemeinschaften: Ein „säkularer" Jude dürfte einer „säkularen" Muslima in vielerlei Hinsicht – etwa bei der Bewertung moralischer Fragen – „ähnlicher" sein als ein „säkularer" einem in einem weit verstandenen Sinn „konservativen" Muslim oder als eine „säkulare" Jüdin einer „orthodoxen" Jüdin. In einer Religionsgemeinschaft treffen, durch ein gemeinsames Bekenntnis lose verbunden, Menschen unterschiedlichster Lebensanschauungen, Lebens-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

wirklichkeiten und Lebenserfahrungen aufeinander.[22] Das kann als Form der Intersektionalität übrigens *vice versa* auch für Kategorisierungen wie „säkular" und „orthodox" gelten.



Ebenso haben die Lebenswirklichkeiten eines griechischen Kleinbauern und eines kleinen deutschen Landwirts möglicherweise mehr gemein als die eines Bauers aus Thessalien und eines Bankers aus Athen, auch wenn die Letztgenannten beide „Griechen" sind.

Und mit Blick auf das menschliche Genom zeigen populationsgenetische Befunde seit den 1970er Jahren, dass die genetische Varianz innerhalb einer „Ethnie" (oder einer lokal definierten Gruppe) deutlich größer ist als jene zwischen grob konstruierten „Rassen".[23] Eine Klassifizierung nach „Rasse" habe „virtually no genetic or taxonomic significance": „Our perception of large differences between races and subgroups, compared to variation within these groups, is a biased perception."[24]

Kurzum: Die Beispiele zeigen – untechnisch formuliert – dass bei der Bildung und dem Vergleich von Gruppen anhand eines Differenzie-

22 *E. V. Towfigh/R. P. Herberg*, Kann im Lebensrecht Religion rechtlich wirksam werden?, ZfL 2017, S. 2, 3; s. auch *C. Engel et al.*, Religion and Tradition in Conflict, Preprint of the Max Planck Institute for Research on Collective Goods, 2021/13.

23 *R. C. Lewontin*, The Apportionment of Human Diversity, in: Dobzhansky/Hecht/Steere (Hrsg.), Evolutionary Biology: Vol. 6, S. 381, 397; *D. J. Witherspoon, et al.*, Genetic Similarities Within and Between Human Populations, Genetics 2007, S. 351; *U. Kattmann*, Reflections on „race" in science and society in Germany, J. Anthropol. Sci. 95 (2017), S. 1; *M. Silcocks et al.,* Indigenous Australian genomes show deep structure and rich novel variations, Nature vol. 624, 7992 (2023), S. 593; s. auch *A. W. F. Edwards*, Human genetic diversity: Lewontin's fallacy, BioEssays 2003, S. 798.

24 *Lewontin,* (Fn. 23), S. 397.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

rungskriteriums (*differentia specifica*) der Verteilung nicht nur *zwischen*, sondern auch *innerhalb* von Gruppen Beachtung geschenkt werden muss, und dass sich nur solche Kategorisierungen als tragfähig erweisen, bei denen die Heterogenität innerhalb einer Kategorie geringer ist als jene zwischen Kategorien.

2. Methodische Lücke

Diese erste Beobachtung und die geschilderten Beispiele führen zur zweiten Beobachtung, die das methodische Werkzeug – die Rechtsdogmatik – betrifft, mit der staatliche Stellen Gleichheitsprobleme lösen sollen: Rechtsdogmatik, und in ihrem Kern Hermeneutik und juristischer Syllogismus, stellen kein geeignetes Instrumentarium für die notwendigen Vergleiche bereit. Hermeneutik hilft bei der Sinnerfassung eines Textes und erlaubt es, aus dem Normtext des Art. 3 GG den *normativen* Maßstab für Gleichheit abzuleiten; und der Syllogismus leitet die Rechtsanwender:innen bei der Subsumtion des *tatsächlichen* Sachverhalts unter den normativen Maßstab an. Für den Zwischenschritt zwischen Normermittlung und Subsumtion aber – nämlich bei der Aufgabe, den tatsächlichen Sachverhalt (die Lebenswirklichkeit) zu erfassen bzw. präziser: die *differentia specifica* methodenstreng und rational zu wählen – bietet die Rechtsdogmatik für eine intersubjektiv überprüfbare Auswahl bislang keine konsistenten, methodisch expliziten Kriterien; sie ist insofern, wie regelmäßig bei der an der Schnittstelle zwischen Norm und Wirklichkeit operierenden Rechtsanwendung,[25] auf extrajuridische Methoden und interdisziplinäre Wissensbestände angewiesen, die sie dann innerhalb der Eigenrationalität des Rechts verarbeiten muss. [26]

Für die Bildung von Begriffen bieten reine Mathematik (Mengen, Relationen, Äquivalenzklassen – *theoretisch*) und Philosophie (insbesondere

25 Etwa die Frage der „Eignung" im Verhältnismäßigkeitsgrundsatz, *N. Petersen/E. V. Towfigh*, § 1 Ökonomik in der Rechtswissenschaft, in: Towfigh/Petersen (Hrsg.), Ökonomische Methoden im Recht, Rn. 24; *J. Krüper,* Kategoriale Unterscheidung von Öffentlichem Recht und Privatrecht? VVDStRL 79 (2020), S. 43, 73; *A. Morell,* Der Beibringungsgrundsatz, S. 65 ff.

26 *N. Petersen*, Braucht die Rechtswissenschaft eine empirische Wende?, Der Staat 2010, S. 435; *C. Engel*, Empirical Methods, JITE 174 (2018), S. 5; *E. V. Towfigh*, Dogmatik und Empirie, Relevanz und Resonanz: Ein Beitrag in drei Thesen, in: Bezemek (Hrsg.), Rechtsdogmatik, S. 383, 386; *ders.*, Empirical arguments in public law doctrine, I-CON 12 (2014), S. 670; *G. Shaffer/T. Ginsburg*, The Empirical Turn in International Legal Scholarship, AJIL 106 (2012) S. 1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Erkenntnis- und Wissenschaftstheorie, aber auch politische Philosophie – *normativ*) gerade mit Blick auf den *genus proximum* geeignete Ansätze.[27] Bei der Beurteilung der Eignung der *differentia specifica* helfen sie jedoch nicht weiter, sie befassen sich nicht mit der Frage, welche Gruppen in der sozialen oder rechtlichen *Wirklichkeit* sinnvoll unterschieden werden *können*; ihnen fehlt insofern die Verknüpfung mit der konkreten Anwendung. Diesen Wirklichkeitsbezug bietet die der angewandten Mathematik entwachsene sozialwissenschaftliche Statistik. Sie kann helfen, Intuition durch Empirie zu leiten und zu stützen.

II. Heuristik

Diese intuitiven Überlegungen lassen sich zu einer einfachen, dogmatisch fruchtbaren Heuristik verdichten, die an der Schnittstelle von theoretischer Beliebigkeit und praktischer Festlegung ansetzt und das bisher implizite Vorgehen des Bundesverfassungsgerichts explizieren und methodisch absichern kann: Bei der Wahl der *differentia specifica* muss sichergestellt sein, dass diese eine empirisch valide Bezugsgröße darstellen die den Ansprüchen an eine methodengerechte statistische Behandlung genügen können.[28] Diesen Anforderungen hält aber nach der vorgestellten Intuition nur eine solche Gruppenbildung stand, bei der die für die rechtliche Beurteilung relevante Varianz zwischen den durch das Differenzierungsmerkmal gebildeten Gruppen (*between-group variance*) größer ist als die Varianz innerhalb einer Gruppe (*within-group variance*).[29] Die Einführung dieses Maßstabs rationalisiert dann ihre Wahl und macht sie interpersonal vermittel- sowie überprüfbar, ohne die Entscheidungskompetenz der Gerichte in Frage zu stellen.

Der hinter dieser „dogmatischen Heuristik" stehende Gedanke lässt sich auch für die Interpretation von Art. 3 Abs. 3 GG nutzbar machen. Greift man bei einer Gruppe von Menschen eine „Identitätsvariable", also ein zur Selbst- oder Fremdzuschreibung verwendetes Merkmal (z. B. Ge-

27 Anders als der Jubilar ist der Autor leider kein Universalgelehrter, so dass er im Folgenden in Ermangelung des erforderlichen Verständnisses nicht mit Einsichten aus diesen Disziplinen aufwarten kann.

28 *A. Glöckner/E. V. Towfigh*, Messgenauigkeit und Fairness in Staatsprüfungen, AnwBl 2016, S. 706.

29 *K. Deater-Deckard, et al.*, Within- and between-person and group variance in behavior and beliefs in cross-cultural longitudinal data, JOA 2018, S. 207.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

schlecht, Religion), das als *differentia specifica* dienen könnte, heraus, dann ist die Varianz bei Berücksichtigung weiterer Merkmale der Angehörigen dieser Gruppe größer als die Varianz zwischen den gebildeten Gruppen. Es ist kein Zufall, dass die oben gebildeten Beispiele auf die Merkmale Religion, Staatsangehörigkeit/Heimat und „Rasse" rekurrieren: All diese Merkmale kreieren Gruppen, deren rechtlich relevante *within-group variance* regelmäßig größer sein wird als die *between-group variance*. Vor diesem Hintergrund lassen sich die Kategorien des Art. 3 Abs. 2 und 3 GG als *Regelbeispiele* für eine verbotene Gruppenbildung im Rahmen einer umfassenden Gleichheitsgewährleistung des Art. 3 GG interpretieren. Das ermöglicht zum einen eine Erweiterung um nicht genannte Kategorien (z. B. sexuelle Orientierung/Identität, sozialer Status, Vermögen, körperliches Erscheinungsbild), so dass die oben identifizierten Schwächen von Art. 3 Abs. 1 und Abs. 3 dogmatisch behoben werden können: Das Dilemma der Differenz lässt sich auflösen und das Problem der Essentialisierung entschärfen. In einem sehr spezifischen Sinne ließe sich diese Interpretation auch als „postkategorial" einordnen:[30] zwar nicht in dem Sinne, dass die Kategorien aufgelöst oder abgeschafft werden, aber eben doch in dem Sinne, dass die Kategorien anders – anhand eines wissenschaftlichen Maßstabs – gebildet werden müssen. In spezifischen normativen Kontexten werden identitär gedachte Kategorien dabei nach wie vor eine sinnvolle Rolle spielen, jedoch nur dort, wo sie sich auch empirisch verifizieren lassen.

Zum anderen muss bei einem solchen Gleichheitsverständnis für die Feststellung eines Gleichheitsverstoßes nicht mehr zwingend ausdrücklich eine Benachteiligung aufgrund einer der genannten Kategorien gerügt werden, es genügt vielmehr die Feststellung, dass ein hoheitlicher Akt den Gleichheitssatz verletzt, weil er die rechtlich relevante Varianz zwischen den gebildeten Gruppen und innerhalb einer Gruppe nicht berücksichtigt hat.

Dogmatisch lässt sich diese Überlegung als Erfordernis einer heuristischen statistischen Plausibilisierung formulieren: Bei Rechtsetzung und Rechtsanwendung (durch die Verwaltung oder bei deren Überprüfung durch die Gerichte) ist zu begründen, dass die *differentia specifica* diesen Anforderungen *allem Anschein nach* genügt; der Anschein kann erschüt-

30 Vgl. *Susanne Baer*, § 5 Das Kategorienproblem und die Herausbildung eines postkategorialen Anti-Diskriminierungsrechts, in: Mangold/Payandeh (Hrsg.), Handbuch Antidiskriminierungsrecht.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tert werden, indem formal (also in einem strengeren Sinne statistisch) gezeigt wird, dass dies nicht der Fall ist. Auch wenn es sich bei der Anwendung der Heuristik lediglich um ein niedrigschwelliges und immer noch recht grobes Instrument handelt, erlaubt es eine interdisziplinäre „Verwissenschaftlichung" und Rationalisierung der Vergleichsgruppenbildung. Gleichzeitig etabliert sie keine unüberwindbaren Hürden für Menschen, die eine Verletzung von Art. 3 GG geltend machen wollen, da insbesondere im Verwaltungsverfahren und im Verwaltungsprozess der Amtsermittlungsgrundsatz gilt. Dieser stellt sicher, dass sie nicht ungebührlich belastet werden, sondern dass Verwaltung und Gerichte von Amts wegen zur Aufklärung der entscheidungserheblichen Umstände verpflichtet sind. Die Heuristik dient damit als widerlegbarer Anscheinsmaßstab, ohne den Zugang zum Rechtsschutz zu erschweren, insbesondere wenn empirische Modelle fehlen oder Daten schwer zugänglich sind; umgekehrt bürdet sie gerade im Rahmen der Amtsermittlung staatlichen Stellen auf, die Daten beschaffen und auswerten zu lassen, die verfügbar sind.

Die dogmatische Funktion dieser Heuristik besteht also nicht darin, eine normative Entscheidung durch Empirie zu ersetzen. Vielmehr liefert sie einen rationalen Prüfungsmaßstab, der es Gerichten erlaubt, die Plausibilität der gewählten *differentia specifica* im Lichte der verfassungsrechtlich geforderten Gleichbehandlung transparent und intersubjektiv nachzuvollziehen. Statistik dient als methodisches Korrektiv: Sie weist auf Kategorien hin, die empirisch zu heterogen sind, um eine rechtlich belastbare Differenzierung zu tragen, und erleichtert die argumentative Rechtfertigung dort, wo stabile Unterschiede bestehen.

III. Formalisierung

Doch wie kann zuverlässig ermittelt werden, ob wir es mit vergleichbaren Gruppen zu tun haben? Die bislang skizzierte Varianzheuristik beruht auf der Intuition, dass *differentia specifica* nur dann tragfähig sind, wenn die für die rechtliche Beurteilung relevante Varianz zwischen den anhand des fraglichen Kriteriums gebildeten Gruppen größer ist als die Varianz innerhalb der Gruppen. Die Entscheidung des Bundesverfassungsgerichts zur Altersgrenze für Vertragsärzte[31] soll im Folgenden als Beispiel dienen und die Formalisierung veranschaulichen:

31 BVerfG NJW, 1998, 1776.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In dem Fall bildet die Gruppe der Vertragsärzt:innen den *genus proximum*, und sie werden durch die *differentia specifica* „Alter *über* bzw. *unter* 68 Jahren" kategorial unterschieden. Die rechtlich relevante Zielgröße, die den sachlichen Rechtfertigungsmaßstab für die Unterscheidung darstellt, liegt in der ärztlichen Leistungsfähigkeit: Nur in Bezug auf diese Größe lässt sich entscheiden, ob Ärzt:innen über 68 Jahren und ihre jüngeren Kolleg:innen „wesentlich gleich" sind. Zeigt das Merkmal „Alter" mit hinreichender Wahrscheinlichkeit Unterschiede in der Leistungsfähigkeit an, kann es als rechtlich tragfähiges Differenzierungsmerkmal gelten. Dazu ist im nächsten Schritt zu prüfen, ob es im Durchschnitt tatsächlich Unterschiede in der Leistungsfähigkeit verschiedener Alterskohorten gibt.[32] Statistisch ließe sich dies etwa durch Mittelwertvergleiche (etwa hinsichtlich auftretender Komplikationen) ermitteln. Hier lassen sich schon innerhalb der Kategorie „Alter" zahlreiche Differenzierungskriterien bilden und prüfen (also beispielsweise neben 68 Jahren auch 50, 67, 69 oder 75). Bei homogenen Gruppen eignet sich hierzu ein simpler *t*-Test, der jedoch bei ausgeprägter *Heteroskedastizität* (Varianzungleichheit) an Aussagekraft verliert.[33]

Im nächsten Schritt ist die Verteilung der Gesamtvarianz der ärztlichen Leistungsfähigkeit zu analysieren. Die Zerlegung in *within-group*-Varianz und *between-group*-Varianz zeigt, ob die Gruppenzugehörigkeit substanzielle Unterschiede erklärt.[34] Wieviel der Unterschiede in der Leistungsfähigkeit wird durch die Gruppenzugehörigkeit (unter 68/über 68) erklärt und welcher Anteil entfällt auf Unterschiede innerhalb der Gruppen? Ist die Heteroskedastizität groß, also die Varianz innerhalb der Altersgruppen deutlich größer als die Varianz zwischen ihnen, sind also nicht alle über 68-jährige gleich leistungsfähig, spricht dies empirisch gegen die Tragfähigkeit des Kriteriums „Alter über 68" als *differentia specifica.*[35]

---

32 vgl. BVerfG Urteil v. 23. September 2025, Az.: 1 BvR 1796/23, Rn. 129, zur Altersgrenze bei Anwaltsnotaren.

33 Dann sind robustere Tests wie der *Welch*-Test oder der *Levene*-Test vorzuziehen; *T. Schäfer*, Methodenlehre und Statistik, S. 205 ff.; *Student*, The Probable Error of a Mean 1908, S. 1; *B.L. Welch,* The generalization of Student's problem when several different population variances are involved, Biometrika 34 (1947), S. 28. *H. Levene,* Robust tests for equality of variances, in: *I. Olkin* (Hrsg.), Contributions to Probability and Statistics, S. 278 ff.

34 *Deater-Deckard et al.,* (Fn. 29), S. 207.

35 z. B. Quantifizierung mittels Intraklassen-Korrelationskoeffizient (ICC), *P. E. Shrout/J. L. Fleiss*, Intraclass Correlations: Uses in Assessing Rater Reliability, Psycho-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Eine hohe Binnenvarianz weist zudem darauf hin, dass die gewählte *differentia specifica* „Alter" möglicherweise ungeeignet ist. Es sollte geprüft werden, ob alternative Gruppenbildungen methodisch und rechtlich überzeugender wären. Statt der pauschalen Altersgrenze könnten mit Hilfe datengetriebener Verfahren wir *Clusteranalysen* oder *Entscheidungsbäumen* homogenere Gruppen identifiziert werden oder mit Hilfe multivariater Ansätze wie *Diskriminanzanalse* oder logistischer Regression andere Variablen wie z. B. Gesundheitszustand oder Fortbildungsaktivitäten auf ihre Eignung als *differentia specifica* getestet werden. Damit wird empirisch und methodisch abgesichert nachvollzogen, was das Bundesverfassungsgericht ohnehin praktiziert: die Prüfung mehrerer möglicher Differenzierungskriterien.

Je mehr Gruppenmerkmale, Subgruppen oder Vergleichsdimensionen in die Analyse eingehen, je mehr Freiheitsgrade ein Modell hat, desto höher ist allerdings auch das Risiko statistischer Verzerrung durch unbeobachtete Störfaktoren (*confounds*).[36] Die empirischen Werkzeuge müssen daher mit Blick auf Modellgüte, Kontrollvariablen und mögliche Verzerrungsquellen kritisch geprüft werden, bevor sie zur Begründung (oder Infragestellung) rechtlicher Differenzierungen herangezogen werden.[37]

Abgeschlossen wird das Prüfprogramm mit der Einbettung der empirischen Ergebnisse in den verfassungsrechtlichen Rahmen. Dabei ersetzt die Statistik die normative Wertentscheidung nicht, die allein dem Recht zu entnehmen ist, sondern wirkt als zusätzlicher Rationalitätsfilter. Sie macht die Wahl der *differentia specifica* transparent und intersubjektiv überprüfbar und verhindert so – empirische Belastbarkeit vorausgesetzt – willkürlich erscheinende oder empirisch nicht tragfähige Differenzierungen.

C. Ausblick

Die hier entwickelte Varianzheuristik versteht sich als maßvolle Weiterentwicklung der bestehenden Gleichheitsdogmatik. Sie knüpft an die Rechtsprechung des Bundesverfassungsgerichts zu den Gleichheitsrech-

logical Bulletin 1979, S. 420; *R. R. McCrae/A. Terracciano*, Personality Profiles of Cultures: Aggregate Personality Traits, J. Pers. Soc. Psychol. 2005, S. 407.

36 *N. Keiding/D. Clayton*, Standardization and Control for Confounding in Observational Studies: A Historical Perspective, Stat. Sci. 2014, S. 529.

37 auch, um sich nicht dem Vorwurf des *p-Value-hunting* auszusetzen.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ten an und offeriert methodische Leitlinien für die Anwendung der Gleichheitsrechte, indem sie ein empirisch fundiertes Programm bereitstellt, das die gleichheitsrechtliche Prüfung rationalisiert und transparent macht. Damit trägt sie dazu bei, dem unterspezifizierten Allgemeinen Gleichheitssatz des Art. 3 Abs. 1 GG überprüfbare Kriterien zuzuordnen und zugleich die Besonderen Gleichheitssätze des Art. 3 Abs. 2 und 3 GG besser zu verstehen, indem die dort normierten Differenzierungsverbote nicht als metaphysische Festschreibungen einer essentialistischen *differentia,* sondern als paradigmatische Hinweise auf empirisch wie normativ ungeeignete Kategorien verstanden werden können.

Damit trägt die Heuristik auch dazu bei, das Dilemma der Differenz neu zu rahmen und ent-essentialisiert gleichheitsrechtliche Diskurse. Sie bietet ein empirisches Instrument, Gruppenbildungen, die tatsächlich auf stereotypische Zuschreibungen oder kontingente historische Konstellationen zurückgehen als empirisch wenig trennscharf zu entlarven und ihre normative Fragwürdigkeit zu unterstreichen. Das ist freilich nicht in jedem Kontext wünschenswert; der hier präsentierte Ansatz trägt natürlich auch dazu bei, dass gewisse Kriterien oder Merkmale weniger sichtbar oder gar unsichtbar werden. Wenn Identitätsfragen verhandelt werden *sollen*, lassen diese sich namentlich mit den in Art. 3 Abs. 3 GG inkriminierten Differenzierungskriterien besonders leuchtend sichtbar machen und diskutieren. Wenn Diskurse und Rechtsgespräche um die Identitäts-Dimension erleichtert werden und versachlicht werden sollen, dann mag auch im Bereich von Art. 3 Abs. 3 GG und des Antidiskriminierungsrechts der statistische Zugriff hilfreich sein.

Insgesamt eröffnet der Rückgriff auf statistische Methoden eine Form der postkategorialen Rationalisierung des Gleichheitssatzes. Kategorien werden nicht abgeschafft, sondern unter den Vorbehalt gestellt, dass sie empirisch plausibel und rechtlich relevant sind. Dadurch lassen sich willkürliche oder historisch belastete Gruppenzuschreibungen vermeiden, ohne den verfassungsrechtlichen Kerngehalt des Diskriminierungsverbots preiszugeben. Art. 3 GG erweist sich in dieser Lesart als ein empiriesensitives Differenzierungsverbot, das normative Vorgaben nicht aus der Statistik ableitet, wohl aber ihre Plausibilität durch empirische Maßstäbe absichert und so zu einer rationalisierten und zugleich gesellschaftlich sensibleren Gleichheitsdogmatik beiträgt.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

IV.
Institutions

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Das „Netz der Behörden“ als Netzwerk?

Florian Becker

A. Einleitung

An einem spätherbstlichen Nachmittag des Jahres 1999 betrat ich zum ersten Mal das Haus an der Poppelsdorfer Allee. Hier war die Max-Planck-Projektgruppe „Recht der Gemeinschaftsgüter“ beherbergt, aus der wenige Jahre später nach einer erfolgreichen Evaluation das Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern entstehen sollte. Eine Referendarskollegin hatte mich auf die neue Gruppe aufmerksam gemacht und dabei vor allem von deren Direktor, Christoph Engel, geschwärmt, den ich bis zu diesem Zeitpunkt nur aus der juristischen Literatur kannte. Er verwendete an diesem Nachmittag sehr viel Zeit auf unser Kennenlernen und nahm mich schließlich in den Kreis seiner Schüler auf. Dieser Zeitpunkt markierte für mich nicht nur den Beginn eines neuen Lebensabschnitts, sondern leitete auch einen Perspektivwechsel ein. Bis dahin war mein Interesse auf die Rechtswissenschaft und insbesondere das Öffentliche Recht fokussiert. Aber wer mit Christoph Engel in Kontakt kommt, wird sehr schnell von seiner Begeisterung für viele weitere wissenschaftliche Probleme und Fragestellungen jenseits der klassischen dogmatischen Rechtswissenschaften angesteckt – ohne dabei notwendigerweise immer und überall mit seiner unendlichen Neugier, mit seinem voranpreschenden Intellekt sowie mit seiner geschliffenen Rhetorik Schritt halten zu können.

Christoph Engel brachte mir das Phänomen des Netzes und des Netzwerkeffekts nahe und schlug mir eine wissenschaftliche Befassung mit deren öffentlich-rechtlicher Einhegung vor. Diese Arbeit ist aus verschiedenen Gründen nicht zustande gekommen.[1] Dennoch entwickelte sich aus den damaligen Überlegungen und Gesprächen ein Rüstzeug für viele spätere Projekte. Denn Christoph Engel schärft den Blick auf Diszipli-

1 Allerdings griff aus dem Kreis von Christoph Engels Schülern *Jörn Lüdemann* wenige Jahre später das Thema auf, in: *Boysen* et.al., Netzwerke, 2007, 266.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

nen jenseits der Rechtswissenschaft und hat in unzähligen Publikationen selbst meisterhaft gezeigt, wie und was man von ihnen lernen kann. Damals war das Netzwerk als Forschungsprojekt – aus öffentlich-rechtlicher Sicht allemal – tatsächlich ein echter Geheimtipp. In der Folge war es dann für einen gewissen Zeitraum in (fast) aller Munde. Zwar ist es inzwischen etwas ruhiger um den Begriff geworden, aber die 2000er Jahre haben ihre Spuren hinterlassen.

B. Netze und Netzwerkeffekte

I. Abstraktionshöhe und Ubiquität

Das Netzwerk setzt sich aus miteinander durch Kommunikation sowie ein- oder wechselseitige Einwirkungsbefugnisse verbundenen Knotenpunkten zusammen. Dabei kann das Netz in seiner Gesamtheit wegen seiner Heterogenität nicht über die Eigenschaft eines Akteurs verfügen. Auch wenn seine Akteure typischerweise einen einheitlichen, wenn auch sehr abstrakten Zweck verfolgen, bleiben die einzelnen Knotenpunkte die entscheidenden Akteure.

Mit dieser Beschreibung ist allerdings noch nicht viel gewonnen. Sie sagt nichts über die Charakterisierung oder die wechselseitigen und -bezüglichen Verhältnisse der Akteure (Hierarchie? Heterarchie?), über die Art ihrer Kommunikation, das Maß und die Folgen der Interaktion oder die Zurechnung des erzeugten Endprodukts aus. Erst recht erfährt man nichts über die Erwartung, die mit der Einordnung eines Phänomens als Netz etwa an dessen interne Dynamiken oder an seinen Output gerichtet wird.

Diese Abstraktionshöhe hat dazu geführt, dass die Begriffe Netz und Netzwerk als Bild, Symbol, Metapher, Allegorie oder schlichte Beschreibung in einer ganzen Reihe von Wissenschaftsdisziplinen und Zusammenhängen präsent sind.[2] Die Begriffe beschreiben eine Vielfalt an Phänomenen, sie werden zu ganz unterschiedlichen Zwecken verwendet und sie befriedigen völlig unterschiedliche Erkenntnisinteressen. Sie finden auch normative Verwendung, um die Erwartungen des Betrachters v. a. an bestimmte Strukturen zu formulieren.

2 Umfassend *Schwind*, Netzwerke im Europäischen Verwaltungsrecht, 2017, 19 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Angesichts dieser Vielgestaltigkeit fällt es schwer, eine auch nur einigermaßen umfassende Definition, eine Kategorisierung des Begriffs oder auch nur eine Aufzählung seiner Verwendungen und der damit einhergehenden deskriptiven oder normativen Aussagen zu bieten, die über Allgemeinplätze hinausgeht. Die auf dieser Ubiquität beruhende Beliebigkeit (und in gewissem Maße auch Konturenschwäche) des Begriffs lässt seine dogmatische Leistungsfähigkeit zweifelhaft erscheinen. Denn über die Rechtsfolgen, die sich aus einer konkreten, entdifferenzierenden Verflechtung ergeben, kann nur das einschlägige Recht Antwort geben.[3] Außerdem bestand und besteht die Gefahr, dass jedwede Kooperation von irgendwelchen Akteuren als „Netzwerk" bezeichnet wird, ohne dass dies einen besonderen Mehrwert hätte.

II. Das Netz als Rechtsbegriff

Ansatzpunkte für ein dogmatisches Verständnis des Netzwerks bestehen dort, wo es als Gesetzesbegriff Verwendung findet. Insbesondere das Wettbewerbsrecht im Allgemeinen, das Regulierungsrecht im Besonderen, kennt Netze schon seit langem. Ihre Präsenz im EnWG, TKG oder AEG[4] macht deutlich, dass es sich heute auch um einen Rechtsbegriff handelt. Für die Energieindustrie, die (sozialen wie althergebrachten) Kommunikationsdienste, die Wasserwirtschaft sowie für das Transportwesen (zu Wasser, Land und in der Luft) spielt das Netz als Transportweg für Energie, Personen, Güter oder Kommunikation bei der Leistungserbringung eine zentrale Rolle. Größen- und Verbundvorteile in diesen Netzen sind die Grundlage für die Bildung natürlicher Monopole.[5] Ein einziges Netz kann den Markt hier effizienter versorgen als eine Mehrzahl.[6] In den Netzen dieser Industrien entfaltet sich ein Netzwerkeffekt:[7] Für gewöhnlich sinkt der Wert eines Guts mit der Menge des Angebots. Dessen Inhaber hat also selten Interesse daran, dass möglichst viele Stücke dieses Guts existieren. Bei bestimmten Gütern ist dies aber genau umgekehrt: Deren Wert steigt für ihre einzelnen Inhaber, je mehr von ihnen auf dem

3 So auch *Schmidt-Aßmann*, Verwaltungsrechtliche Dogmatik, 2. Aufl. 2023, 206.
4 Bspw. §§ 4, 6 ff., 11 EnWG, §§ 2 Abs. 2 Nr. 1, 3 Nr. 42, 51 Abs. 1 TKG, §§ 2b f. AEG.
5 *Fritsch*, Marktversagen und Wirtschaftspolitik, 10. Aufl. 2018, 163 ff.
6 *Schmidt/Haucap*, Wettbewerbspolitik und Kartellrecht, 10. Aufl. 2023, 44 f.
7 Zu dem Folgenden *Lemley/McGowan*, Cal. L. Rev. Bd. 86 (1998), 479 (483 ff.).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Markt verfügbar sind. Die Wertschätzung bzw. der Wert des Gutes wächst mit der Anzahl weiterer Nutzer. Dieser Effekt ist in netzgebundenen Industrien zu beobachten: Ein Telefon, das an ein Telefonnetz angeschlossen ist, wird mit jeder weiteren ebenfalls angeschlossenen Person wertvoller. Der Netzwerkeffekt stärkt die Wettbewerbsposition großer Unternehmen gegenüber neu in den Markt eintretenden, weil die Kunden eher dem alten Netz vertrauen, dort mehr Verbindungen herstellen können und sich ihm deswegen anschließen.[8] Der Netzwerkeffekt erschwert somit den Markteintritt neuer Unternehmen, sorgt auf diese Weise für wettbewerbsrechtliche Probleme und regulatorische Herausforderungen. Er begründet zudem Pfadabhängigkeiten, die die weitere technische Entwicklung beeinflussen.[9] Solche natürlichen Monopole müssen natürlich nicht immer jedwedem Wettbewerb auf allen Ebenen der Wertschöpfungskette entgegenstehen. Der Staat kann für einen Wettbewerb verschiedener Anbieter im Netz durch Zugangs- und Zusammenschlussregelungen, für einen Wettbewerb um den zeitweisen exklusiven Betrieb eines Netzes durch Ausschreibung oder aber durch die Förderung substitutiver Netze für einen Wettbewerb zwischen Netzen sorgen.[10]

Aber auch der Staat macht sich den Netzwerkeffekt zunutze, der etwa in der Idee der Sozialversicherung abgebildet wird und dort die Tragfähigkeit der ihr zugrundeliegenden Idee auch jenseits sozialstaatlicher Aspiration illustriert: Bestünde die gesetzliche Krankenversicherung nur aus einem einzelnen beitragspflichtigen Mitglied, wäre die Organisation bei dessen ernsthafter und kostenträchtiger Krankheit sofort wirtschaftlich überfordert. Je mehr Menschen aber der Sozialversicherung angehören, desto eher bildet sich in der Mitgliedschaft der durchschnittliche gesundheitliche Status der Gesamtbevölkerung ab und desto eher können auch erhebliche Ausgaben zugunsten einzelner durch die Solidargemeinschaft getragen werden. Das System funktioniert indes nicht mehr ohne steuerliche Querfinanzierung, wenn jedes Mitglied für sich (und seine

8 *Lemley/McGowan*, Cal. L. Rev. Bd. 86 (1998), 479 (495 ff.).

9 *Liebowitz/Margolis*, JLEO Bd. 11 (1995), 205.

10 Siehe schon umfassend zu den Konzepten *Klein*, Competition in Network Industries, World Bank Policy Research Working Paper 1591 (1996); Zugangsverpflichtungen z. B. in § 21 TKG für Unternehmen mit dominierender Marktmacht über Telekommunikationsnetze; § 20 Abs. 1 EnWG für Netzbetreiber der Energiewirtschaft; Zugang zu Eisenbahnanlagen § 10 ERegG; zur exklusiven Vergabe von Personenverkehrsdiensten auf der Schiene s. VO (EG) 1370/2007 vom 23. 10. 2007, ABl. 2007 L 315/1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Mitversicherten) mehr an Leistungen beansprucht, als im Durchschnitt an Beiträgen erbracht wird.

C. Organisationen als Netzwerke

Begriff und Bauform des Netzes finden aber auch in ganz anderen Zusammenhängen und jenseits der greifbaren und gut erforschten Netzwerkindustrien Verwendung. So gibt es Versuche, den Staat insgesamt als Netzwerk – aus Behörden, Vereinigungen, Individuen – zu rekonstruieren.[11] Eine ähnliche Beschreibung oder Erwartung kommt dem Begriff des Netzwerks zu, wenn er im Bereich der Verwaltungsorganisation im weitesten Sinne verwendet wird. Solche Behördennetzwerke existieren vor allem im europäischen Mehrebenensystem.[12] Anders als bei den sehr theoretischen und abstrakten Konzeptionalisierungen finden sich für deren Konstruktion und Funktionieren allerdings naturgemäß Regelungen im positiven Recht.

I. Entdifferenzierung

Solche Netzwerke sind in mehrfacher Hinsicht Orte der Entdifferenzierung.[13]

Sie greifen über organisatorisch voneinander getrennte Einheiten (Kommunen, Länder, Bund, EU) hinaus. Auf verfassungsrechtlicher Ebene reiben sie sich damit an dem grundgesetzlichen Verbot der Mischverwaltung.[14] Dieses Verbot spielt im Schnittbereich des Unions-

11 Vgl. *Teubner*, in: FS für Ingeborg Maus, 1999, 346 ff.; *Ladeur*, Globalization and the Conversion of Democracy to Polycentric Networks: Can Democracy Survive the End of the Nation State?, in: ders. (Hrsg.), Public Governance in the Age of Globalization, 2004, 89 ff., ebenfalls abgedruckt in EUI Working Paper Law 2003/4.

12 Hierzu *Schwind*, Netzwerke im Europäischen Verwaltungsrecht, 2017; *Simantiras*, Netzwerke im europäischen Verwaltungsverbund, 2016.

13 Die einzelnen Aspekte finden sich bei *Möllers*, in: Oebbecke (Hrsg.), Nicht-Normative Steuerung in Dezentralen Systemen, 2005, 285 (295 ff.); *Schwind*, Netzwerke im Europäischen Verwaltungsrecht, 2017, 132 ff.

14 Dazu BVerfGE 119, 331 (Arge); Dürig/Herzog/Scholz/*F. Kirchhof*, 107. EL März 2025, GG Art. 83 Rn. 117; Huber/Voßkuhle/*Trute*, 8. Aufl. 2024, GG Art. 83 Rn. 28 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

verfassungsrechts mit den Verfassungsordnungen der Mitgliedstaaten zwar kaum eine Rolle, aber dort setzt der Grundsatz der begrenzten Einzelermächtigung (Art. 5 EUV) das freihändige Verschieben und Vermischen von Zuständigkeiten zwischen den verschiedenen Ebenen unter Rechtfertigungsvorbehalt.[15]

Die Differenzierungen zwischen Staat und Privat und die daraus abzuleitenden Unterschiede zwischen staatlicher Kompetenz und privater Handlungsfreiheit, Grundrechtsbindung und -berechtigung, gesetzlich bestimmter Gemeinwohlverpflichtung und legitimem Eigennutz sowie zwischen öffentlichem und privatem Recht[16] werden in Netzwerken ebenfalls zumindest teilweise aufgehoben, wenn in ihnen – was regelmäßig der Fall ist – Akteure des i. w. S. staatlichen mit denen des gesellschaftlichen Funktionsbereichs zusammenwirken.

Ausdruck des organisationsrechtlich geprägten Gegensatzes von staatlichem/hoheitlichem und privatem Funktionsbereich ist die Differenzierung zwischen kooperativem und hierarchischem Handlungs- und Entscheidungsmodus, die in den Handlungsformen Rechtsnorm und Verwaltungsakt einerseits und Vertrag andererseits idealisiert wird. Eine Auflösung dieser Gegensätze bringt es zugleich mit sich, dass die Zusammenarbeit im Netzwerk weniger einen hierarchischen als vielmehr einen heterarchischen Charakter aufweist.

Netzwerke sind des Weiteren durch die Auflösung des Gegensatzes von formellen und informellen Handlungs-, Entscheidungs- und Interaktionsformen gekennzeichnet. Zwar bestehen oft gesetzliche Vorgaben für Zuständigkeit, Kooperation und Informationsfluss. Aber neben diese formalen Regeln treten auch Übungen und wechselseitige Abstimmungen jenseits institutionalisierter Kanäle, die sich im Rahmen eines tatsächlichen Näheverhältnisses zwischen den Akteuren mit der Zeit entwickeln. Wenn dieses Näheverhältnis vor allem zwischen behördlichen und privaten Akteuren zu eng wird, entsteht in regulierten Lebensbereichen ein „regulatory capture".[17]

Netzwerkartige Organisationen können sich (sogar auch im staatlichen Funktionsbereich) spontan (evolutiv) bilden. Sie wirken dann zwar informell zusammen, aber ihr tatsächlicher Einfluss in bestimmten Hand-

15 Grabitz/Hilf/Nettesheim/*Bast*, 85. EL Mai 2025, EUV Art. 5 Rn. 15.
16 Dazu *Becker*, in: Kahl/Ludwigs (Hrsg.), VerwR-HdB, Bd. VI 2024, § 160.
17 Allg. *Masing*, AöR 128 (2003), 558 (592 ff.).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lungsfeldern kann enorm sein.[18] Häufiger werden Netzwerke staatlicher Funktionsträger (v. a. Behörden) aber durch Gesetz oder Vertrag gebildet und mit einem Handlungsauftrag sowie entsprechenden Handlungsinstrumenten ausgestattet. Dies ist dann ein Fall (fremd-)gesteuerter Netzwerkbildung.

Netzwerke können offen oder geschlossen sein. Entweder steht ihr Teilnehmerkreis fest und kann dann auch nicht ohne weiteres verändert werden. Hier können dann bestenfalls noch Akteure informell in das Netzwerk eintreten. Oder aber die Netzwerke bilden eine offene Partizipationsstruktur und erlauben spontane Veränderungen in der Akteurszusammensetzung durch das Hinzutreten oder den Austritt einzelner Mitglieder.

II. Zweckverwirklichung

Das Zusammenwirken im Netzwerk ist einem bestimmten Zweck gewidmet, dessen Verwirklichung die einzelnen Akteure dienen. Dabei ist zwischen jenem übergeordneten Zweck des Netzwerks insgesamt – seiner Existenzberechtigung – und dem individuellen Motiv für die Teilnahme an dem Netzwerk zu differenzieren. Dies wird bei einem Zusammenwirken von Staat und Privaten im Netzwerk besonders deutlich: Obschon staatliche Akteure stets auch institutionelle Interessen verfolgen, haben sie ihre Tätigkeit doch grundsätzlich an einem heteronomen gesetzlichen Auftrag und den hierauf aufbauenden, den Auftrag konkretisierenden einzelnen rechtlichen Vorgaben auszurichten. Private Akteure sind demgegenüber grundsätzlich nicht fremdbestimmt und keinem fremden Zweck verpflichtet. Ihre Motivation zur aktiven Teilnahme an einem Netzwerk ist typischerweise grundrechtlich geschützt und bedarf keiner Rechtfertigung. Damit wirken bei einem Netzwerk, das sich aus staatlichen wie auch aus privaten Akteuren zusammensetzt, gesetzlich determinierte, gemeinwohlorientierte Motivation einerseits sowie rationaleigennützige Motivation andererseits zusammen, um auf diese Weise ein gewünschtes Ergebnis zu erreichen.

Die Akteure wirken an der Verfolgung eines gemeinsamen – selbst gesetzten oder heteronom bestimmten – Ziels mit, ohne dass sie dabei

18 So etwa zum Basler Ausschuss für Bankenaufsicht *Zaring*, Texas International Law Journal 33 (1998), 281 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ihre individuelle, akteurspezifische Motivation notwendigerweise aufgeben (müssen).[19] Dies ist nur anders zu beurteilen, wenn die Privaten durch staatlichen Zwang in das Netzwerk integriert werden. Der Akt der Zwangsintegration stellt nicht nur einen Grundrechtseingriff dar, sondern der Gesetzgeber muss auch Instrumente bereithalten, um dauerhaft sicherzustellen, dass sich die Privaten bei ihren Handlungen dem Netzwerkzweck unterordnen und ihre Position nicht ausschließlich zur Verfolgung eigener Zwecke zu dessen Lasten missbrauchen. Soweit hierbei ebenfalls hoheitlicher Zwang angewendet wird, ist auch dieser an den Grundrechten der Beteiligten zu messen.

D. Das „Netz von Behörden" im Wettbewerbsrecht

Netzwerke öffentlicher Akteure gibt es – hier einmal mehr, dort einmal weniger verdichtet – inzwischen in einigen Bereichen des Verwaltungsrechts im Mehrebenensystem,[20] so etwa bei der Aufsicht über Finanzinstitute.[21] Das bekannteste Beispiel für ein von öffentlichen Akteuren gebildetes und als solches bezeichnetes Netz ist das aus der Kommission sowie den mitgliedstaatlichen Wettbewerbsbehörden zusammengesetzte „Netz von Behörden", dessen Akteure die hoheitliche Durchsetzung von Art. 101 f. AEUV gegenüber Unternehmen im Binnenmarkt verantworten.[22]

Dieses durch die VO (EG) 1/2003[23] etablierte Netzwerk dient nach seiner ursprünglichen Bestimmung der hoheitlichen Durchsetzung des Wettbewerbsrechts durch Untersuchungen, Anordnungen und Sanktionen, mithin unter Nutzung hoheitlicher Handlungsinstrumente. Das Netzwerk soll eine effiziente Arbeitsteilung zwischen seinen Gliedern – den mitgliedstaatlichen Behörden, der Kommission sowie den Gerich-

19 *Schwind*, Netzwerke im Europäischen Verwaltungsrecht, 2017, 131 ff.

20 Weitere Beispiele bei Schoch/Schneider/*Schoch*, 6. EL November 2024, Verwaltungsrecht Einleitung VwVfG Rn. 540 ff.; *Schwind*, Netzwerke im Europäischen Verwaltungsrecht, 2017, §§ 7, 8, 9; *Sydow*, Verwaltungskooperation in der Europäischen Union, 2004, 129 ff, 160 ff., 190 ff., 222 ff.

21 Calliess/Ruffert/*Häde*, 6. Aufl. 2022, AEUV Art. 127 Rn. 61 ff.

22 Erwägungsgründe 15 ff. und Art. 11 ff. VO (EG) 1/2003; s. a. § 50a GWB.

23 VO (EG) 1/2003 vom 16. 12. 2002, ABl. 2003 L 1/1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ten[24] – ermöglichen und zugleich die einheitliche Anwendung des Wettbewerbsrechts sicherstellen.[25]

Der europäische Gesetzgeber verspricht sich von einer solchen Organisations- und Interaktionsstruktur nicht nur, dass die Vorteile dezentraler Rechtsanwendung und das Bedürfnis nach unionsweiter Uniformität miteinander in Einklang gebracht werden, sondern auch, dass der Kommission die Konzentration auf die wichtigen Verfahren und Entscheidungen ermöglicht und vorbehalten bleibt.[26] Um aber diese Vorteile des Netzwerks auch auszuspielen, sind Regelungen über Zuständigkeiten sowie über gegenseitige Informationspflichten und -rechte und Regelungen zur Vermeidung kollidierender Entscheidungen erforderlich.

In die erstgenannte Kategorie fallen diejenigen Vorschriften, die einzelnen Elementen des Netzwerks bestimmte Aufgaben routinemäßig zuweisen. So begründen Art. 5 und 6 VO (EG) 1/2003 die grundsätzliche Zuständigkeit der nationalen Behörden und Gerichte zur Verfolgung und Sanktionierung von Verstößen gegen Art. 101 und Art. 102 AEUV. Schicken sich mehrere Behörden an, einen Verstoß zu untersuchen, ermöglicht Art. 13 VO (EG) 1/2003 die Aussetzung einzelner Verfahren. Die Auswahl der – möglichst alleine – verfahrensführenden Behörde(n) richtet sich nach der Netzwerkbekanntmachung und der Gemeinsamen Erklärung.[27] Neben die mitgliedstaatliche Verfolgung von Wettbewerbsbeschränkungen tritt die Zuständigkeit der Kommission, die jedes Verfahren in jedem Stadium voraussetzungslos an sich ziehen und dadurch die Zuständigkeit der mitgliedstaatlichen Wettbewerbsbehörden beenden kann (Art. 11 Abs. 6 VO (EG) 1/2003). Diese hierarchische Funktion ist untypisch für ein Netzwerk.

Solche Zentralisierungsentscheidungen sind nur möglich, wenn die Kommission über mitgliedstaatliche Verfahren ins Bild gesetzt wird. Eine

---

24 Trotz der Bezugnahme auf „Behörden" spielen allerdings auch die mitgliedstaatlichen und europäischen Gerichte eine Rolle in dem Netzwerk; vgl. Art. 6 VO (EG) 1/2003 und Erwägungsgrund 33.

25 Erwägungsgrund 15 zur VO (EG) 1/2003.

26 Erwägungsgrund 3 ff. zur VO (EG) 1/2003.

27 Bekanntmachung der Kommission über die Zusammenarbeit innerhalb des Netzes der Wettbewerbsbehörden, ABl. 2004 C 101/43; Gemeinsame Erklärung des Rates und der Kommission zur Arbeitsweise des Netzes der Wettbewerbsbehörden (https://competition-policy.ec.europa.eu/antitrust-and-cartels/european-competition-network/documents_en; letzter Abruf 14.10.2025); zur Exklusivität vgl. Netzwerkbekanntmachung Rn. 7 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

wichtige Funktion aller Netzwerke ist der – mehr oder weniger vertrauensvolle – Austausch von Informationen und Wissen zwischen den Akteuren. Der Informationsfluss findet in informellen Arrangements statt, wird aber auch durch Informationsansprüche und -pflichten beflügelt. In diesem Sinne regeln Art. 11 bis 15 VO (EG) 1/2003 Einzelheiten der Zusammenarbeit von Kommission und den nationalen Wettbewerbsbehörden und insbesondere die Verpflichtung der mitgliedstaatlichen Behörden, die Kommission nach einer Verfahrenseröffnung hierüber in Kenntnis zu setzen (Art. 11 Abs. 3 VO (EG) 1/2003) und ihr auch rechtzeitig einen Entscheidungsentwurf zu überlassen (Art. 11 Abs. 4 VO (EG) 1/2003). Daneben ermöglicht Art. 12 VO (EG) 1/2003 einen Austausch von Informationen zwischen Kommission und mitgliedstaatlichen Behörden.

Wo eine Vielzahl unterschiedlicher Behörden und Gerichte an der Verfolgung eines einheitlichen Ziels beteiligt ist, besteht naturgemäß die Gefahr divergierender Rechtsanwendung, die die Effizienz der Zweckverfolgung und sogar die Stabilität des Netzes beeinträchtigen kann. Auch die klammheimliche Verschonung nationaler Champions durch mitgliedstaatliche Behörden ist nicht auszuschließen. Daher bedarf es in einem Netzwerk nicht nur der regelhaften Verteilung, sondern auch der Sicherstellung nicht nur minimaler Uniformität und der Durchsetzung des Vorrangs des Unionsrechts, den Art. 3 VO (EG) 1/2003 unterstreicht. In diesem Sinne ergänzt Art. 16 VO (EG) 1/2003 das Selbsteintrittsrecht der Kommission zur Sicherstellung einer unionsweiten Einheitlichkeit bei der Anwendung des Wettbewerbsrechts durch die mitgliedstaatlichen Behörden im Netzwerk. Nach dieser Vorschrift dürfen Behörden und Gerichte der Mitgliedstaaten keine Entscheidungen treffen, die denen der Kommission zuwiderlaufen.

Der Begriff des „Netzwerks" lenkt den Blick auf die Interaktion und die Interdependenz der beteiligten Behörden. Doch neben diese kooperierende Funktion tritt in dem konkreten Fall der europäischen Behördennetzwerke auch eine hierarchische.

E. Integration Privater und ihrer Interessen in das Netzwerk

Im Wettbewerbsrecht und in dem Verfahren seiner Durchsetzung lässt sich ein weiterer Aspekt beobachten, der bei der theoretischen Erschließung von Netzwerken eine Rolle spielt: Es geht um die Integration privater Handlungsbeiträge zur Erreichung des Netzwerkzwecks.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I. Dezentrale und privatisierte Rechtsdurchsetzung im Allgemeinen

In vielen Bereichen erfolgt die Durchsetzung des Unionsrechts in und gegenüber den Mitgliedstaaten nicht von oben herab durch die Kommission im Wege des ihr an die Hand gegebenen, aber sehr aufwendigen Vertragsverletzungsverfahrens. Erst recht in einer gewachsenen und dadurch immer unübersichtlicher gewordenen Union mit 27 Mitgliedstaaten bauen EuGH und Kommission (auch) auf eine dezentrale Durchsetzung des Unionsrechts in den Mitgliedstaaten. Die Basis dieser dezentralen (und privaten) Rechtsdurchsetzung bilden unmittelbare Anwendbarkeit sowie Vorrang des Gemeinschafts- und jetzt des Unionsrechts[28], deren Wirksamkeit durch den europarechtlich initiierten Staatshaftungsanspruch für Unionsrechtsverstöße abgerundet wurde.[29]

Diese Entwicklungen ermöglichten es dem einzelnen Gemeinschafts- bzw. Unionsbürger, seine aus den Verträgen fließenden Rechte unmittelbar und selbst – d. h. ohne Einschaltung der Kommission – vor mitgliedstaatlichen Gerichten geltend zu machen. Hierin lag nicht nur eine Ermächtigung bzw. Subjektivierung des Einzelnen, sondern zugleich auch ein Hebel zur dezentralen Durchsetzung des Gemeinschaftsrechts und damit zur Überwindung mitgliedstaatlicher Souveränität.

In Bereichen, in denen besondere Vollzugsdefizite zu beobachten waren, traten als weitere Elemente dezentraler Rechtsdurchsetzung sekundärrechtlich begründete individuelle Auskunfts- und Verbandsklagerechte hinzu. Hier ist insbesondere für das in den siebziger und achtziger Jahren mit dem Ruf eines notorischen Vollzugsdefizits[30] belastete Umweltrecht auf die Rechtsentwicklung zu verweisen, die durch die im Jahr 1998 im Rahmen der United Nations Economic Commission for Europe unterzeichnete Åarhus-Konvention ausgelöst wurde.[31] Die Bürger der Konventionsstaa-

28 EuGH, Urt. v. 5.2.1963 – Rs. 26/62 (van Gend & Loos) – Slg. 1963, 3; EuGH, Urt. v. 15.7.1964 – Rs. 6/64 (Costa/E.N.E.L.) – Slg. 1964, 1141 (1269 f.); *Polzin*, in: Ludwigs (Hrsg.), EU-WirtschaftsR-HdB, 63. EL Mai 2025, § 4 Rn. 2 ff.

29 EuGH, Urt. v. 19.11.1991 – Rs. C-6/90 u. 9/90 (Francovich) – Slg. 1991, I-5403; EuGH, Urt. v. 5.3.1996 – Rs. C-46/93 u. C-48/93 (Brasserie du Pêcheur) – Slg. 1996, I-1131; BeckOK BGB/*Kümper*, 75. Ed. 1.8.2025, BGB § 839 Rn. 101 ff.; *Knops*, NJW 2023, 2225 ff.

30 Hierzu *Ramsauer*, in: Koch/Hofmann/Reese (Hrsg.), UmweltR-HdB, 6. Aufl. 2024, § 3 Rn. 80 f.

31 Übereinkommen über den Zugang zu Informationen, die Öffentlichkeitsbeteiligung an Entscheidungsverfahren und den Zugang zu Gerichten in Umweltangelegenheiten v. 25.6.1998, ABl. 2005 L 124, 4.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ten und damit auch die Unionsbürger sollen durch Informations-, Beteiligungs- und Klagerechte zur Durchsetzung des EU-Rechts, insbesondere des Umweltrechts gegenüber ihrem Staat „mobilisiert" werden.[32] Insoweit bieten vor allem Umweltverbände die Gewähr für hohen Sachverstand und verfügen über ein erhebliches Partizipationsinteresse. Die Gefahr späterer Klageverfahren kann schon zu Kooperationen zwischen Verbänden und Behörden während des Verwaltungsverfahrens führen, die dann in eine besser informierte und weniger klageanfällige Planungs- oder Genehmigungsentscheidung münden. Aber auch wenn es nicht von der Hand zu weisen ist, dass die frühzeitige Integration solcher Verbände in den Prozess der Entscheidungsfindung zu einer höheren Akzeptanz einer eigentlich nicht gewünschten Entscheidung führen kann, steht doch hier die Idee der dezentralen Durchsetzung von Unionsrecht gegenüber mitgliedstaatlichen Entscheidungen mit Umweltrelevanz im Vordergrund. Der Private wird auf diese Weise zumindest zu einem informell an der Entscheidung Beteiligten und soll – idealerweise – der Garant für die Einhaltung von Rechtmäßigkeitsvoraussetzungen sein.

Hier sind vorsichtige Ansätze des Netzwerksgedankens zu identifizieren: Eine umweltrechtlich relevante Entscheidung wird im Zusammenwirken privater und behördlicher (sowie gegebenenfalls auch gerichtlicher) Netzwerkknotenpunkte getroffen, die im Zuge der Entscheidungsfindung miteinander kommunizieren und mit den ihnen jeweils zugewiesenen Rechten und Befugnissen aufeinander einwirken. Die Zurechnung des „Endprodukts" – der umweltrelevanten Planungs- oder Zulassungsmaßnahme – bleibt aber mit der Erlassbehörde eindeutig (und muss dies aus Gründen des Rechtsschutzes und der Legitimation auch sein).

II. Private Enforcement im Wettbewerbsrecht

1. Dimensionen des Private Enforcement

Vor inzwischen mehr als zwei Jahrzehnten hat sich auch für das EU-Wettbewerbsrecht die Erkenntnis durchgesetzt, dass dessen Durchsetzung al-

32 *Markus/Verheyen*, in: Koch/Hofmann/Reese (Hrsg.), UmweltR-HdB, 6. Aufl. 2024, § 1 Rn. 112 ff.; siehe auch das Gesetz über ergänzende Vorschriften zu Rechtsbehelfen in Umweltangelegenheiten nach der EG-Richtlinie 2003/35/EG (Umwelt-Rechtsbehelfgesetz) mit der Möglichkeit altruistischer Verbandsklagen. Dazu *Schwerdtfeger*, in: Koch/Hofmann/Reese (Hrsg.), UmweltR-HdB, 6. Aufl. 2024, § 4 Rn. 69 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lein durch mitgliedstaatliche Behörden und die Kommission nicht einmal annähernd umfassend möglich sein dürfte.[33] Kapazitätsmängel und daraus resultierende Priorisierungserwägungen der Wettbewerbsbehörden führten damals wie heute dazu, dass diese nicht alle Wettbewerbsverstöße ermitteln und verfolgen können. Was lag also näher, als die Aufgabenerfüllung des Netzwerks der Behörden durch die Integration neuer Akteure wirksamer werden zu lassen?

Bereits die VO (EG) 1/2003 hatte den Unternehmen als Adressaten des Wettbewerbsrechts eine wichtigere Rolle bei dessen Anwendung (weniger bei dessen Durchsetzung) zugewiesen: Vor deren Erlass oblag es den Unternehmen lediglich, Vereinbarungen zur Freistellung bei der Kommission anzumelden (Art. 4 Abs. 1 VO 17/62). Diese hatte dann die ausschließliche Befugnis, die Vereinbarung unter die Freistellungsvoraussetzung von (jetzt) Art. 101 Abs. 3 AEUV zu subsumieren (Art. 9 Abs. 1 VO 17/62). Mit der VO (EG) 1/2003 ist diese Einbeziehung deutlich stärker ausgeprägt worden. Die Verordnung hat die Selbstverantwortung der Marktteilnehmer deutlich erhöht: Sie müssen nunmehr aufgrund von Art. 1 Abs. 1 und 2 VO (EG) 1/2003 eine eigene Einschätzung über die Rechtmäßigkeit von wettbewerbsbeschränkenden Vereinbarungen treffen (s. a. Erwägungsgrund Nr. 3 VO (EG) 1/2003) und das aus dieser Entscheidung resultierende Risiko eventueller Fehleinschätzungen tragen. Zwar waren die Unternehmen schon vor der Novelle an der Anwendung des Wettbewerbsrechts beteiligt, doch ist der Schritt von der Notwendigkeit einer Anmeldung zur selbstverantwortlichen Entscheidung groß.

Eine weitere, zum Teil parallel angestoßene Entwicklung birgt das Potential zu einer noch grundlegenderen Veränderung bei der Durchsetzung des Wettbewerbsrechts. Dabei handelt es sich um die Integration derjenigen in die Durchsetzung der Ziele des Wettbewerbsrechts, die durch wettbewerbsbeschränkende Vereinbarungen geschädigt werden. Solche Vereinbarungen haben negative Auswirkungen auf die Gesamtwohlfahrt.[34] Entlang der Wertschöpfungskette können Unternehmen wettbewerbswidrige Einkaufspreise vielleicht noch an die nächste Stufe der Wertschöpfung weiterreichen. Aber spätestens der einzelne Verbraucher bekommt die Auswirkungen der rechtswidrigen Vereinbarung individuell zu spüren. Denn typischerweise sind die Verbraucher nicht

33 Grünbuch Schadenersatzklagen wegen Verletzung des EU-Wettbewerbsrechts, KOM (2005) 672 endg. vom 19.12.2005, 3 ff.

34 Vgl. Immenga/Mestmäcker/*Zimmer*, 7. Aufl. 2025, AEUV Art. 101 Abs. 1 Rn. 4 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

nur mit geringerer Qualität und weniger Auswahl konfrontiert, sondern erleiden einen Schaden in Form von überhöhten Preisen.

Spätestens seit der Veröffentlichung eines Grünbuchs[35] verfolgt die Kommission das Anliegen, das Private Enforcement durch die Geltendmachung von Schadensersatzansprüchen auch in der EU neben die hoheitliche Durchsetzung des Wettbewerbsrechts zu stellen und so die Vereinbarung von Wettbewerbsbeschränkungen wirtschaftlich zumindest potentiell unattraktiver zu machen. Während private Schadensersatzklagen in den USA schon lange zu dem Instrumentenkasten des Antitrust-Enforcement gehört,[36] schlummerte diese Option in der EU recht lange vor sich hin.

In seiner insoweit wegweisenden Entscheidung Courage/Crehan betonte der EuGH, dass die volle Wirksamkeit des Verbots wettbewerbsbeschränkender Vereinbarungen nach (inzwischen) Art. 101 Abs. 1 AEUV und die hierauf aufbauende Nichtigkeit entsprechender Abreden (Art. 101 Abs. 2 AEUV) in ihrer Effektivität beeinträchtigt wären, „wenn nicht jedermann Ersatz des Schadens verlangen könnte, der ihm durch einen Vertrag, der den Wettbewerb beschränken oder verfälschen kann, oder durch ein entsprechendes Verhalten entstanden ist. Ein solcher Schadensersatzanspruch erhöht nämlich die Durchsetzungskraft der gemeinschaftlichen Wettbewerbsregeln und ist geeignet, von – oft verschleierten – Vereinbarungen oder Verhaltensweisen abzuhalten, die den Wettbewerb beschränken oder verfälschen können".[37]

Der Gerichtshof macht deutlich, dass die Geltendmachung von Schadensersatzansprüchen nicht allein dem einzelnen Anspruchsteller dient, sondern auch durch Entfaltung einer spezial- wie generalpräventiven Wirkung die Zielerreichung des Wettbewerbsrechts (und damit des Netzwerkzwecks) allgemein befördert. Jede Erleichterung wettbewerbsrechtlicher Schadenersatzklagen führt prima facie zu einem verbesserten Schutz der Wettbewerbsordnung, da sie potenzielle Rechtsverletzer abschreckt.[38]

Die Geschädigten mussten solche Schadensersatzansprüche zunächst mangels entsprechender unionsrechtlicher Vorgaben im Sinne des Au-

35 Grünbuch Schadenersatzklagen wegen Verletzung des EU-Wettbewerbsrechts, KOM (2005) 672 endg. vom 19.12.2005.

36 Hierzu *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 74ff.

37 EuGH, Urt. v. 20.9.2001 – Rs. C-453/99 (Courage/Crehan) – Slg. 2001, I 6297 Rn. 25ff.

38 EuGH, Urt. v. 20.9.2001 – Rs. C-453/99 (Courage/Crehan) – Slg. 2001, I 6297 Rn. 27.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tonomieprinzips auf der Grundlage nationalen Zivilverfahrens- und Deliktsrechts geltend machen.[39] Insoweit bot sich vor dem Austritt des Vereinigten Königreichs insbesondere aufgrund seiner vorprozessualen discovery-Regelungen die dortige Gerichtsbarkeit an,[40] während die deutschen Gerichte aus verschiedenen Gründen lange mit solchen Klagen fremdelten, da hier das prozessuale und das materielle Recht traditionell für den Kläger in solchen komplexen Zusammenhängen nicht günstig ist.

2. Herausforderungen für das Private Enforcement

Trotz der von Kommission und EuGH geäußerten Erwartung waren wichtige Aspekte der Durchsetzung lange ungeklärt, was die Entwicklung des Private Enforcement hemmte.

Die privatrechtliche Durchsetzung des Wettbewerbsrechts durch Geltendmachung deliktischer Schadensersatzansprüche setzt eine gewisse Risikobereitschaft der Geschädigten voraus.[41] Wenn diese Verbraucher sind, bekommen sie es mit übermächtigen Gegnern zu tun. Handelt es sich um Unternehmen, möchten sie nicht durch die Geltendmachung von Schadensersatzansprüchen einen Schatten auf andauernde Geschäftsbeziehungen werfen. Unabhängig von diesen eher atmosphärischen Aspekten erleiden gerade einzelne Verbraucher nur einen geringen (und kaum dokumentierten) Schaden, wenn die Wettbewerbsbeschränkung durch Preisverabredungen über Bier, Fußballtrikots oder Kaffee[42] erfolgt („Streuschäden"[43]). Die Durchsetzungsbereitschaft der Betroffenen hängt nicht zuletzt von einem kräftigen finanziellen Anreiz sowie von einer hohen Obsiegenswahrscheinlichkeit ab. Diese wird nicht nur von der materiellen Rechtslage, sondern auch von den prozessualen Verwirklichungsbedingungen bestimmt. Verbraucher haben dann auch kaum einen

39 So auch noch Art. 4 RL 2014/104/EU.

40 *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 66 f.

41 Zu den einzelnen Aspekten, die den Anspruch begründen sowie deren Durchsetzung: *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 28 ff.

42 BGH v. 13.7.2020 – KRB 99/19, NJW 2021, 395 (Bierkartell); OLG Düsseldorf v. 28.2.2018 – 4 Kart 3/17 OWi, BeckRS 2018, 39581 (Kartell der Kaffeeröster); UK OFT v. 1.8.2003, Case CP/0871/01 (Fußballtrikots; abrufbar auf gov.uk).

43 *Mestmäcker/Schweitzer*, Europäisches Wettbewerbsrecht, 3. Aufl. 2014, § 23 Rn. 60.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Anreiz, sich in einem teuren, aufwendigen und nicht zuletzt risikobehafteten Verfahren selbst um einen (geringen) Schadensersatz zu bemühen.

Die Kommission empfahl[44] den Mitgliedstaaten zunächst Möglichkeiten für einen kollektiven Rechtsschutz bereitzustellen, dabei aber das in den USA durch die Kombination von treble damages, „Opt-out"-Prinzip und besonderen Prozesskostenregelungen zu beobachtende Missbrauchspotential auszuschalten.[45] Die Kombination aus multipliziertem Schadensersatz zwecks Abschreckung (treble damages) und dem gruppenvergrößernden Prinzip, dass diejenigen automatisch Teil der im kollektiven Rechtsschutzverfahren vertretenen Gruppe werden, die einen Schaden aufgrund desselben oder eines ähnlichen Wettbewerbsverstoßes geltend machen und nicht ausdrücklich aus der Gruppe ausgeschieden sind („Opt-out"-Prinzip), birgt die Gefahr übermäßiger Haftung.[46] Sie bietet zudem Missbrauchsanreize in Form zweifelhafter oder überhöhter Klageforderungen, die zwecks Risikominderung durch beklagte Unternehmen im Wege kostenintensiver außergerichtlicher Vergleiche abgewendet werden.[47]

Diese Fehlanreize will die Empfehlung der Kommission vermeiden. In Deutschland ist sie mit dem Gesetz zur Einführung einer zivilprozessualen Musterfeststellungsklage im Jahr 2018 umgesetzt worden.[48] Seitdem ist die Musterfeststellungsklage für die Geltendmachung kartellrechtlicher Schadensersatzansprüche durch geschädigte Verbraucher zugänglich,[49] auch wenn ihre Eignung, die individuelle Betroffenheit und den individuellen Schadenseintritt verbindlich zu klären, in Zweifel gezogen wird.[50] Sie ermöglicht es eingetragenen Verbraucherschutzverbänden, zentrale Voraussetzungen eines Schadensersatzanspruchs zugunsten

44 Kommission, Empfehlung v. 11. 6. 2013: Gemeinsame Grundsätze für kollektive Unterlassungs- und Schadensersatzverfahren in den Mitgliedstaaten bei Verletzung von durch Unionsrecht garantierten Rechten, ABl. 2013 L 201/60; Siehe besonders auch Kommission, Mitteilung: Auf dem Weg zu einem allgemeinen europäischen Rahmen für den kollektiven Rechtsschutz v. 11.6.2013, COM (2013) 401 final.

45 *Mestmäcker/Schweitzer*, Europäisches Wettbewerbsrecht, 3. Aufl. 2014, § 23 Rn. 62.

46 Siehe i.e. *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 50 ff.

47 Vgl. Kommission, Mitteilung: Auf dem Weg zu einem allgemeinen europäischen Rahmen für den kollektiven Rechtsschutz v. 11.6.2013, COM (2013) 401 final, 3.

48 Gesetz vom 12.7.2018 – BGBl. I 2018, Nr. 26, S. 1151; BT Drs. 19/2507, S. 1.

49 *Fuchs*, in: Fuchs/Weitbrecht (Hrsg.), PrivKartellRDurchs.-HdB, 1. Aufl. 2019, § 1 Rn. 11.

50 *Mallmann/Erne*, NZKart 2019, 77.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

der Verbraucher feststellen zu lassen. Das Schutzsystem der überindividuellen Rechtsbehelfe wurde in Umsetzung der RL (EU) 2020/1828 über Verbandsklagen zum Schutz der Kollektivinteressen der Verbraucher erneut erheblich erweitert. Heute findet sich eine einheitliche Kodifikation von Abhilfe- und Feststellungsklagen der Verbraucherschutzverbände im Verbraucherrechtedurchsetzungsgesetz (VDuG).[51]

Die materielle Anspruchsgrundlage für die Geltendmachung deliktischer Ansprüche, die Schäden aus wettbewerbswidrigem Verhalten abgelten soll, ist in Deutschland nunmehr § 33a GWB. Um den abschreckenden, präventiven Zweck der Haftung[52] zu verstärken, verzichtet § 33a GWB auf die einschränkende Voraussetzung des § 823 Abs. 2 BGB.[53] Die Schwierigkeiten in der Umsetzung dieser Schadensersatzansprüche liegen aber vor allem in der zunehmenden Zersplitterung der Schäden auf unteren Vertriebsebenen und dem zunehmend anspruchsvolleren Kausalitätsnachweis.[54]

§ 33 Abs. 4 GWB räumt zudem bestimmten Verbänden Unterlassungs- und Beseitigungs-, jedoch keine Schadensersatzansprüche ein. Diese Verbände können nach § 34a GWB bei vorsätzlichen Kartellverstößen die rechtswidrig erlangten wirtschaftlichen Vorteile zugunsten des Bundeshaushalts abschöpfen, sofern dies nicht bereits durch die Kartellbehörde (§ 34 GWB) oder Schadensersatzleistungen der Unternehmen (vgl. § 34a Abs. 2 GWB) geschehen ist.

3. Mit- und Gegeneinander von Public und Private Enforcement

Die genannten Herausforderungen im Rahmen des Private Enforcement können zum Teil durch ein Zusammenwirken mit der behördlichen Durchsetzung des Wettbewerbsrechts bewältigt werden.

Im deutschen Zivilprozessrecht liegen Darlegungs- und Beweislast für anspruchsbegründende Tatsachen bei dem Kläger.[55] Hiervon sind insbesondere das Vorliegen einer Vereinbarung, deren Wirkung auf den Wettbewerb und der kausale Schadenseintritt betroffen. Diese Informa-

51 Gesetz vom 8.10.2023 – BGBl. I 2023, Nr. 272, S. 2.

52 *Mestmäcker/Schweitzer*, Europäisches Wettbewerbsrecht, 3. Aufl. 2014, § 23 Rn. 62.

53 Zur Entwicklung Immenga/Mestmäcker/*Franck*, 7. Aufl. 2024, GWB § 33a Rn. 3.

54 *Mestmäcker/Schweitzer*, Europäisches Wettbewerbsrecht, 3. Aufl. 2014, § 23 Rn. 37.

55 *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 63 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tionen sind nicht ohne weiteres zugänglich – geschweige denn, dass sie offen zutage liegen. Aus diesem Grunde ist es für den Geschädigten sinnvoll, mit der gerichtlichen Geltendmachung von Schadensersatzansprüchen so lange abzuwarten, bis ein Wettbewerbsverstoß im Wege des Public Enforcement festgestellt worden ist – zumal jedenfalls der Verbraucher von der Existenz einer wettbewerbswidrigen Abrede regelmäßig nicht einmal etwas ahnt. Im Rahmen des behördlichen Verfahrens, das auf die Unterbindung des Verstoßes und die Erteilung einer Sanktion abzielt, trägt indes die Behörde mit den ihr zur Verfügung stehenden hoheitlichen Mitteln die für das erfolgreiche Private Enforcement benötigten Informationen zusammen. Insoweit besteht dann die Notwendigkeit, den Aktenzugang der Geschädigten zu den behördlichen Informationen mit einem angemessenen Ausgleich zwischen deren Interessen und dem ebenfalls legitimen Interesse an dem Schutz von Geschäftsgeheimnissen zu ermöglichen.[56]

Das Public Enforcement wird allerdings häufig erst durch die Zuarbeit eines Kronzeugen angestoßen.[57] Denn weil die Teilnehmer in einem Kartell nicht nur wirtschaftliche Interessen an dessen Aufrechterhaltung haben, sondern auch Sanktionen fürchten, werden diese Abreden institutionell so eingebettet, dass sie nur schwer „von außen" zu entdecken sind. Um diese Konspiration zu schwächen, gibt es Kronzeugenregelungen (vgl. § 81h ff. GWB und Art. 4a VO 773/2004). Diese wirken sowohl präventiv als auch repressiv.[58] Einerseits bedrohen sie durch das ihnen zugrundeliegende Windhundprinzip die innere Stabilität des Kartells und damit auch das wechselseitige Vertrauen seiner Teilnehmer. So machen sie wettbewerbswidrige Abreden ein wenig unattraktiver. Andererseits ermöglichen sie dann aber nach Aufdeckung des Kartells auch die effektivere Verfolgung der übrigen Mitglieder, weil sich der Kronzeuge seine Immunität nur mit der umfassenden Offenlegung aller Informationen über das Kartell verdient.

56 Hierzu Vgl. EuGH, Urt. v. 14.6.2011, Rs C-360/09 (Pfleider AG/Bundeskartellamt), Slg 2011 I-5161, Rn. 30f.; EuGH, Urt. v. 6.6.2013, Rs C-536/11 (Bundeswettbewerbsbehörde/Donau Chemie AG ua), Rn. 30ff.

57 *von Schreitter/Matt*, NZKart 2024, 144.

58 Die beiden folgenden Aspekte finden sich bei Immenga/Mestmäcker/*Bueren*, 7. Aufl. 2024, GWB § 81h Rn. 4; s. a. Immenga/Mestmäcker/*Biermann*, 7. Aufl. 2025, VO (EG) 1/2003 Art. 23 Rn. 232ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die Angewiesenheit des Public Enforcement auf die Mitwirkung von Kronzeugen führt nun in einen Teufelskreis:[59] Das wirksame Private Enforcement ist auf die vorgelagerte behördliche Aufarbeitung und Verfolgung eines Rechtsverstoßes angewiesen. Für die Einleitung eines solchen Verfahrens bedarf die Behörde wiederum der Mitarbeit eines an der Vereinbarung beteiligten Unternehmens. Dessen Motivation beruht darauf, dass es einerseits schon von der wettbewerbswidrigen Vereinbarung profitiert hat und andererseits als Kronzeuge – anders als die übrigen Wettbewerber – von behördlichen Sanktionen verschont bleibt. Drohen allerdings empfindliche Schadensersatzansprüche im Wege des Private Enforcement, könnte sich dies auf die grundsätzliche Bereitschaft, jedenfalls aber auf den Zeitpunkt auswirken, zu dem ein an der wettbewerbswidrigen Vereinbarung Beteiligter durch seine Meldung als Kronzeuge die Einleitung eines Verfahrens ermöglicht. Je einfacher den Geschädigten also die zivilrechtliche Durchsetzung von Schadensersatzansprüchen gemacht wird, desto weniger werden Unternehmen bereit sein, als Kronzeugen in Erscheinung zu treten.[60] Tendenziell wachsen daher die durch das Kartell entstehenden Wohlfahrtsverluste.

4. Antwort des Unionsgesetzgebers

Damit hatte der Unionsgesetzgeber einen gordischen Knoten zu durchschlagen: Auf der einen Seite sollte eine Verbesserung und Vereinheitlichung des Private Enforcement der Wettbewerbsordnung zu einer besseren Durchsetzung verhelfen. Auf der anderen Seite würde jede Verbesserung und Vereinheitlichung die behördliche Durchsetzung des Wettbewerbsrechts erschweren, auf die das Private Enforcement wiederum angewiesen ist.

Der Unionsgesetzgeber erließ zur Vereinheitlichung der nationalen Regelungen für die Durchsetzung privatrechtlicher Schadensersatzklagen die Richtlinie 2014/104/EU. Deren Ziel war es, die Wirksamkeit der Art. 101 f. AEUV sicherzustellen und den Schutz einzelner Geschädigter zu gewährleisten.[61] Art. 3, 12 RL 2014/104/EU sehen einen umfassenden

59 Hierzu ausf. *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 173 f.

60 *Kammin*, Reforming Private Antitrust Enforcement in Europe: Between Harmonisation and Regulatory Competition, 2014, 177.

61 Erwägungsgründe 1, 3 RL 2014/104/EU.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Schadensersatzanspruch für „jede natürliche oder juristische Person" vor, „die einen durch eine Zuwiderhandlung gegen das Wettbewerbsrecht verursachten Schaden erlitten hat". Der Anspruch ist auf eine Totalreparation gerichtet und umfasst Ersatz der Vermögenseinbußen, des entgangenen Gewinns und Zinsen (Art. 3 Abs. 2 RL 2014/104/EU). Eine Verpflichtung zur Einführung von Sammelklagen begründet die Richtlinie indes nicht. Allerdings kann der Grundsatz der Effektivität eine durchsetzungsfreundliche Auslegung des nationalen Rechts gebieten, wenn ansonsten die mitgliedstaatlichen Regeln entsprechende Hilfskonstruktionen (etwa Sammelzessionen) und dadurch eine wirksame Durchsetzung von Schadensersatzansprüchen v. a. für Streuschäden unmöglich machen.[62]

Ein zentrales Anliegen der Richtlinie war die bessere Abstimmung zivilrechtlicher Durchsetzung des Wettbewerbsrechts mit ihrem hoheitlichen Pendant. Auch hier sind Informationsregelungen wieder von besonderer Bedeutung; so etwa die Vorschriften über den Zugang zu Unterlagen der Wettbewerbsbehörden.[63]

Um das Informationsdefizit des privaten Klägers zu neutralisieren, sieht die Richtlinie auch insoweit eine Verzahnung des Private mit dem Public Enforcement vor.[64] Nach Art. 5, 6 RL 2014/104/EU können nicht nur der Beklagte, sondern auch Dritte zur Offenlegung von Beweismitteln gerichtlich verpflichtet werden. Dies umfasst nach Maßgabe von Art. 6 RL 2014/104/EU auch Akten von Wettbewerbsbehörden. Allerdings genießen Kronzeugenerklärungen einen besonderen Schutz. Einsicht in sie ist für die Zwecke der privaten Rechtsverfolgung weitgehend ausgeschlossen (Art. 6 Abs. 6 RL 2014/104/EU). Beweise, die unter Verstoß gegen diese Regelung offengelegt wurden, dürfen nicht verwertet werden (Art. 7 Abs. 1 RL 2014/104/EU). Hier wird der besondere Wert dieser Aussagen für die behördliche Verfolgung deutlich, die den Schutz der Kronzeugen – auch zu Lasten des Private Enforcement – erforderlich macht.

Um das Public Enforcement nicht zu stark zu beeinträchtigen, werden Kronzeugen auch durch formelle und materielle Privilegierung im Haftungsrecht geschützt.[65] Die Richtlinie ergänzt den wirtschaftlichen Schutz des Kronzeugen gegen die von der Kommission auferlegten Geld-

62 EuGH, Urt. v. 28.1.2025 – Rs. C-253/23 (ASG 2) – ECLI:EU:C:2025:40, Rn. 69 f.; s. a. für Deutschland *Riehm*, NJW 2024, 2713.

63 Erwägungsgrund 6 zur RL 2014/104/EU.

64 *Keßler,* VuR 2015, 83 (87).

65 *Keßler,* VuR 2015, 83 (84).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

bußen durch zumindest eine Privilegierung auch bei der zivilrechtlichen gesamtschuldnerischen Haftung (vgl. Art. 11 Abs. 4 RL 2014/104/EU). Zu einer umfassenden deliktsrechtlichen Immunität konnte sich der Unionsgesetzgeber hingegen nicht durchringen.

Aus diesen Regelungen und aus dem starken Schutz der Kronzeugen ergibt sich allerdings auch, dass nach der unionsrechtlichen Konzeption die öffentlich-rechtliche Durchsetzung des Wettbewerbsrechts immer noch Vorrang vor dem Private Enforcement genießt.[66] Dieses rechtssystematische Verhältnis ist aus der Perspektive einer effektiven Rechtsdurchsetzung sinnvoll. Denn nur dadurch, dass ein Kronzeuge das Kartell publik macht, wird regelmäßig überhaupt erst ein Anknüpfungspunkt für zivilrechtliche Schadensersatzklagen eröffnet. Die Beobachtung von rückläufigen Kronzeugenanträgen führt in Deutschland zu der berechtigten Diskussion, ob eine weitergehende Privilegierung der Kronzeugen im Haftungsrecht oder etwa auch im Vergaberecht geboten ist.[67]

Das in Teilen aufgrund mangelnder Abstimmung zu beobachtende dysfunktionale Nebeneinander von behördlicher Durchsetzung und Private Enforcement erodiert indes nicht die Zweckwidmung des Netzwerks oder weckt Zweifel an der Netzwerkstruktur, sondern stellt vielmehr seine Kohärenz sowie seine interne Funktionalität in Frage.

F. Ein hybrides Netzwerk?

Ergibt sich nun aber angesichts dieser Organisations- und Interaktionsstrukturen bei der Durchsetzung des Wettbewerbsrechts tatsächlich ein gewisser Mehrwert aus der (gesetzlichen) Charakterisierung der Kooperation als „Netzwerk" – nicht mehr nur „der Behörden", sondern auch mit der Integration Privater?

Der Begriff des Netzwerks beschreibt in der Verwaltungsorganisation i. w. S. die Aufhebung von Gegensätzen, die eigentlich der Gewährleistung einer klaren Handlungszurechnung als Grundlage für Kontrolle, Rechtsschutz und Legitimation dienen.[68] Selbst wenn eine bestimmte Handlung

66 *Mestmäcker/Schweitzer*, Europäisches Wettbewerbsrecht, 3. Aufl. 2014, § 23 Rn. 59l.

67 Zu den tatsächlichen Grundlagen s. Monopolkommission, Wettbewerb 2022, XXIV. Hauptgutachten, Rn. 305 ff.; ausführliche Diskussion bei *von Schreitter/Matt*, NZKart 2024, 144 (145 ff.).

68 *Schwind*, Netzwerke im Europäischen Verwaltungsrecht, 2017, 132.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

einem Akteur formal eindeutig zuzurechnen ist, haben oft andere Akteure des Netzwerks im Vorfeld der Entscheidungsfindung einen so starken materiellen Einfluss auf das nach außen tretende Ergebnis genommen oder nehmen können, dass die formale Zuordnung nur vordergründig Sinn ergibt. Verlangt man von einem Netzwerk eine heterarchische Struktur, lässt sich diese in dem wettbewerbsrechtlichen „Netzwerk der Behörden" nicht nachweisen. Zwar mag es hier auch formellen wie informellen Informationsaustausch auf Augenhöhe geben. Aber die Belange der hoheitlichen Durchsetzung des Wettbewerbsrechts genießen Vorrang vor der Effektivierung des Private Enforcement und die Befugnisse der Kommission gehen denen der mitgliedstaatlichen Wettbewerbsbehörden vor. Dies sind erkennbar hierarchische Strukturen.

Sowohl das Netzwerk der Behörden als auch das durch Hinzutreten von Unternehmen und Vereinigungen erweiterte Netzwerk zur Anwendung und Durchsetzung des Wettbewerbsrechts erlaubt überdies eine klare Zuordnung einzelner Handlungen. Es gibt zwar Kooperation und Informationsfluss, aber sowohl die behördlichen Funktionsträger als auch die beteiligten Privaten bestimmen ihre Beiträge zur Erreichung des Netzwerkzwecks selbstständig. Insbesondere auf der staatlichen Seite erlauben die Regelungen von Zuständigkeit und Selbsteintritt eine klare Zuweisung der Verantwortung für getroffene Entscheidungen. Die Entscheidungen über Sanktionen oder auch über die Geltendmachung von Schadensersatz- bzw. Unterlassungsansprüchen sind keine nach außen tretenden „Verbundprodukte", deren Zuordnung zu einem einzelnen Akteur zweifelhaft ist. Insgesamt bleibt auch in der Kooperation die jeweilige Einordnung der Akteure in unterschiedliche Funktionsbereiche erkennbar; es erfolgt keine Entdifferenzierung von öffentlichen und privaten Akteuren.

Dies alles zeigt, dass Vorsicht geboten ist. Das Recht der Verwaltungsorganisation, die Modalitäten privaten/staatlichen Zusammenwirkens, die Formulierung und Durchsetzung des Gemeinwohls funktionieren nicht mehr wie vor 100 Jahren, weil die realen Wirkungsbedingungen des Handelns immer komplexer und immer schneller geworden sind. Deswegen ist es einerseits sinnvoll, sich neuen Organisations- und Handlungsmodalitäten zu öffnen, die auf diese Veränderungen reagieren können. Andererseits nimmt es diesen Neuigkeiten auch ihre Funktionalität, wenn sie zu unkritisch verwendet werden. Wenn alles und jede Form der institutionalisierten Kooperation ein „Netz" ist, dann illustriert man zwar einerseits bestimmte äußerlich erkennbare Kooperationsstrukturen,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

beraubt sich aber andererseits des spezifischen Erklärungswerts des Netzwerks. Dann verliert der Begriff an Kontur, Erklärungswert und Überzeugungskraft, weil er letztlich nur eine institutionalisierte Kooperation zur Verfolgung eines übergeordneten Zwecks beschreibt.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Governing the Asphalt Commons. Recht und Akzeptanz der Parkraumbepreisung

Konstantin Chatziathanasiou & Christian Traxler

A. Einleitung

Für *Christoph Engel* ist das Recht ein Instrument sozialer Ordnung. Die Rechtswissenschaft versteht er dabei als angewandte Sozialwissenschaft.[1] Natürlich hat der Ansatz Grenzen.[2] Gerade das spricht dafür, das Recht als sozialwissenschaftlichen Forschungsgegenstand nicht einfach den Ökonomen, Soziologen oder Psychologen zu überlassen. Zum einen stellen diese nur selten die Fragen, die die Juristen beschäftigen. Zum anderen muss auch die Bewertung vorliegender Evidenzen für die – sorgfältig zu rekonstruierenden – normativen Belange des Rechts sensibel sein.[3] Sozialwissenschaftlich gerüstet kann die Rechtswissenschaft dann eher dazu beitragen, dem Steuerungsanpruch des Rechts tatsächliche Geltung zu verleihen: durch verhaltenswissenschaftlich fundierte Vorschläge, die im Konzert konkurrierender Expertisen nicht untergehen.

Christoph Engel hat seinen Ansatz in der Frühphase seines Instituts am Alltagsproblem des Abfalls exemplifiziert.[4] Die damit verbundenen regulatorischen Probleme lassen sich mit Hilfe der Sozialwissenschaften präzise beschreiben und unterscheiden. Die Eingrenzung des normativen Problems, das aus juristischer Sicht als wesentlich gelten muss, erlaubt dann den nächsten Schritt: die Anwendung sozialwissenschaftlicher Methoden unter Wahrung der jeweiligen disziplinären Standards. Das Referenzgebiet des Abfallrechts war mit Bedacht gewählt. Schon aufgrund des expliziten Steuerungswillens des Gesetzgebers lag die Rezeption ökomischer Methoden nahe – und der Vorwurf des „methodischen Imperialismus'" fern. Seither ist *Christoph Engel* nicht müde geworden, auf

1 Engel 1998; 2024.
2 Engel 2007; 2008.
3 Engel 2018.
4 Engel 2002.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

weitere Rechtsgebiete und Fragestellungen hinzuweisen, deren interdisziplinäre Aufarbeitung lohnen würde.[5] In unserem Beitrag folgen wir dieser Stoßrichtung und widmen uns einer Thematik, die von einer verstärkten interdiszplinären Bearbeitung profitieren würde: der Regulierung des Straßenverkehrs und insbesondere des Parkens.

Die Regulierung des Straßenverkehrs zielt darauf ab, die Nutzung knapper Ressourcen in einem Umfeld zu lenken, das von zahlreichen Externalitäten geprägt ist. Die Gemeingutproblematik begegnet uns hier auf mehreren Ebenen. Unser Beitrag legt den Fokus auf das Parken von Pkw, das die strukturelle Übernutzung des öffentlichen Straßenraums veranschaulicht (B). Der aktuelle Rechtsrahmen hilft bei der Problemlösung nur bedingt.[6] Soziale Normen – das gefühlte Recht auf „freies Parken" – und die tatsächliche rechtliche Privilegierung des Dauerparkens vertiefen das Problem sogar (C). Die Gemeinden setzen auf Parkraumbewirtschaftung, stoßen aber auf rechtliche Grenzen (D). Wie die Ergebnisse aus einem Survey-Experiment nahelegen, besteht hier Raum, um mit verhaltenswissenschaftlich informierten Kommunikations- und Informationsstrategien starre Positionen zu überwinden (E). Dies könnte die Basis für eine angemessene Verkehrsrechtspolitik schaffen (F).

B. Das Problem der *asphalt commons*

Der Straßenverkehr stellt sich auf vielfältige Art als Gemeingutproblematik dar. Verkehr produziert negative Externalitäten, die Gemeingüter betreffen, etwa Lärm- und Schadstoff-Emissionen. Unter den Verkehrsteilnehmern stellen sich damit vielfältige Kooperationsprobleme: Wenn Einzelne durch überhöhte Geschwindigkeit Zeit zu gewinnen suchen, geschieht dies zulasten der Sicherheit anderer;[7] wenn öffentliche Straßen übermäßig beansprucht werden, entstehen Staus. Risiken der Übernutzung betreffen aber nicht nur den fließenden Verkehr. Parkende Pkws beanspruchen erhebliche Teile des öffentlichen Raums, der damit für alternative Nutzungen nicht zur Verfügung steht.[8] Sie beeinträchtigen Sichtachsen und mindern so die Sicherheit der übrigen Verkehrs-

5 Engel 2020b.
6 Die Ausführungen zum Rechtsrahmen schöpfen aus: Chatziathanasiou 2025.
7 Bauernschuster & Traxler 2021.
8 Gössling u. a. 2016; Szell 2018; Creutzig u. a. 2020.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

teilnehmer. Hinzu kommt, dass private Pkw den Großteil des Tages ungenutzt bleiben – in Deutschland durchschnittlich über 23 Stunden.[9] Dies geht auch zulasten jener, die lediglich für kurze Zeit parken möchten: Durchschnittlich 40 bis 70 Stunden pro Jahr verbringen Autofahrer in deutschen Großstädten mit der Parkplatzsuche[10] – ein Problem, das bekanntlich schon besungen wurde.[11] In jüngerer Zeit wird zudem der Einfluss parkender Fahrzeuge auf das Mikroklima diskutiert, da sie als Wärmespeicher wirken können.[12]

Es gehört zu den Kerneinsichten der Verkehrsforschung, dass mehr Parkplätze – ebenso wie mehr Straßen – mehr Verkehr und mehr private Autos nachsichziehen.[13] Das gilt erst recht, wenn Parkplätze nicht oder sehr niedrig bepreist werden. In den Worten von *Donald Shoup*: „Free curb parking is an asphalt commons: just as grazing cattle compete in their search for scarce grass, drivers compete in their search for scarce curb parking spaces. Drivers waste time and fuel, congest traffic, and pollute the air while cruising for curb parking, and after finding a space, they have no incentive to economize on how long they park."[14]

Aus ökonomischer Sicht sollte diesem Steuerungsproblem mit Preisen begegnet werden. Preise können zu einer Internalisierung der sozialen Kosten des Parkens führen.[15] Dabei gibt es verschiedene Wege, Kostenwahrheit zu erreichen. Nach *William Vickrey* sollte die Nutzung der Straße nachfrageorientiert, in Abhängigkeit vom Ort und Tageszeit, bepreist werden.[16] Hieran anknüpfend hat sich ein theoretischer Literaturstrang für eine dynamische Bepreisung des Bordsteinparkens stark gemacht.[17] Auch die empirische Mobilitäts- und Verkehrsforschung spricht dafür, dass eine angemessene Bepreisung des Bordsteinparkens die Parkplatzsuche in Stadtzentren erleichtert.[18] Im Zeitalter der Digitalisierung erscheint eine dynamische Bepreisung der Straßennutzung, die sich an

9 Canzler 2024: 4; Nobis & Kuhnimhof 2018.
10 Agora Verkehrswende 2022.
11 Grönemeyer 1984: „Du sitzt bei Kaffee und Kuchen, und ich muss weiter suchen".
12 Matias u. a. 2025.
13 Topp 1993: 83; Yin u. a. 2018.
14 Shoup 2004: 754.
15 Baumol 1972; Inci 2015.
16 Vickrey 1969.
17 Verhoef u. a. 1995; Arnott & Inci 2006; Shoup 2004; 2011.
18 Ottosson u. a. 2013; Cats u. a. 2016; Ostermeijer u. a. 2022.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

der tatsächlichen Nachfrage orientiert und gleichzeitig variable soziale Kosten berücksichtigt, grundsätzlich auch realisierbar.

C. Soziale Norm und rechtliches Privileg

Was aus ökonomischer und verkehrswissenschaftlicher Sicht naheliegend erscheint, stößt bei der Implementierung jedoch auf starke Widerstände. Schon allgemein ist eine Bepreisung der Straßennutzung in Deutschland unbeliebt.[19] Nicht nur in jenen Generationen, die mit dem populären Brettspiel Monopoly und dem Feld „Frei Parken" aufgewachsen sind, geht die Abneigung tief. Die Verkehrsforschung diagnostiziert, zu viele Menschen verspürten, dass ihnen ein Recht auf Straße im Allgemeinen – und ein Recht auf Parken im Besonderen – zustände.[20] Im Bereich des Parkens wird diese soziale Norm durch bestimmte rechtliche Weichenstellungen gar stabilisiert.

Nutzungsintensives Dauerparken ist rechtlich privilegiert. Die Straßengesetze der Länder unterscheiden zwischen Gemeingebrauch und Sondernutzung. Gemeingebrauch ist der Gebrauch der öffentlichen Straßen im Rahmen ihrer Widmung für den Verkehr. Eine Sondernutzung ist eine Nutzung zu vorwiegend anderen Zwecken.[21] Letztere ist genehmigungspflichtig und kann mit Auflagen verbunden werden.[22] Die Rechtsprechung schlägt parkende Fahrzeuge dem Gemeingebrauch zu – es handele sich um sog. *ruhenden Verkehr*.[23] Die Länder können hier keine Ausnahmen treffen, der Bund hat den Begriff des Verkehrs abschließend geregelt.[24] Ist ein parkendes Fahrzeug zugelassen und betriebsbereit, wird es zum Verkehr gezählt. Etwas anderes gilt nur dann, wenn das Fahrzeug nach objektiver Sichtweise zu anderen Zwecken aufgestellt wurde, etwa zur Werbung[25] oder als Verkaufswagen[26]. Die objektive Sichtweise ist sogar maßgeblich, wenn es um geparkte Mietwagen geht, die zu kommer-

19 Traxler & Wegrich 2023.
20 Taylor 2021: „folk legality"; Gössling u. a. 2022: 250: „entitlement"; Canzler 2024: 6.
21 § 14 Abs. 1 Satz 1 bzw. § 14 Abs. 3 Satz 1 StrWG NRW.
22 § 18 Abs. 2 Satz 2 StrWG NRW.
23 BVerfGE 67, 299 (315 ff) [1984].
24 BVerfGE 67, 299 (320 ff) [1984]: „Vorbehalt des Straßenverkehrsrechts".
25 OVG NRW, Urt. v. 17.9.2020, Az.: 11 A 2961/19.
26 BVerwGE 80, 36 (42) [1988].

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ziellen Zwecken vorgehalten werden.[27] Solange ein Fahrzeug also dem allgemein üblichen Straßenbild entspricht, wird ausgeblendet, welche Lasten dem öffentlichen Raum und den anderen Verkehrsteilnehmern aufgebürdet werden. Um beim Fall der geparkten Mietwagen zu bleiben: Der Betreiber der Mietwagenfirma, der den öffentlichen Raum für das Abstellen seiner Fahrzeuge nutzen kann, hat keinen Anreiz den öffentlichen Raum zu schonen.

D. Parkraumbewirtschaftung

In Anbetracht des fehlenden politischen Willens, hier etwas zu verändern, liegen die Hoffnungen der Gemeinden auf der Parkraumbewirtschaftung. Dies betrifft das Kurzzeit-, aber auch das Anwohnerparken. Das Anwohnerparken stellt durch die dauerhafte Raumverknappung eine besonders intensive Belastung für die Städte dar.[28] Vereinzelt ist versucht worden, hier stärker zu steuern. So führte die Stadt Freiburg für das Anwohnerparken – auch im Sinne des Klimaschutzes – eine Gebührensatzung ein, die die Gebühren nicht nur erheblich erhöhte, sondern auch nach der Länge des jeweiligen Pkw differenzierte.[29] Die Normalgebühr für ein Jahr sollte 360 € betragen, sich für Fahrzeuge unter 4,21 m Länge auf 240 € reduzieren und sich für solche über 4,70 m Länge auf 480 € erhöhen.[30] Während der VGH Mannheim diese Regelungen als noch von der gesetzlichen Gebührenermächtigung gedeckt ansah, sah das Bundesverwaltungsgericht diese Grenze als überschritten an.

Tatsächlich erwähnt die Ermächtigungsnorm den Klimaschutz nicht.[31] Der VGH Mannheim ließ sich aber überzeugen, dass die darin genannten Gesichtspunkte zur Gebührenbemessung nicht abschließend seien.[32] Die „Erreichung des staatlichen Klimaschutzziels des Art. 20a GG und der Schutz von Grundrechten vor den Gefahren des Klimawan-

27 BVerwG, Urt. v. 3.6.1982, Az.: 7 C 73/79.

28 van Ommeren u. a. 2011: 30.

29 Satzung der Stadt Freiburg über die Erhebung von Bewohnerparkgebühren vom 14.12.2021.

30 Zum Beispiel würde ein klassischer Honda Civic mit 4,14 m Länge noch knapp in die reduzierte Gebührenkategorie fallen.

31 § 6a Abs. 5a S. 2 bis 5 StVG in Verbindung mit der Parkgebührenverordnung.

32 Unter Berufung auf u. a. BT-Drs. 19/19132, S. 10ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dels durch eine Reduktion des Kfz-Verkehrs und die Verringerung des hierdurch bedingten CO_2-Ausstoßes" seien zulässige Lenkungsziele, die bei der Bemessung der Bewohnerparkgebühren berücksichtigt werden durften.[33] Auch die Staffelung der Bewohnerparkgebühr sei nicht zu beanstanden. Es gehe um eine „möglichst effizient[e] Nutzung der im Innenstadtbereich stets knappen öffentlichen Parkplätze"; als Instrument der Verkehrssteuerung trage die Gebührenerhebung dazu bei, Parkplatzsuchverkehr zu minimieren und setze Anreize, auf andere Verkehrsmittel, wie ÖPNV oder Fahrrad, umzusteigen.[34] Der VGH verstand die Bewohnerparkgebühren schließlich als Gebühren für eine Flächennutzung.[35]

Das Bundesverwaltungsgericht setzte dem Freiburger Vorstoß jedoch Grenzen. Es erkannte in der Gebührenstaffelung, die sich nach der Länge der Autos richtete, eine Verletzung des allgemeinen Gleichheitssatzes (Art. 3 Abs. 1 GG). Die gestaffelten Gebühren in Höhe von 240 €, 360 € oder 480 € beeinträchtigen die „Gebührenschuldner in deutlich unterschiedlichem Ausmaß in ihrer durch Art. 2 Abs. 1 GG gewährleisteten persönlichen Freiheitsentfaltung im vermögensrechtlichen Bereich".[36] Die Sprünge zwischen den Gebührenstufen seien beträchtlich und diese Ungleichbehandlung sei nicht zu rechtfertigen. Die Länge des Fahrzeugs spiele gemessen am Vorteil, überhaupt parken zu können, eine untergeordnete Rolle und die erheblichen Sprünge zwischen den Gebühren stünden nicht im Verhältnis zur Bedeutung der Fahrzeuglänge. Vor allem aber ermächtige die Ermächtigungsgrundlage zum Erlass der Gebührenverordnung nicht zur Berücksichtigung klimaschutz- und gesundheitspolitischer Lenkungsziele.[37]

Rechtsdogmatisch verwundert das nicht. Auch in der Gebührenerhebung liegt eine Freiheitsbeschränkung, die einer klaren gesetzlichen Grundlage bedarf. Rechtspolitisch ist die Entscheidung jedoch bedauerlich, aus einer ökonomischen Perspektive sogar irritierend – zumal größere Autos größere (negative) Externalitäten produzieren. Sie verbrauchen mehr Raum und fordern bestehende Infrastrukturen hinaus: Dank *autobesity* sind moderne Pkws häufig übermäßig lang und breit.[38] Das

33 VGH BW, Urt. v. 13.7.2022, Az.: 2 S 808/22, Rn. 106.
34 VGH BW, Urt. v. 13.7.2022, Az.: 2 S 808/22, Rn. 107.
35 VGH BW, Urt. v. 13.7.2022, Az.: 2 S 808/22, Rn. 113.
36 BVerwG, Urt. v. 13.6.2023, Az.: 9 CN 2.22 -2 S 808/22, Rn. 71.
37 BVerwG, Urt. v. 13.6.2023, Az.: 9 CN 2.22 -2 S 808/22, Rn. 78.
38 Vgl. The Economist 2024.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Parken von größeren Fahrzeugen führt damit häufiger zu Behinderungen des fließenden Verkehrs. Ein Gebührenregime, das den Flächenverbrauch abbildet, erscheint aus dieser Perspektive geboten. Gebührensprünge in der Höhe von 10 € pro Monat scheinen dabei auch mit Blick auf die aktuelle Bußgeldkatalog-Verordnung (BKatV), die sich durch erhebliche Gebührensprünge auszeichnet, vertretbar.[39]

Mehr Flexibilität haben die Kommunen begrüßenswerterweise beim Kurzzeitparken. § 6a Abs. 6 S. 1 StVG erlaubt es, für das Parken im Innenstadtbereich Gebühren zu verlangen. Der Wortlaut ist offen. Er erlaubt es, Anreize zum Umstieg auf andere Verkehrsmittel zu setzen. Zudem „bietet die Ermächtigung Raum auch für eine digitale Abwicklung (etwa Smartphone-Bezahlsysteme) und sogar für an Verkehrsaufkommen und Uhrzeiten anknüpfende Differenzierungen".[40] Hier sind bloß etwaige Gebührenordnungen der Länder zu berücksichtigen, soweit sie Vorgaben zu Gebührenbemessung machen. Tatsächlich gehen die Städte mit dieser Freiheit unterschiedlich um. Das betrifft die Zahl und Größe der kostenpflichtigen Parkzonen, der gebührenpflichtigen Zeiträume und die Tarife.[41] Die Deutsche Umwelthilfe hat dokumentiert, dass die maximale Parkgebühr der meisten Städte zwischen 1,00 € und 2,50 € pro Stunde liegt.[42] Ein Wille zur verstärkten Anreizsteuerung ist angesichts dieser Werte nicht zu erkennen.

E. Normwandel

Hinter rechtlichen Privilegierungen, fehlenden Ermächtigungsgrundlagen und zögerlicher Anreizsetzung steht fehlender politischer Wille. Der fehlende politische Wille ist auf die Angst zurückzuführen, unpopuläre Maßnahmen zu ergreifen, die der sozialen Norm des freien Parkens zuwiderlaufen. Soziale Normen sind jedoch nicht unveränderlich. Sie können sich wandeln – gerade auch in Interaktion mit Recht. In unserem konkreten Fall stellen sich insbesondere zwei empirische Fragen: Wie unpopulär wäre eine Ausweitung der Parkraumbewirtschaftung? Und wie ließe sich

39 Traxler u. a. 2018.

40 Fehling 2024: 65.

41 Canzler 2024: 5.

42 https://www.duh.de/fileadmin/user_upload/download/Projektinformation/Verkehr/Anwohnerparken/250422_Kurzzeitparkgebühren_Infografik.pdf.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

die Akzeptanz höherer Parkgebühren steigern? Ein aktuelles Forschungsprojekt untersucht diese Fragen für den Fall Berlins mit einem Survey-Experiment.[43]

In Kooperation mit einem Online-Panel wurden rund 1.400 Berliner Erwachsene zu ihren Einstellungen zum Thema urbane Mobilität befragt.[44] Erfasst wurden dabei auch die Positionen zu den möglichen Erhöhungen der Gebühren für das Kurzzeitparken und für den Anwohnerparkausweis, die seit dem Jahr 2022 diskutiert werden. Die zwei zentralen Fragen lauteten:

(1) „Die Gebühren für Kurzzeitparken werden in Berlin, je nach Bezirk, um einen Euro auf 2, 3 bzw. 4 € pro Stunde erhöht. Wie stehen Sie zur Erhöhung der Kurzpark-Parkgebühren?"

(2) „Die Anwohner-Parkgebühren sollen in Berlin auf 120 € pro Jahr erhöht werden (dort, wo der Parkraum bereits jetzt gebührenpflichtig ist). Wie stehen Sie zur Erhöhung der Anwohner-Parkgebühren?"

Im Rahmen des Survey-Experiments dienten diese beiden Fragen als Baseline (*Kontrollgruppe*). Aufbauend auf ihren Formulierungen wurden weitere Versionen getestet. In einem Treatment-Arm wurde betont, dass die höheren Einnahmen in die Stärkung des öffentlichen Nahverkehrs investiert werden sollen (*Einnahmenrecycling*). In einer weiteren Variation wurde auf die hohen Gebühren in Städten wie Amsterdam oder Paris verwiesen (*Referenzwerte*).[45] Als zweite Dimension eines faktoriellen 2 x 3 Designs wurden Informationen über den *status quo* hinzugefügt: Bei der Frage zum Kurzzeitparken wurde die Stabilität der aktuellen Gebühren („Dies ist der erste Preisanstieg seit 18 Jahren."), bei der Frage zum Anwohnerparken ihr niedriges Niveau („…von derzeit etwa 10 € auf 120 € …") salient gemacht. Insgesamt gab es also je sechs Versionen der beiden Fragen. Die Befragten wurden über unabhängige Randomisierungen mit jeweils einer Version der ersten und einer Version der zweiten Frage konfrontiert.

Abbildungen 1 und 2 visualisieren die über alle Treatment-Variationen gepoolten Antworten zu den beiden Fragen. Die Ergebnisse doku-

43 Traxler 2025.

44 Es handelt sich hier um eine altersrepräsentative Convenience-Stichprobe, die die Bevölkerungsverteilung über die Berliner Bezirke abbildet.

45 Die Formulierungen lauteten: „Zum Vergleich: in Städten wie Amsterdam oder Paris liegen die Parkgebühren in Kurzparkzonen zwischen 4 und 7 €."; „Zum Vergleich: in Städten wie Amsterdam, Paris, oder Stockholm liegen die Parkgebühren für Anwohner über 500 € pro Jahr."

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

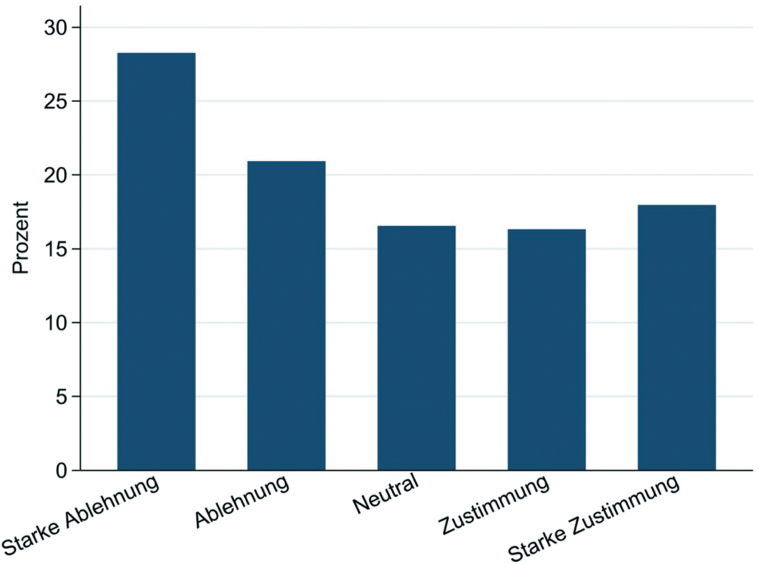


Abbildung 1 – Positionen zur Erhöhung der Gebühren in Kurzparkzonen

mentieren zunächst einen ausgeprägten Widerstand gegen höhere Parkgebühren. Während 33 % der Befragten einer Erhöhung der Kurzpark-Gebühren um 1 € zustimmen, lehnen dies 50 % der Befragten ab (29 % der Befragten gar stark). Bei der Erhöhung der Anwohner-Parkgebühren zeigt sich eine Ablehnungsrate von 47 % bei 31 % Zustimmung. Die Zahlen weisen damit auf ein hohes Ausmaß an Polarisierung hin. Zusätzliche Analysen zeigen, dass höhere Parkgebühren in Berlin noch unpopulärer wären als die Einführung einer City-Maut oder die Ausweitung von Tempo-30 Zonen.[46]

Lässt sich dieses polarisierte Meinungsbild durch unterschiedliche Kommunikationsstrategien beeinflussen? Die Analyse der unterschiedlichen Variationen der Fragen legt nahe, dass dies der Fall ist. Wird etwa betont, dass die zusätzlichen Mittel in den öffentlichen Nahverkehr investiert werden (*Einnahmenrecycling*), steigt die Zustimmungsrate für höhere Gebühren für das Kurzparken um 8 bis 9 Prozentpunkte (Tabelle 1,

46 Vgl. Traxler & Wegrich 2023. Eine Mehrheit der Befragten würde jedoch den Ausbau der Radverkehrsinfrastruktur und von verkehrsberuhigten Zonen unterstützen (‚Kiezblocks').

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

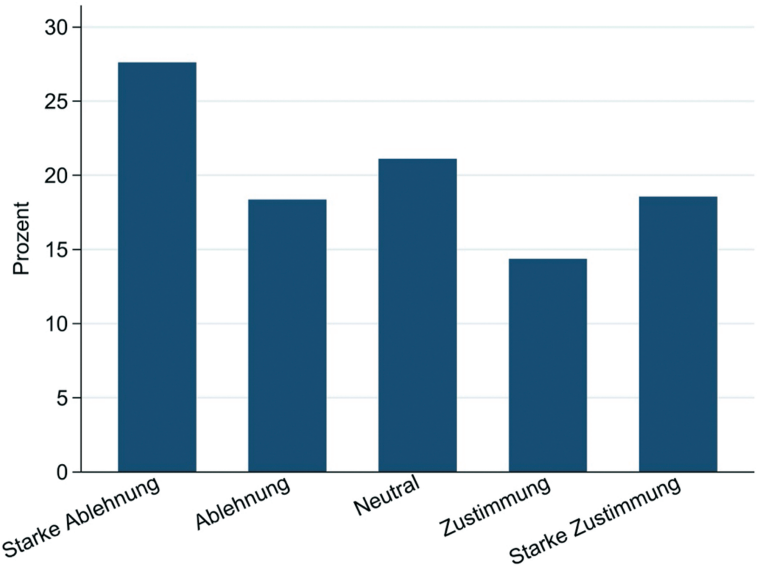


Abbildung 2 – Positionen zur Erhöhung der Anwohner-Parkgebühren

Spezifikation 1 und 2).[47] Für die Zustimmung zum Anwohnerparken zeigt sich ein positiver Effekt von 6 Prozentpunkten (Spezifikation 3 und 4). Dieses Ergebnis deckt sich qualitativ mit vergleichbaren Resultaten aus der Literatur.[48]

Die in Tabelle 1 aufgeführten Schätzungen zeigen, dass das Anführen von hohen *Referenzwerten* zu ähnlich starken Effekten führt: Wird auf die Parkgebühren in Amsterdam oder Paris verwiesen, so steigen die Zustimmungsraten um 5 bis 7 Prozentpunkte. Das Betonen des institutionellen *status quo* führt nur im Fall des Kurzparkens (wo hervorgehoben wurde, dass es sich um die erste Gebührenerhöhung seit 18 Jahren handeln würde) zu einem positiven Effekt (Spezifikation 1 und 2). Für das Anwohnerparken (wo die niedrigen aktuellen Gebühren erwähnt wurden) zeigt sich kein statistisch signifikanter Effekt.

Die Effekte auf die Zustimmungsraten zeigen sich, spiegelbildlich, auch für Ablehnungsraten und sind, wie auch in Tabelle 1 illustriert, durch

47 Relativ zur gemessenen Zustimmungsrate in der Kontrollgruppe entspricht dies einem 30-prozentigen Anstieg.

48 Arlinghaus u. a. 2025.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

das randomisierte Design robust hinsichtlich der Kontrolle auf unterschiedlichste sozio-ökonomische Hintergrundcharateristika. Für die methodischen Feinschmecker lassen sich die Ergebnisse auch in geordneten Probit- und Logit-Modellen replizieren. Die Analyse von typenspezifischer Effektheterogenität steht noch aus.[49]

Nach alledem führten die Variationen in der Information bzw. im Framing der Fragen also zu starken Effekten. Hervorzuheben ist hierbei, dass die geschätzen Effekte additiv sind.[50] Für die Gebührenerhöhung in Kurzparkzonen bedeutet dies etwa, dass das gleichzeitige Betonen von Einnahmenrecycling, hohen Referenzwerten und seltenen Gebührenanpassungen die Zustimmungsrate von 26 % auf 46 % erhöht und so die relativen Mehrheiten verschiebt. Die Ergebnisse legen somit nahe, dass unterschiedliche Kommunikationsstrategien einen wesentlichen Einfluss auf die artikulierten Positionen der Bürger haben.

Tabelle 1 – Schätzung der Treatment-Effekte

	(1)	(2)	(3)	(4)
	Kurzparkzonen		*Anwohnerparken*	
Einnahmenrecycling	0.078**	0.089***	0.063**	0.062**
	(0.031)	(0.029)	(0.031)	(0.028)
Referenzwerte	0.064**	0.053*	0.060**	0.069**
	(0.031)	(0.029)	(0.030)	(0.029)
Status-quo Information	0.056**	0.055**	0.013	0.016
	(0.025)	(0.024)	(0.025)	(0.024)
Konstante	0.258***	n.a.	0.266***	n.a.
	(0.025)		(0.024)	
Kontrollvariablen	Nein	Ja	Nein	Ja
Beobachtungen	1,370	1,370	1,370	1,370

Hinweise: Die Tabelle zeigt die geschätzen Koeffizienten sowie Heteroskedastie-robuste Standardfehler aus linearen Wahrscheinlichkeitsmodellen. Die abhängige Variable ist ein Indikator für eine (normale/starke) Zustimmung zur Erhöhung der jeweiligen Parkgebühr. Spezifikationen (2) und (4) inkludieren Kontrollvariablen (u. a. Alter, Familienstruktur, Bildungniveau, Autobesitz, und Einkommen). ***/**/* markieren statistische Signifikanz auf einem 1%-/5%-/10%-Niveau.

49 Dazu Engel 2020a.

50 Interaktionseffekte scheinen vernachlässigbar zu sein.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

F. Fazit

Das Schicksal der *asphalt commons* muss nicht tragisch sein. Zwar halten Gerichte und Politik an einem Leitbild fest, das private Pkw und deren Halter – selbst im ruhenden Verkehr – privilegiert. Viele Bürger sind jedoch durchaus bereit, eine progressivere Mobilitätspolitik mitzutragen – sofern eine sorgfältige Kommunikationsstrategie die Akzeptanz der Maßnahmen stärkt, die der Übernutzung des öffentlichen Raumes entgegenwirken. Auf gute Kommunikation und kluge Worte sollten aber auch Taten folgen. Dass dies trotz anfänglichem Widerstand möglich ist, haben andere europäische Großstädte gezeigt: In Lyon, Paris und Stockholm haben mutige verkehrspolitische Maßnahmen Meinungen verändert.[51] Diese Beispiele zeigen, wie politisches Handeln und rechtliche Regelungen soziale Normen verschieben können. Hier setzt auch das Potenzial einer empirisch fundierten und verhaltenswissenschaftlich informierten Rechtswissenschaft an. Sie weist einerseits die Grenzen der Problemlösungskraft des geltenden Rechts aus und zeigt andererseits konstruktiv auf, welche Lösungspfade sich anbieten. *Christoph Engel* hat einer solchen Rechtswissenschaft den Weg bereitet.

Literatur

Agora Verkehrswende (2022). *Umparken – den öffentlichen Raum gerechter verteilen. Zahlen und Fakten zum Parkraummanagement*. 4. Auflage. Berlin. https://www.agora-verkehrswende.de/fileadmin/Projekte/2022/Umparken/Agora-Verkehrswende_Factsheet_Umparken_Auflage-4.pdf

Arlinghaus, J., Konc, T., Mattauch, L., & Sommer, S. (2025). The Effect of Framing on Policy Support: Experimental Evidence from Urban Transport Policies. *SSRN*. https://doi.org/10.2139/ssrn.5359102

Arnott, R., & Inci, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, *60*(3), 418–442.

Bauernschuster, S., & Traxler, C. (2021). Tempolimit 130 auf Autobahnen: Eine evidenzbasierte Diskussion der Auswirkungen. *Perspektiven der Wirtschaftspolitik*, *22*(2), 86–102.

Baumol, W. J. (1972). On Taxation and the Control of Externalities. *American Economic Review*, *62*(3), 307–322.

51 Börjesson u. a. 2012; siehe auch Manderscheid 2022: 70.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Börjesson, M., Eliasson, J., Hugosson, M. B., & Brundell-Freij, K. (2012). The Stockholm congestion charges – 5 years on. Effects, acceptability and lessons learnt. *Transport Policy*, *20*, 1–12.

Canzler, W. (2024). *Benamo Hintergrundpapier Nr. 3 zum Policy Paper Nr. 1: „Parken – der ruhende Verkehr und die Verkehrswende"*. Berlin: Nexus/WZB.

Cats, O., Zhang, C., & Nissan, A. (2016). Survey methodology for measuring parking occupancy: Impacts of an on-street parking pricing scheme in an urban center. *Transport Policy*, *47*, 55–63.

Chatziathanasiou, K. (2025). Das Recht des „ruhenden Verkehrs": Herausforderung für die effiziente Nutzung des öffentlichen Raums. *SSRN*. https://doi.org/10.2139/ssrn.5201767

Creutzig, F., Javaid, A., Soomauroo, Z., Lohrey, S., Milojevic-Dupont, N., Ramakrishnan, A., Sethi, M., Liu, L., Niamir, L., Bren d'Amour, C., Weddige, U., Lenzi, D., Kowarsch, M., Arndt, L., … Zausch, J. M. (2020). Fair street space allocation: ethical principles and empirical insights. *Transport Reviews*, *40*(6), 711–733.

Engel, C. (Hrsg.) (1998). *Methodische Zugänge zu einem Recht der Gemeinschaftsgüter*. Baden-Baden: Nomos.

Engel, C. (2002). *Abfallrecht und Abfallpolitik*. Baden-Baden: Nomos.

Engel, C. (2007). Herrschaftsausübung bei offener Wirklichkeitsdefinition. In: C. Engel & W. Schön (Hrsg.), *Das Proprium der Rechtswissenschaft* (S. 206–240). Tübingen: Mohr Siebeck.

Engel, C. (2008). The difficult reception of rigorous descriptive social science in the law. In: N. Stehr & B. Weiler (Hrsg.), *Who Owns Knowledge? Knowledge and the Law* (S. 169–213). Routledge.

Engel, C. (2018). Empirical Methods for the Law. *Journal of Institutional and Theoretical Economics*, *174*(1), 5–23.

Engel, C. (2020a). Estimating heterogeneous reactions to experimental treatments. *Journal of Economic Behavior & Organization*, *178*, 124–147.

Engel, C. (2020b). Rechtswissenschaft als Sozialwissenschaft – Spurensuche im Jahrgang 2019 der ZGR. In: *Festschrift für Christine Windbichler zum 70. Geburtstag* (S. 33–50). Berlin: De Gruyter.

Engel, C. (2024). The Law – An Impossible Discipline. *Review of Law & Economics*, *20*(3), 325–340.

Fehling, M. (2024). Rechtliche Gestaltungsoptionen für eine Verkehrswende in den Kommunen (Teil 2). *InfrastrukturRecht*, 64–68.

Gössling, S., Humpe, A., Hologa, R., Riach, N., & Freytag, T. (2022). Parking violations as an economic gamble for public space. *Transport Policy*, *116*, 248–257.

Gössling, S., Schröder, M., Späth, P., & Freytag, T. (2016). Urban Space Distribution and Sustainable Transport. *Transport Reviews*, *36*(5), 659–679.

Grönemeyer, H. (1984). Mambo. In: *Bochum*.

Inci, E. (2015). A review of the economics of parking. *Economics of Transportation, 4*(1), 50–63.

Manderscheid, K. (2022). *Soziologie der Mobilität*. Stuttgart: utb.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Matias, M., Mills, G., Silva, T., Girotti, C., & Lopes, A. (2025). The underestimated impact of parked cars in urban warming. *City and Environment Interactions*, *28*, 100232.

Nobis, C., & Kuhnimhof, T. (2018). *Mobilität in Deutschland– MiD: Ergebnisbericht*. https://elib.dlr.de/125879/1/MiD2017_Ergebnisbericht.pdf

Ostermeijer, F., Koster, H., Nunes, L., & Van Ommeren, J. (2022). Citywide parking policy and traffic: Evidence from Amsterdam. *Journal of Urban Economics*, *128*, 103418.

Ottosson, D. B., Chen, C., Wang, T., & Lin, H. (2013). The sensitivity of on-street parking demand in response to price changes: A case study in Seattle, WA. *Transport Policy*, *25*, 222–232.

Shoup, D. C. (2004). The ideal source of local public revenue. *Analysis of Urban Land Markets and the Impact of Land Market Regulation*, *34*(6), 753–784.

Shoup, D. C. (2011). *The High Cost of Free Parking [2005]*. Updated edition. Routledge.

Szell, M. (2018). Crowdsourced quantification and visualization of urban mobility space inequality. *Urban Planning*, *3*(1), 1–20.

Taylor, E. (2021). Free parking for free people: German road laws and rights as constraints on local car parking management. *Transport Policy*, *101*, 23–33.

The Economist. (2024, September 7). *America's killer cars*. https://www.economist.com/interactive/united-states/2024/08/31/americans-love-affair-with-big-cars-is-killing-them

Topp, H. H. (1993). Parking policies to reduce car traffic in German cities. *Transport Reviews*, *13*(1), 83–95.

Traxler, C. (2025). Communication and support for the mobility transition: Survey evidence from Berlin. *Mimeo*.

Traxler, C., & Wegrich, K. (2023). *Einstellungen zur Mobilitätswende: Ergebnisse einer Befragung in Berlin, Hertie School – Policy Brief*. https://www.hertie-school.org/fileadmin/2_Research/5_Policy_Briefs/Policy_Brief_Einstellungen_zur_Mobilitaetswende_DE.pdf

Traxler, C., Westermaier, F. G., & Wohlschlegel, A. (2018). Bunching on the Autobahn? Speeding responses to a ‚notched' penalty scheme. *Journal of Public Economics*, *157*, 78–94.

van Ommeren, J., Wentink, D., & Dekkers, J. (2011). The real price of parking policy. *Journal of Urban Economics*, *70*(1), 25–31.

Verhoef, E., Nijkamp, P., & Rietveld, P. (1995). The economics of regulatory parking policies: The (IM)possibilities of parking policies in traffic regulation. *Transportation Research Part A: Policy and Practice*, *29*(2), 141–156.

Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review*, *59*(2), 251–260.

Yin, C., Shao, C., & Wang, X. (2018). Built environment and parking availability: Impacts on car ownership and use. *Sustainability*, *10*(7), 2285.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Staatliche Opportunität
Kontrollmodi der Wirtschaftsaufsicht

Stefanie Egidy

A. Einleitung

Kontrollen privater Unternehmen stehen im öffentlichen Wirtschaftsrecht an vorderster Front der Rechtsdurchsetzung.[1] Aufgrund von Budgetrestriktionen können die zuständigen Behörden jedoch nicht alle Betriebe laufend kontrollieren.[2] Sie müssen immer wieder die schwierige Entscheidung treffen, welche Betriebe sie inspizieren. Kontrolliert die Behörde die Einen, entgehen die Anderen einer Inspektion. Unzureichende Kontrollen begünstigen die Verwirklichung von Gefahren, insbesondere für die Angestellten und die Kunden eines Unternehmens.[3] Beispielsweise kann der Konsum von unter unhygienischen Bedingungen hergestellten Produkten zu Gesundheitsschäden führen.[4] Gleichzeitig belasten Kontrollen die betroffenen Unternehmen, gerade wenn Beanstandungen über Transparenzplattformen oder durch die Behörden selbst, wie vor einigen Jahren für Lebensmittelkontrollen von Restaurants in Form der sogenannten „Pankower Ekelliste",[5] veröffentlicht werden.[6] Obschon Kontrollen damit für Unternehmen wie für Verbraucher grundrechtliche Relevanz entfalten, überantwortet der Gesetzgeber die Steuerung der Auswahl und die Intensität der Kontrollen weithin dem Ermessen der Verwal-

1 *May/Wood*, Journal of Public Administration Research and Theory 13 (2002), S. 117; *Buckley*, Food Policy 51 (2015), S. 74 (75); *Ho*, UC Irvine L. Rev. 7 (2017), S. 401 (407).

2 *Stigler*, Journal of Political Economy 78 (1970), S. 526 (526 f.).

3 Empirisch zeigen *Levine/Toffel/Johnson*, Science 336 (2012), S. 907 die Effektivität von Routinekontrollen der U.S. Occupational Safety and Health Administration.

4 *Antle*, in: Gardner/Rausser (Hrsg.), Handbook of Agricultural Economics, Vol. I, 2001, S. 1083 (1086–1088); zu der supranationalen Gefahrendimension siehe *F. Wollenschläger*, VERW 52 (2019), S. 1 (2 f.).

5 Krit. siehe *Spiecker gen. Döhmann*, Zeitschrift für Kultur- und Kollektivwissenschaft 1 (2015), S. 49 (60–66).

6 Siehe zudem zur Grundrechtsrelevanz von Warnungen *F. Wollenschläger*, VERW 52 (2019), S. 1 (17–19).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tung.[7] Welche Kontrollmodi besonders erfolgversprechend sind – sei es durch die Entdeckung von gegenwärtigen Missständen, sei es durch ihre abschreckende Wirkung für die Zukunft – identifizieren ökonomische Modelle und empirische Evidenz. Das weite Ermessen der Verwaltung bietet hinreichend Spielraum, um derartige Erkenntnisse zu integrieren. Als Anleitung dieser Ermessensausübung eignet sich angesichts der Bedeutung der Kontrollen gerade mit Blick auf die knappen behördlichen Kontrollressourcen die Handlungsform des Kontrollkonzeptes. Indem dieser Beitrag sie in den Mittelpunkt stellt, verbindet er – wie der Jubilar – Verwaltungsrechtswissenschaft und Sozialwissenschaft.[8]

Den Ausgangspunkt meiner Betrachtung bildet deshalb eine ökonomische Analyse von Kontrollkonzepten zur Rechtsdurchsetzung (B.). Kontrastierend hierzu zeige ich, dass die bisher angenommene rechtliche Steuerung der Kontrollen defizitär ist (C.). Drittens aktiviert mein Lösungsansatz die von der Rechtsprechung entwickelte Konzeptpflicht im Rahmen der Ermessensausübung (D).

B. Kontrollen im öffentlichen Wirtschaftsrecht aus ökonomischer Perspektive

I. Regulierungs- und Durchsetzungsbedarf

Das öffentliche Wirtschaftsrecht dient der Gefahrenabwehr und normiert deshalb zahlreiche Verhaltenspflichten. Sie lassen sich aus wohlfahrtsökonomischer Perspektive als Reaktion auf ein Marktversagen beschreiben.[9] Zwischen den Wirtschaftsunternehmen und ihren Kunden besteht eine Informationsasymmetrie.[10] Die innerhalb von Unternehmen wurzelnden Gefahrenquellen – wie die Salmonellengefahr in Restaurants – sind von

7 Hierzu *Eifert*, VERW 39 (2006), S. 309 (317–320).

8 Siehe nur für das Ermessen *Engel*, in: ders./Schön (Hrsg.), Das Proprium der Rechtswissenschaft, 2007, S. 205 (220 f., 229–231).

9 *Pigou*, The Economics of Welfare, 4. Aufl. 1932, S. 127 ff., 243 ff., 329 ff.; krit. *Posner*, Bell Journal of Economics and Management Science 5 (1974), S. 335 (335–341); zu wohlfahrtsökonomischen Kosten und Nutzen *Antle*, in: Gardner/Rausser (Hrsg.), Handbook of Agricultural Economics, Vol. I, 2001, S. 1083 (1119–1129).

10 Siehe nur *Focker/van der Fels-Klerx*, Current Opinions in Food Science 36 (2020), S. 18 (21); *Antle*, in: Gardner/Rausser (Hrsg.), Handbook of Agricultural Economics, Vol. I, 2001, S. 1083 (1085 f.); so auch BVerfGE 148, 40 Rn. 29.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

außen kaum beobachtbar.[11] Signale wie die Reputation des Restaurants oder eigene Sinneseindrücke von den verzehrten Produkten sind meist nicht zuverlässig.[12] Damit verursachen Verstöße gegen Hygienevorschriften soziale Kosten, die nicht internalisiert werden.[13] Dies führt zu einer adversen Selektion.[14] Das heißt: Weil die Konsumenten nicht erkennen können, ob eine Speise hygienisch produziert ist, sind sie nicht bereit, den entsprechenden Preis für sie zu zahlen. Dies nimmt wiederum den Produzenten den Anreiz, die Hygienevorschriften einzuhalten und verschlechtert die im Markt angebotene Qualität der Lebensmittel.[15] *George Akerlof* hat dieses Problem im Gebrauchtwagenhandel als „market for lemons" (also ein „Markt für Montagsautos") beschrieben.[16]

Um die Regulierungsziele zu erreichen, bedarf es staatlicher Verhaltensvorschriften und ihrer Durchsetzung. Wie *Gary Becker* zeigte, ist es für Unternehmen individuell rational, Verhaltenspflichten zu verletzen, wenn die erzielten Einsparungen größer sind als die erwarteten Kosten im Falle der Entdeckung.[17] Diese Kosten ergeben sich im einfachsten Fall aus der Sanktionshöhe multipliziert mit der Entdeckungswahrscheinlichkeit.[18] Hier besteht wiederum eine Informationsasymmetrie – nunmehr zwischen dem regulierten Unternehmen und der Behörde.[19] Denn durch Lebensmittel verursachte Krankheiten lassen sich nur selten eindeutig auf einen Verursacher zurückführen. Dies illustriert die Suche nach dem

11 *Lüdemann*, in: Augsberg (Hrsg.), Extrajuridisches Wissen im Verwaltungsrecht, 2013, S. 121 (130); *Ho/Altenburger*, JITE 175 (2018), S. 98 (101) betonen, wie schwierig es ist, Lebensmittelvergiftungen zuzuordnen; *Crumley*, Review of Litigation 31 (2012), S. 353 (370).

12 Siehe zu Vertrauensgütern *Lüdemann*, in: Augsberg (Hrsg.), Extrajuridisches Wissen im Verwaltungsrecht, 2013, S. 121 (130 f.); *Antle*, in: Gardner / Rausser (Hrsg.), Handbook of Agricultural Economics, Vol. I, 2001, S. 1983 (1113 f.).

13 *Crumley*, Review of Litigation 31 (2012), S. 353 (371 f.).

14 *Akerlof*, Quarterly Journal of Economics 84 (1070), S. 488 (493); siehe auch *Crumley*, Review of Litigation 31 (2012), S. 353 (370 f.).

15 Zum *moral hazard* siehe auch *Rodi*, Ökonomische Analyse des Öffentlichen Rechts, 2014, S. 45.

16 *Akerlof*, Quarterly Journal of Economics 84 (1970), S. 488 (489 f.).

17 *Becker*, Journal of Political Economy 76 (1968), S. 169 (176–179); siehe auch *Polinsky/Shavell*, Journal of Economic Literature 38 (2000), S. 45 (47 f.).

18 *Becker*, Journal of Political Economy 76 (1968), S. 169 (176 f.); komplexer beziehen *Gray/Shimshack*, Review of Environmental Economics and Policy 5 (2011), S. 3 (10) die subjektive Einschätzung der Entdeckungs- und Sanktionsgefahr sowie der Sanktionshöhe mit ein.

19 *Baron/Besanko*, RAND Journal of Economics 15 (1984), S. 447 (447 f.).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Auslöser der EHEC-Epidemie im Sommer 2011 – der letztlich nicht Tomaten, Gurken oder Blattsalat, sondern Sprossen waren.[20] Die Behörde erfährt von Normverstößen deshalb vor allem durch Kontrollen.[21] Dabei bedarf es einer staatlichen Rechtsdurchsetzung, weil eine privatrechtliche Ahndung von Rechtsverstößen zumeist daran scheitert, dass den Geschädigten die notwendigen Informationen über die Identität des Schädigers sowie die schadensbegründenden Umstände fehlen.[22] Will die Behörde die gesetzten Verhaltenspflichten durchsetzen, muss sie daher – unabhängig von der Sanktionshöhe – mit Hilfe von Kontrollen sicherstellen, dass die erwartete Entdeckungswahrscheinlichkeit größer als Null ist. Der Staat verfügt für diese Kontrollen nur über begrenzte Ressourcen, insbesondere Personal- und Sachmittel.[23] Unter den geltenden Budgetrestriktionen kann die Verwaltung keine kontinuierliche Kontrolle aller Betriebe leisten.[24] Ökonomische Modelle können dazu beitragen, die Effizienz der Kontrollstrategie zu erhöhen.

Zunächst muss die Verwaltung entscheiden, welches Ziel sie optimieren möchte. Vor dem Hintergrund der haushaltsrechtlichen Budgetierung sowie *William Niskanens* Postulat, dass die Bürokratie ein Eigeninteresse an der Maximierung ihrer Budgets hat,[25] liegt es nicht im Interesse der Behörde, die Durchsetzungskosten zu minimieren – abgesehen von Wirtschaftlichkeitserwägungen nach dem Neuen Steuerungsmodell.[26] Stattdessen besteht ihr Ziel meist darin, mit den verfügbaren Ressourcen eine

20 Siehe *Bernhard* et al., Journal für Verbraucherschutz und Lebensmittelsicherheit 6 (2011), S. 483.

21 Zudem gibt es ein unionsrechtliches Warnsystem, das Gesundheitsgefahren identifiziert, siehe *F. Wollenschläger*, VERW 52 (2019), S. 1 (4–13).

22 *Crumley*, Review of Litigation 31 (2012), S. 353 (375–378); siehe hierzu *Polinsky*/*Shavell*, in: dies. (Hrsg.), Handbook of Law and Economics, 2007, Vol. 1, S. 403 (406); *Shavell*, Journal of Legal Studies 9 (1980), S. 1 (4–5); *Lüdemann*, in: Augsberg (Hrsg.), Extrajuridisches Wissen im Verwaltungsrecht, 2013, S. 121 (130); *Schmidt-Kessel*, GewArch 2016, S. 169 (169).

23 Zu Kontrolldefiziten durch Personalmangel siehe Foodwatch, Kontrolle ist besser, 2019, abrufbar unter https://www.foodwatch.org/fileadmin/-DE/Themen/Lebensmittelkontrollen/2019-12_foodwatch_Kontrolle-ist-besser_final.pdf (letzter Abruf 1. Dez. 2025).

24 Das Problem beschreibt *Stigler*, Journal of Political Economy 78 (1970), S. 526 (526 f.).

25 *Niskanen*, American Economic Review 58 (1968), S. 293.

26 Siehe nur *Schmidt-Aßmann*, Das allgemeine Verwaltungsrecht als Ordnungsidee, 2. Aufl. 2006, S. 23 f. Rn. 41 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

effektive Rechtsdurchsetzung zu erreichen, wobei der Zuschnitt und die Dichte von Kontrollen sowie die nachfolgende Vollstreckung bzw. Sanktionierung wichtige Stellschrauben darstellen.[27] Hierzu stehen der Verwaltung verschiedene Kontrollmodelle zur Verfügung.

II. Ökonomische Evidenz zu Kontrollmodellen

Idealtypisch kann die Verwaltung zwischen zwei Kontrollmodellen wählen.[28] Erstens kann sie Zufallskontrollen durchführen, indem sie eine randomisierte Stichprobe aus der Gesamtheit der Kontrollsubjekte zieht. Zweitens kann sie risikobasiert kontrollieren, indem sie für jedes Kontrollsubjekt einen individuellen Risikowert anhand verschiedener Faktoren berechnet und damit die Kontrollhäufigkeit festlegt. Die beiden Modelle lassen sich miteinander kombinieren und in ein komplexeres Kontrollkonzept integrieren.[29]

Die empirische ökonomische Literatur hat sich intensiv mit der Frage beschäftigt, wie sich der Einsatz dieser Modelle auf die Aufdeckungsrate und das Befolgungsverhalten der Kontrollierten auswirkt. Sie widmet sich einerseits der Durchsetzung von umwelt-, verbraucher- und arbeitsschutzrechtlichen Vorschriften gegenüber Unternehmen.[30] Andererseits untersucht sie – auch experimentell – das Steuerbefolgungsverhalten von Privatpersonen,[31] das allerdings nur begrenzt Einsicht in die Mechanismen der Durchsetzung des öffentlichen Wirtschaftsrechts gegenüber Unternehmen gewährt. Im Fokus stehen im Folgenden daher die empirischen Studien zu Wirtschaftskontrollen.

Sie identifizieren verschiedene Elemente effizienter Kontrollmodelle. So erreicht im Vergleich eine gezielte Auswahl der Kontrollsubjekte eine

27 *De Benedetto*, European Journal of Risk Regulation 9 (2018), S. 391 (405–407).

28 Siehe etwa *Blanc*/*Faure*, Journal of Risk Research, 23 (2020), S. 1405 (1407–1410); empirisch im Vergleich *Levine*/*Toffel*/*Johnson*, Science 336 (2012), S. 907.

29 Vgl. den experimentellen Vergleich zwischen zwei risikobasierten Kontrollmodellen und einem Zufallskontrollmodell *Clark*/*Friesen*/*Muller*, Economic Inquiry 42 (2004), S. 69.

30 Siehe nur *Helland*, Review of Economics and Statistics 80 (1998), S. 141; *Shimshack*/*Ward*, Journal of Environmental Economics and Management 50 (2005), S. 519; *Jung*/*Makowsky*, Journal of Regulatory Economics 45 (2014), S. 1; *Faure* et al., Loyola Consumer L. Rev. 20 (2008), S. 361.

31 Siehe im Überblick *Mascagni*, Economic Surveys 32 (2018), S. 273.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

höhere Aufdeckungsrate als zufällige Kontrollen.[32] Eine Gesamtschau zeigt, dass Kontrollen auch die zukünftige Normbefolgung erhöhen.[33] Dabei steigt die Normbefolgung, je höher die erwartete Aufdeckungsrate ist.[34] Wie stark Kontrollen abschrecken, hängt auch von den Eigenschaften der Betriebe ab,[35] etwa wie kostspielig es für sie ist, Normen einzuhalten.[36] Wichtig ist aber, dass es einer Mindestzahl stichprobenartiger Kontrollen bedarf, um eine hinreichende abschreckende Wirkung zu erzeugen.[37] Zudem müssen die Kontrollen selbst bestimmte Eigenschaften erfüllen, insbesondere unangekündigt stattfinden und schwer vorhersehbar sein.[38] So waren bei Lebensmittelkontrollen in Las Vegas die aufgedeckten Normverstöße um ein Drittel höher, wenn ein Betrieb in einem großen Einkaufszentrum oder Hotelkomplex als erstes kontrolliert wurde und somit nicht vorgewarnt war.[39]

Neben ihrer Kontrolleffizienz unterscheiden sich die Modelle in ihrer Anwendungskomplexität. Stichprobenartige Kontrollen sind ohne vorherige Informationen und Berechnungen umsetzbar, während der Staat für eine risikobasierte Bewertung darauf angewiesen ist, die notwendigen Informationen von den Unternehmen selbst zu erhalten oder durch eigene Kontrollen zu erzeugen, um eine angemessene Risikoabschätzung zu leisten.[40] Dieser Informationsbedarf kennzeichnet das Überwachungsrechtsverhältnis zwischen Staat und Unternehmen.[41]

32 Siehe etwa *Levine/Toffel/Johnson*, Science 336 (2012), S. 907.

33 Im Überblick *Gray/Shimshack*, Review of Environmental Economics and Policy 5 (2011), S. 3 (12–17) m. w. N.

34 Ebd., S. 15 f.

35 *Earnhart*, Journal of Regulatory Economics 36 (2009), S. 247.

36 *Gray/Shimshack*, Review of Environmental Economics and Policy 5 (2011), S. 3 (19 f.).

37 Ebd.

38 So auch BVerwG, Beschl. v. 28. Jan. 1998, Az.: 1 B 5.98.

39 *Makofske*, Journal of Economic Behavior & Organization 190 (2021), S. 348.

40 Zu den Modi administrativer Wissensgenerierung siehe *Augsberg*, VERW 51 (2018), S. 351 (355 f., 358).

41 *Gröschner*, Das Überwachungsrechtsverhältnis, 1992, S. 305–330; *Di Fabio*, Risikoentscheidungen im Rechtsstaat, 1994, S. 455–457; *Huber*, in: Voßkuhle/Eifert/Möllers (Hrsg.), Grundlagen des Verwaltungsrechts, Bd. II, 3. Aufl. 2022, S. 1313 (Rn. 7–11, 52, 58, 72, 147–161).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. Entscheidungstheoretische Durchsetzungsdefizite

Kontrollen unterliegen allerdings entscheidungstheoretischen Durchsetzungsdefiziten. Bereits im rational-theoretischen Modell der ökonomischen Regulierung identifiziert *George Stigler* das Problem der *regulatory capture*.[42] Es beschreibt die Gefahr, dass Aufsichtsbehörden durch mächtige Unternehmen vereinnahmt werden, insbesondere durch finanzielle Anreize, und so die Rechtsdurchsetzung ihnen gegenüber defizitär bleibt. Ein Kontrolleur mag deshalb einen Betrieb mit wachsender Vertrautheit weniger intensiv kontrollieren, um negative Reaktionen zu vermeiden.[43] So konstatierte der damalige Leiter der Bankenaufsicht der BaFin im Rahmen des Wirecard-Skandals vor dem Untersuchungsausschuss selbstkritisch, es habe der Aufsicht die „nötige kritische Distanz" gefehlt.[44] Zudem stellt das innerbehördliche Prinzipal-Agent-Verhältnis vor Herausforderungen, indem es für individuelle Amtswalter rational sein kann, die Kontrollen nicht mit gleichbleibender Intensität durchzuführen, etwa weil sie spieltheoretisch als wiederholte Spieler (*repeat player*) an einer kooperativen Kontrollbeziehung interessiert sind.[45]

Wie *Herbert Simon* und andere herausarbeiten, unterliegt das Kontrollverhalten der Amtswalter zudem verhaltensökonomischen Einschränkungen.[46] Ihre Gefahrenprognosen weichen von der Erwartungsnutzentheorie (*rational choice*) ab, indem sie Vorurteilen und kognitiven Verzerrungen unterliegen.[47] Diese *biases* können sowohl die Auswahl der Kontrollziele als auch die Durchführung der Kontrollen betreffen.[48]

---

42 *Stigler*, Bell Journal of Economics and Management Science 2 (1971), S. 3 (3–7).

43 Siehe auch *Bernstein*, Regulating Business by Independent Commissions, 1955, S. 165 ff.

44 Beschlussempfehlung und Bericht des 3. Untersuchungsausschusses der 19. Wahlperiode gemäß Artikel 44 des Grundgesetzes, BT-Drs. 19/30900, S. 1114 f. (Zitat der Aussage von Raimund Röseler, Stenografisches Protokoll 19/36 der 36. Sitzung am 13. April 2021, S. 47).

45 *Egidy*, in: Engel et al. (Hrsg.), Öffentliches Recht als Verhaltensordnung, S. 229 (245).

46 *Simon*, Administrative Behavior, 4. Aufl. 1997, S. 72 ff., 118 ff.; im Überblick *Ip*, Common Law World Rev. 46 (2017), S. 171 (173–177).

47 *Simon*, Administrative Behavior, 4. Aufl. 1997, S. 72 ff., 118 ff.; im Überblick aus verwaltungswissenschaftlicher Sicht *Battaglio* et al., Public Administration Rev. 79 (2019), S. 304.

48 *Ibanez/Toffel*, Management Science 66 (2019), S. 2396 zeigen den Einfluss von früheren Kontrollergebnissen sowie am Kontrolltag der Kontrollreihenfolge und Auslastung auf die Kontrollergebnisse.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Bei wiederholten Kontrollen durch den gleichen Amtswalter kann etwa ein *familiarity bias* entstehen.[49] Ihm zufolge erscheinen bekannte Kontrollstätten als weniger riskant bzw. gefahrenträchtig als unbekannte Kontrollobjekte. Tatsächlich weisen empirische Studien zu Gaststättenkontrollen darauf hin, dass Kontrolleure bei erstmaligen Inspektionen eines Unternehmens strengere Bewertungsmaßstäbe anlegen als bei späteren.[50] Empirische Studien zeigen außerdem, dass Risikobewertung und Kontrollintensität auch von der Ethnizität der Restaurants abhängen, etwa für asiatische Restaurants in den USA.[51]

C. Rechtliche Regelung der behördlichen Kontrolltätigkeit

Das öffentliche Wirtschaftsrecht reguliert die behördlichen Kontrollen auf verschiedenen Ebenen.

I. Kontrollkompetenzen zur Gefahrenabwehr im öffentlichen Wirtschaftsrecht

Es räumt den Überwachungsbehörden zunächst Kontrollkompetenzen zur Gefahrenabwehr ein, beispielsweise den Behörden der Finanzaufsicht,[52] der Umweltaufsicht[53] und für Arbeitsschutz.[54] Aufgrund ihrer intensiven auch empirischen Aufarbeitung sollen das Gaststättenrecht und das Lebensmittelrecht auch im Folgenden als Referenzgebiete dienen.

§ 22 Abs. 2 GastG ermächtigt die Aufsichtsbehörden dazu, im Wege der Nachschau die Geschäftsräume der Gaststättenbetriebe zu betreten, um „dort Prüfungen und Besichtigungen vorzunehmen".[55] Sie können dabei insbesondere die gaststättenrechtliche Zuverlässigkeit des Betreibers anhand von Indizien prüfen, etwa ob er dem Alkoholmissbrauch Vorschub

49 Auch „mere exposure effect", siehe *Zajonc*, Current Directions in Psychological Science 10 (2001), S. 224; *O'Hara*, Western New England L. Rev. 39 (2017), S. 101 (117).

50 Siehe etwa *Jin/Lee*, Journal of Law and Economics 61 (2018), S. 159.

51 *Ho*, UC Irvine L. Rev. 7 (2017), S. 401 (407–417).

52 Siehe z. B. § 44 KWG.

53 Siehe z. B. § 52 BNatSchG.

54 § 22 ArbSchuG.

55 Zur verfassungsrechtlichen Dimension der Nachschau siehe *Ennuschat*, AöR 127 (2002), S. 252 (253–256); *Tille*, GewArch 2017, S. 184 (184 f.).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

leistet oder Hygienevorschriften missachtet.[56] Eine entsprechende Kompetenz zur Lebensmittelüberwachung gewährt Art. 9 der EU-Kontrollverordnung. Auch sie dient der Ermittlung von Verstößen gegen Hygienevorschriften. Ein Vergleich dieser beiden Regelungssysteme ist instruktiv, weil sie zwar eine große inhaltliche Schnittmenge aufweisen, aber die Durchführung von Kontrollen ganz unterschiedlich regulieren.

II. Inneradministrative Steuerungsmechanismen der Kontrollhäufigkeit

Jenseits dieser Ermächtigungsnormen unterliegen Kontrollen inneradministrativen Steuerungsmechanismen. Das Gaststättenrecht stellt es in das Ermessen der Behörde, welche Betriebe sie wie oft kontrolliert.[57] Dies gilt besonders für Routinekontrollen, die nicht auf Gefahranzeigen beruhen.[58] Die zuständigen Kontrolleure müssen keinem Kontrollkonzept folgen, sondern dürfen sich vielmehr in ihrem Vorgehen von Zweckmäßigkeitserwägungen sowie subjektiven Eindrücken leiten lassen.[59]

Im Kontrast hierzu gestaltet das Lebensmittelrecht die Kontrolltätigkeit näher aus. In Reaktion auf die BSE-Krise vor 20 Jahren vereinheitlichte die EU das Lebensmittelrecht, um insbesondere die Effektivität und Kohärenz der Kontrollen zu verbessern.[60] Die aktuelle Kontrollverordnung verpflichtet die nationalen Überwachungsbehörden dazu, „alle Unternehmer regelmäßig risikobasiert und mit angemessener Häufigkeit amtlichen Kontrollen" zu unterziehen.[61] Diesen „risikobasierten" Ansatz

56 §§ 22 Abs. 2, 15 Abs. 2, 4 Abs. 1 Nr. 1 GastG; siehe *Ehlers*, in: ders./Fehling/Pünder (Hrsg.), Besonderes Verwaltungsrecht, Bd. 1, 4. Aufl. 2019, Rn. 59; *Metzner/Thiel*, in: dies. (Hrsg.), Gaststättenrecht 7. Aufl. 2023, § 22 Rn. 19.

57 § 22 Abs. 2 GastG, § 40 VwVfG; so auch *Metzner/Thiel*, in: dies. (Hrsg.), Gaststättenrecht 7. Aufl. 2023, § 22 Rn. 18, 23.

58 Bei Gefahranzeigen kann sich das Ermessen auf Null reduzieren.

59 Zu der Divergenz hierbei siehe jedoch *Ho*, UC Irvine L. Rev. 7 (2017), S. 401 (410–417).

60 Verordnung (EG) Nr. 178/2002 des Europäischen Parlaments und des Rates v. 28. Jan. 2002 zur Festlegung der allgemeinen Grundsätze und Anforderungen des Lebensmittelrechts, zur Errichtung der Europäischen Behörde für Lebensmittelsicherheit und zur Festlegung von Verfahren zur Lebensmittelsicherheit. ABl. L 31 v. 1. Feb. 2002, S. 1; siehe *Borraz* et al., Regulation & Governance 16 (2022), S. 274 (275, 278).

61 Art. 9 Abs. 1 Verordnung (EU) 2017/625 des Europäischen Parlaments und des Rates v. 15. März 2017 (Verordnung über amtliche Kontrollen), ABl. L 095 v. 7. April 2017, S. 1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

formt in Deutschland eine Verwaltungsvorschrift genauer aus.[62] Sie benennt quantitative und qualitative Faktoren, die das potentielle Schadensausmaß und die Wahrscheinlichkeit des Schadenseintritts betreffen.[63] Hierzu gehören die Risikoeinstufung der Produkte sowie frühere Hygieneverstöße des Betriebsinhabers.[64] Auf dieser Basis ermitteln die Behörden einen betriebsspezifischen Risiko-Wert, der die Kontrollfrequenz bestimmt, die von mindestens wöchentlich bis zu dreijährlich reicht und neben anlasslose Kontrollen tritt.[65] Zwar weisen Lebensmittelkontrollen aufgrund der verbleibenden Spielräume sowie der unterschiedlichen Ausstattung der Aufsichtsbehörden eine geographische und sektorspezifische Varianz auf.[66] Allerdings ist die Steuerungsdichte aufgrund der normativ verankerten Kontrollkriterien wesentlich größer als im Gaststättenrecht.

III. Durchsetzungsdefizite

Nach derzeit herrschendem Rechtsverständnis müssen sich die Behörden für ihre Kontrollentscheidungen kaum gerichtlich verantworten. Das öffentliche Wirtschaftsrecht gewährt den Verbrauchern in der Regel keine subjektiven Rechte, um effektive Kontrollen zu fordern.[67] Ihr Schutz gilt allein als Rechtsreflex der Überwachungskompetenzen.[68] Auch die kontrollierten Unternehmen können umgekehrt nicht geltend machen, dass andere Betriebe systematisch weniger häufig oder weniger intensiv kon-

62 Allgemeine Verwaltungsvorschrift über Grundsätze zur Durchführung der amtlichen Überwachung der Einhaltung der Vorschriften des Lebensmittelrechts, des Rechts der tierischen Nebenprodukte, des Weinrechts, des Futtermittelrechts und des Tabakrechts (AVV Rahmen-Überwachung) v. 20. Jan. 2021, BAnz AT 26.01.2021, B6.

63 § 7 i. V. m. Anlage 1 AVV Rahmen-Überwachung, z. B. in Form der „Art und Anzahl aller verwaltungsrechtlichen Maßnahmen (Ordnungsverfügungen, Beschränkungen oder Widerruf von Zulassungen, Bußgeldverfahren, Strafverfahren) innerhalb eines bestimmten Zeitraumes".

64 Ebd.

65 § 7 Abs. 4 AVV Rahmen-Überwachung.

66 Siehe Foodwatch (Fn. 23), Kontrolle ist besser, 2019; auf Unionsebene siehe *Borraz* et al., Regulation & Governance 16 (2022), S. 274.

67 Siehe nur *Huber*, in: Voßkuhle/Eifert/Möllers (Hrsg.), Grundlagen des Verwaltungsrechts, Bd. II, 3. Aufl. 2022, S. 1313 (Rn. 111 f.); *Wahl*/*Schütz*, in: Schoch/Schneider (Hrsg.), VwGO, Stand: 47. EL Feb. 2025, § 42 Rn. 227.

68 Vgl. ausdrücklich § 4 Abs. 4 FinDAG.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

trolliert werden, sofern nicht ein klarer Fall von Willkür vorliegt.[69] So entgehen systematische Kontrolldefizite und Kontrollexzesse einer gerichtlichen Prüfung.

Die Ausgestaltung der Kontrollen wird vielmehr dem Verwaltungsinnenrecht zugerechnet. § 22 VwVfG stellt es im Gaststättenrecht in das Aufgreifermessen der zuständigen Behörde, zu entscheiden, welches konkrete Verwaltungsverfahren sie gegen wen eröffnet. Dogmatisch wird dieser Entscheidung kaum Bedeutung zugemessen, weshalb sie nicht weiter rechtlich ausgeformt ist.[70] Die Kontrolle selbst beruht auf den soeben zitierten gesetzlichen Ermächtigungen im Gaststätten- und Lebensmittelrecht. Sie sind materiell faktisch voraussetzungslos und erlauben auch stichprobenartige Routinekontrollen.[71] Hat die Behörde aber einen Betrieb als kontrollbedürftig eingeordnet, so liegen mit dieser Vorentscheidung auch die tatbestandlichen Voraussetzungen einer Kontrolle vor. Deshalb fallen das Aufgreifermessen, das Verfahren gegenüber einem bestimmten Unternehmen zu eröffnen, und das Auswahlermessen, die Kontrolle als Verfahrensabschluss durchzuführen, praktisch zusammen. Aufgrund dieser Pfadabhängigkeit kommt bereits der Verfahrenseröffnung eine wesentlich größere Bedeutung zu als das Recht reflektiert.

D. Konzeptpflichten als Lösungsansatz

Welche Regulierungsperspektive eröffnet nun das öffentliche Wirtschaftsrecht?

I. Ableitung einer Konzeptpflicht aus §§ 22, 40 VwVfG

Dogmatisch bietet die verwaltungsrechtliche Regulierung von Ermessensentscheidungen einen Ansatz, die behördliche Entwicklung von Kontroll-

69 *Metzner/Thiel*, in: dies. (Hrsg.), Gaststättenrecht, 7. Aufl. 2023, § 22 Rn. 23; für das Kartellrecht EuG, Urt. v. 17. Dez. 2003, Rs. T-219/99, ECLI:EU:T:2003:343, Rn. 65 f. – *British Airways/Kommission*.

70 *Kreifels*, Die Prioritätensetzung der Europäischen Kommission beim Aufgreifen kartellrechtlicher Fälle, 2019, S. 21.

71 Insbesondere bedarf es keines konkreten Verdachts, siehe nur *Metzner/Thiel*, in: dies. (Hrsg.), Gaststättenrecht, 7. Aufl. 2023, § 22 Rn. 7, 18; vgl. zur verdachtslosen Kontrolle im Polizeirecht *Waechter*, DÖV 1999, S. 138 (142); *Möllers*, NVwZ 2000, S. 382 (386).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

konzepten auch jenseits bereits bestehender unionsrechtlicher Vorgaben, wie im Lebensmittelrecht, *de lege ferenda* zu fordern. Diese Kontrollkonzepte bieten eine formalisierte Anleitung für die behördliche Auswahl der Kontrollsubjekte und der Kontrollfrequenz.[72] Der normative Ausgangspunkt liegt in §§ 22 und 40 VwVfG, die das Aufgreifermessen sowie das Auswahlermessen normieren. In Verbindung mit dem Willkürverbot des Art. 3 Abs. 1 GG bedarf die Verwaltung ohne weitere Vorgaben bisher lediglich eines sachlichen Grundes für die Eröffnung eines Verfahrens sowie für die spätere Auswahl eines Maßnahmeadressaten.[73] Aufgrund der voraussetzungslosen Kontrollermächtigungen ist bereits der Verweis auf den Routinecharakter einer Kontrolle ein ausreichender sachlicher Grund.

Allerdings hat die Verwaltungsrechtsdogmatik jedenfalls für das Auswahlermessen die Figur der „Selbstbindung der Verwaltung" teilweise zu einer Konzeptpflicht verdichtet.[74] Möchte die Verwaltung zum Beispiel gegen einen von mehreren sog. Schwarzbauten vorgehen, fordern die Gerichte das Vorliegen eines Konzeptes für die Auswahl.[75] Ähnlich begründet *Christoph Möllers*, dass die Durchführung anlassloser Polizeikontrollen eines Konzeptes bedarf.[76] Eine vergleichbare Rechtspflicht zum Einsatz evidenzbasierter Risikobewertungskonzepte statuierte zuletzt der Europäische Gerichtshof für Menschenrechte in Bezug auf die polizeiliche Risikobeurteilung von Gewalttätern.[77] In all diesen Fällen ist das Verwaltungshandeln inhaltlich kaum durch gesetzliche Vorgaben determiniert.[78]

72 Grundlegend *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 197–216; *Herzmann*, VerwArch 2013, S. 429 (433–445).

73 *Schmitz*, in: Stelkens/Bonk/Sachs (Hrsg.), VwVfG, 10. Aufl. 2023, § 22 Rn. 6, 12; *Geis*, in: Schoch/Schneider (Hrsg.), VwVfG, Stand: 7. EL Mai 2025, § 40 Rn. 68 f.; für das Gaststättenrecht *F. Wollenschläger*, in: Schmidt/ders. (Hrsg.), Kompendium Öffentliches Wirtschaftsrecht, 2016, S. 51 Rn. 88–93; weitergehend mit Blick auf Konzeptpflichten aber *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 211 f.

74 Siehe hierzu *Schmidt-Aßmann*, Das allgemeine Verwaltungsrecht als Ordnungsidee, 2. Aufl. 2006, S. 334 f. Rn. 99; mit weiteren Beispielen *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 200 f.

75 Siehe m. w. N. zur Rspr. *Sachs*, in: Stelkens/Bonk/Sachs (Hrsg.), VwVfG (Hrsg.), 10. Aufl. 2023, § 40 Rn. 92.

76 *Möllers*, NVwZ 2000, S. 382 (386 f.).

77 EGMR, Urt. v. 15. Juni 2021, Nr. 62903/15, Rn. 167–176 – *Kurt v. Austria*.

78 Die „zurückgenommene Programmsteuerung" betont *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 216; siehe zum Regulierungsrecht *Broemel*, JZ 69 (2014), S. 286.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die Eingriffsqualität schafft jedoch ein Bedürfnis danach, die Auswahl der Kontrollierten evidenzbasiert zu steuern.

Ich argumentiere nun dafür, den Ermessensvorgaben auch für die Durchführung von Kontrollen im nationalen öffentlichen Wirtschaftsrecht eine Konzeptpflicht zu entnehmen, jedenfalls wenn die Eingriffs- und Schutzwirkung der Kontrolle eine bestimmte Erheblichkeit überschreitet.[79] Anknüpfungspunkt ist der Umstand, dass die Entscheidung über die Eröffnung eines Verwaltungsverfahrens funktional äquivalent zur eigentlichen Eingriffsentscheidung in Form der Kontrolle selbst ist. Obwohl Kontrollen im öffentlichen Wirtschaftsrecht für die Kontrollierten wie für die Geschützten grundrechtsrelevant sind, sind sie inhaltlich kaum vorbestimmt und unterliegen Einflüssen, die ihre Effektivität und Gleichmäßigkeit beeinträchtigen können. Deshalb ist auch für diese Kontrollen im Rahmen des Ermessens eine Konzeptpflicht abzuleiten. Dies gilt insbesondere dann, wenn die Behörde durch den Einsatz automatisierter Datenanalysen bestimmte Weichenstellungen für ihre Kontrolltätigkeit selbst vorab festlegt.[80] Eine solche Konzeptpflicht wäre justiziabel, denn jeder Adressat einer Kontrolle könnte das Fehlen eines Konzepts oder aber seine diskriminierende Ausrichtung gerichtlich geltend machen.[81] Die Kontrollkonzepte fungieren dann als behördliche Entscheidungsanleitung nach innen und Selbstbindung nach außen.[82]

II. Rationalität einer Konzeptpflicht

Eine Konzeptpflicht fordert eine behördliche Eigenprogrammierung[83], die inhaltlich an dem Zweck der Kontrollen auszurichten ist. Es erscheint sinnvoll, sie ausdrücklich gesetzlich zu verankern, und dabei bestimmte Rahmenbedingungen vorzugeben. Als Vorbild können neben der

79 Ebenso für die personenbezogene Prävention *Rusteberg*, in: Münkler (Hrsg.), Dimensionen des Wissens im Recht, 2019, S. 233 (263).

80 Hierzu *Egidy*, in: Rose-Ackerman (Hrsg.), Public Administration and Expertise in Democratic Governments, 2024, S. 46 (57–59); *Coglianese/Lehr*, Georgetown L. J. 105 (2017), S. 1147 (1169); zum Sonderfall der steuerrechtlichen Kontrolle *Hashimzade* et al., Journal of Economic Behavior & Organization 124 (2016), S. 130.

81 *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 215 f.; *Broemel*, JZ 69 (2014), S. 286 (293 f.).

82 *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 197 f.

83 *Eifert*, VVDStRL 67 (2008), S. 286 (317).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lebensmittelrechtlichen Ausformung die gesetzlichen Konzeptpflichten zur Überwachung des Marktes für Tabakerzeugnisse[84] sowie zur Prüfung von Steuererklärungen[85] im nationalen Recht dienen. Die Rationalität der Konzepte folgt der zu Beginn entwickelten ökonomischen Analyse.

Ein taugliches Steuerungskonzept sollte folgende drei Grundanforderungen erfüllen. Erstens muss das Konzept einen Ausgleich zwischen Genauigkeit und Gleichheit finden, der mit Art. 3 GG vereinbar ist. Besonders relevant sind die in das Konzept einfließenden Daten, aus denen sich ein Diskriminierungspotential ergeben kann. Zweitens sollte ein Konzept weithin resistent gegen strategische Manipulation sein, sowohl durch die Kontrollsubjekte selbst als auch durch ihre Konkurrenz. Hierfür kann eine Pflicht zur Geheimhaltung der Kontrollkriterien beitragen, wie sie etwa § 88 Abs. 5 S. 4 Abgabenordnung für die automatisierte Identifizierung fehlerhafter Steuererklärungen vorschreibt. Drittens kann ein in das Konzept integrierter Zufallsmechanismus, wie ihn § 88 Abs. 5 Abgabenordnung vorsieht, eine gleichmäßigere Kontrollwirkung sichern und zusätzliche Daten für eine kontinuierliche Prüfung und Fortentwicklung des Konzeptes erzeugen.

III. Subjektive und objektive Prüfung der Kontrollkonzepte

Die so ausgestalteten Kontrollkonzepte lassen sich allerdings erst durchsetzen, wenn sich das kontrollierte Unternehmen als Adressat eines Eingriffs gegen die Kontrolle wehrt. Das zuständige Gericht müsste dann das behördliche Kontrollkonzept überprüfen, ggf. mit Sachverständigen und bei besonderen Geheimhaltungsbedürfnissen womöglich gar in einem *in camera* Verfahren.[86] Allerdings wären selbst standardisierte Kontrollkonzepte in jedem einzelnen Gerichtsverfahren neu zu prüfen. Auch würden sie weiterhin keine Schutzwirkung für Verbraucher entfalten.

Vor diesem Hintergrund besteht ein Bedarf für eine ergänzende objektiv-rechtliche Kontrolle. Eine solche könnte der Gesetzgeber *de lege ferenda* normieren. Einen Ansatz hierfür enthält § 28 Abs. 1 S. 3 Tabaker-

84 § 28 Abs. 1 Tabakerzeugnissegesetz.

85 § 88 Abs. 5 AO.

86 Ähnlich auch *B. Wollenschläger*, Wissensgenerierung im Verfahren, 2009, S. 199; zum Verfahren *in camera* siehe § 99 Abs. 2 VwGO; hierzu im Überblick *Neumann*, DVBl 2016, S. 473; *Häußler*, DVBl 2024, S. 1333.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

zeugnissegesetz, der eine regelmäßige Überprüfung und Bewertung der Wirksamkeit des vorgeschriebenen Überwachungskonzepts, mindestens alle vier Jahre, vorsieht.

E. Fazit

Kontrollkonzepte rationalisieren verwaltungsinterne Entscheidungsprozesse und ermöglichen eine Steuerung der Kontrollentscheidungen im öffentlichen Wirtschaftsrecht. Die empirischen Sozialwissenschaften bieten das hierzu notwendige Instrumentarium. Dies wirft die Frage auf, wie weit diese Verrechtlichung des Verwaltungsinnenrechts vordringen darf. Die bisher geforderte Flexibilität der Verwaltung steht im Konflikt mit dem Anliegen einer stärkeren Regulierung der Kontrollen im öffentlichen Wirtschaftsrecht. Allerdings ist zu beachten, dass die Verwaltung ihre Auswahlkriterien und Kontrollmaßstäbe häufig bereits selbst vorab festlegt, insbesondere wenn sie Risikobewertungen mit Hilfe von Analysesoftware durchführt. Hier ließe sich das Potenzial einer empirisch informierten Steuerung des Ermessens zum Wohl der Schutzsubjekte wie der Kontrollierten verwirklichen.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

„Heterogenität als Problem für Steuerung“, Einzelfallgerechtigkeit und das Referenzklassenproblem im Kartellschadensersatzrecht

Alexander Morell

Meinen Weg in die Gruppe von Christoph Engel am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern fand ich über die „Heterogenität als Problem für Steuerung“. Ich war neugierig geworden auf diesen Max-Planck-Direktor, der Jurist ist, aber rechnet und auch Ameisenforscher in sein Seminar lädt. Der Tristesse der Examensvorbereitung beschloss ich also, mit der reichen Abwechslung zu begegnen, die ein Seminar bei diesem vielseitigen Denker versprach. Ich erhielt das Thema „Heterogenität als Problem für Steuerung“ – und verstand es nicht. Dass Christoph Engel mein stetes Bemühen dennoch mit einem erfreulichen Seminarschein honorierte, ermunterte mich nach überwundenem Examen zu einer Promotionsanfrage. Seitdem hatte ich am Institut immer wieder Gelegenheit, darüber nachzudenken, was Christoph Engel sich von diesem Thema erhofft hatte, auf das er immer wieder zurückgekommen ist.

Kürzlich meine ich, das Thema in einem angewandten Kontext wiedererkannt zu haben: In den Problemen um den Beweis der Kartellschadenshöhe. Der Rahmen meines Aufsatzes dazu[1] ließ für die Meta-Gedanken, denen man am MPI in Bonn wie an keinem anderen Ort im akademischen Betrieb nachgehen konnte, keinen Raum, so dass ich die Gelegenheit nutze, meine unfertigen Hintergrundüberlegungen hier zu platzieren. Vielleicht findet Christoph Engel Freude am Prozess, den er ausgelöst hat und vielleicht auch an einigen Gedanken, die dieser Prozess hat entstehen lassen.

1 Morell 2025a, 2025b.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

A. Heterogenität als Problem für die Bemessung des (steuernden) Kartellschadensersatzes

Der Kartellschadensersatz dient nach heute wohl allgemeiner Ansicht zumindest auch der Verhaltenssteuerung (Prävention, private Rechtsdurchsetzung).[2] Der Beweis des Kartellschadens ist allerdings schwer.[3] Niemand kümmert sich um den Ersatz des dead weight loss.[4] Den zu bemessen, ist zu kompliziert. Der Fokus liegt auf der sogenannten „overcharge", den kartellbedingten Mehrkosten, die Kunden während der Kartelldauer für die tatsächlich abgenommene Menge bezahlen mussten. Auch die Bemessung dieses Schadens führt regelmäßig zu wenig ergiebigen „Gutachterschlachten"[5], die vor allem die Gerichte, aber wohl auch die Kläger, gern vermeiden würden. Die Beklagten mögen diese „Gutachterschlachten" als Mittel zur Abschreckung schätzen, aber deren Kosten bezahlen auch sie. Schön wäre es daher, könnte man sich bei der von § 287 ZPO grundsätzlich eröffneten Schätzung der „Overcharge" behelfen, indem man – ggf. prima facie – einen mittleren Schaden unterstellte.[6] Studien dazu liegen vor. Der Medianschaden von (aufgedeckten) Kartellen liegt bei zwischen 16 % und 18 % des unter dem Kartell bezahlten Preises.[7]

Zumindest noch[8] verschließen sich Gerichte aber der Schätzung der Schadenshöhe unter direktem Rückgriff auf die Metastudien zu Kartellschäden. Eine Begründung ist die Heterogenität von Kartellen. Kartelle seien zu „vielgestaltig"[9], als dass aus dem Aggregat auf den „Einzelfall" geschlossen werden könne.[10] Und da ist sie, die Heterogenität als Problem für den (auch steuernden) Kartellschadensersatz. Ist es tatsächlich unzulässig, bezüglich der Schadenshöhe zu typisieren, weil die Grundgesamtheit, aus der man abstrahiert, „vielgestaltig" ist?

2 Weitbrecht 2017, S. 1574; Franck 2024, Rn. 3–4; Roth 2016, S. 1136.
3 Galle 2016; Stancke 2018, Vor Fn. 52; Martinek 2023, S. 144; Weitbrecht 2024, S. 170; Hornkohl 2020, S. 661; Roth 2018, S. 29.
4 Oxera und Komninos 2009, S. 15.
5 Heusel et al. 2024, S. 386; Schweitzer und Woeste 2022, S. 72.
6 Das versuche ich in Morell 2025b zu begründen.
7 Morell 2025b, S. 198.
8 Die Tür zu einer derartigen Schätzung hat jüngst der BGH mit seinem LKW IV-Urteil aufgestoßen, BGH, Urteil vom 09.07.2024, KZR 98/20.
9 BGH, Urteil vom 11.12.2018, 57 (KZR 26/17)
10 BGH, Urteil vom 11.12.2018, 57; BGH, Urteil vom 28.01.2020, S. 41.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

B. Heterogenität als Problem für Steuerung und das Referenzklassenproblem

Das Problem der Heterogenität als Problem für Steuerung ist eng verwandt, vielleicht gar identisch, mit dem Problem der Referenzklassenauswahl[11] zur Bildung von Anfangswahrscheinlichkeiten.[12]

Will man aufklären, ob eine Tatsache vorliegt (ein durch das Kartell verursachter Mindestschaden von x, zum Beispiel), behilft man sich zur Vermeidung von Widersprüchen bei der Auswertung der Evidenz sinnvollerweise der Theorie der (subjektiven) Wahrscheinlichkeit: Man bildet eine Anfangswahrscheinlichkeit, aktualisiert diese durch die Integration von Indizien nach Bayes Regel („updating“) und prüft, ob die Endwahrscheinlichkeit über der kritischen Wahrscheinlichkeit liegt, die das gesetzliche Beweismaß zum Kriterium eines gelungenen Beweises erklärt.[13] Weil man oft nicht sehr viele Indizien hat, bleibt die Anfangswahrscheinlichkeit in vielen Fällen für die Endwahrscheinlichkeit relevant. Auch die (subjektive) Anfangswahrscheinlichkeit muss daher sinnvoll und sorgfältig bestimmt werden. Dazu sollte sie an (objektiven) historischen Häufigkeiten des gesuchten Merkmals in einer für den Entscheidungskontext relevanten Referenzklasse bemessen werden.[14]

Für den Kartellschadensersatz hieße das: Zunächst geht man davon aus, dass ein Kartell mit einer gerade überwiegenden Wahrscheinlichkeit eine Overcharge von mindestens 16 % verursacht hat.[15]

C. Die Schranke der Einzelfallgerechtigkeit als Verbot übermäßiger Abstraktion

Dieses Vorgehen lässt bei vielen Juristen die Alarmglocken schrillen. Die Kartelle in den Metastudien seien eben „zu vielgestaltig“.[16] Sie unterschieden sich zu sehr in den Overcharges, die für sie gemessen wurden, und den

11 Schweizer 2015, S. 258–266; Morell 2022, S. 123–128.

12 Ich spreche im Folgenden der Anschaulichkeit halber nur von „Anfangswahrscheinlichkeit“. Das hier geäußerte gilt genauso für die Bestimmung des Gewichts von Indizien und Beweismitteln aus der Lebenserfahrung (Bayesianisch: „Likelihoods“).

13 Schweizer 2015, S. 287; Morell 2022, S. 269.

14 Morell 2022, S. 95.

15 Morell 2025b.

16 BGH, Urteil vom 11. 12. 2018, 57; Inderst und Thomas 2021, S. 434.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Eigenschaften, die sie auszeichneten, als dass man sie alle, und sei es nur im Ausgangspunkt, gleich behandeln könnte. Scherte man alle Kartelle – und sei es nur in erster Näherung – über einen Kamm, so werde man dem „Einzelfall" nicht „gerecht"![17]

Ist das ein valider Einwand gegen den (vorläufigen) Schluss von allen vermessenen Kartellen auf ein Kartell, der die subjektive Beweislast für geringere Schäden auf den beklagten Kartellanten wechseln ließe? Dafür ist kurz zu klären, was Einzelfallgerechtigkeit bedeutet. Die Einzelfallgerechtigkeit ist ein Begriff, der in der Forschungsgruppe von Christoph Engel keine Prominenz genoss. Für die Beobachtung nur eines einzelnen Falls kann man keine Signifikanztests rechnen und keine Graphen zeichnen. Kurz: Die Freude, die man mit einem Einzelfall haben könnte, schien uns begrenzt. Außerhalb des Instituts scheint die Einzelfallgerechtigkeit in der Welt der Juristen aber so etwas wie der Goldstandard zu sein. Eigentlich überraschend für einen Begriff, dessen Wortsinn ein Oxymoron darstellt. „Gerecht" ist es, Gleiches gleich und Ungleiches ungleich zu behandeln.[18] Zur Feststellung von Gerechtigkeit müssen also mindestens zwei Fälle betrachtet und verglichen werden[19] – und in der Regel noch viele mehr. Wie kann dann die Betrachtung nur eines „einzelnen" Falls eine Aussage über Gerechtigkeit erlauben? Ein Verständnis, nach dem man im Dienst der Einzelfallgerechtigkeit die richtige Lösung nur aus dem konkret betrachteten Fall herausarbeitet, kann also nicht richtig sein. Es ist widersprüchlich.

Sinnvoller beschreibt Gabriele Britz in einer der wenigen Abhandlungen zum Thema die Einzelfallgerechtigkeit als Verbot einer übermäßigen Generalisierung.[20] Dabei weist sie darauf hin, dass man, wenn man die Rechtsfolge für einen heterogenen Kreis von Normadressaten generalisiert (Berufsende mit 65, statt Berufsende für alle, die nicht mehr leistungsfähig sind), wissentlich Fälle auf eine Weise behandelt, die das Normziel missbilligt (man versetzt leistungsfähige 66-Jährige in den Ruhestand).[21] Sie nennt diese Fälle „Generalisierungsunrecht".[22] Obwohl sie

17 Coppik und Heimeshoff 2021, S. 67; Mallmann und Leinemann 2023, S. 50; BGH, Urteil vom 28.01.2020, S. 41.

18 Aristotles 1911, S. 95.

19 Aristotles 1911, S. 95.

20 Zusammenfassend Britz 2008, S. 212.

21 Britz 2008, S. 12–13.

22 Britz 2008, S. 3. Sie sieht Verstöße gegen die Einzelfallgerechtigkeit auch in zwei weiteren Konstellationen: Das Anknüpfen an generalisierende Umständen könne auch

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

schon von Unrecht spricht, wenn das Telos die Entscheidung des Falls nicht deckt, will sie über die Frage der Rechtsmäßigkeit von Generalisierung erst in einer Verhältnismäßigkeitsprüfung – wieder im Einzelfall – entscheiden.[23]

Zum Maßstab zulässiger Generalisierung kann man mehr sagen, als dass sie verhältnismäßig sein sollte. Die Definition von Britz ist wie für einen Fehlerkostenansatz gemacht.[24] Eine Entscheidung, die mit dem Telos nicht übereinstimmt, ist eine im Hinblick auf das Regelungsziel fehlerhafte Entscheidung. Weil das Regelungsziel hoffentlich sinnvoll ist, hat sie Kosten. Zugleich hat die Generalisierungen einen Nutzen. Britz nennt vor allem die „Vereinfachung“.[25] Doch die Dinge liegen grundsätzlicher.[26] Generalisierung ist ein Mittel zum Umgang mit Unwissenheit. Wer nicht weiß, welche Gesundheitskosten ein Versicherter im Laufe seines Lebens akkumulieren wird, generalisiert und schließt aus seiner Erfahrung mit ähnlichen Fällen auf eine unsichere, aber nützliche Vorhersage. Der Richter, der bei Prozessbeginn nicht weiß, ob der Beklagte das Rennrad des Klägers wie behauptet absichtlich zerstört hat, generalisiert auch, obwohl er die Tatsachen des Einzelfalls aufklären soll. Er bildet sich eine Überzeugung über die Grundwahrscheinlichkeit, dass Menschen an einem bestimmten Ort einander Rennräder zerstören (oder setzt sie „neutral“ auf 50:50) und wertet dann den Beweiswert der Indizien und Beweismittel anhand seiner Lebenserfahrung. Wie häufig haben sich Parteien wirklich gestritten, wenn (nur) zwei Zeugen mit persönlicher Bindung zum Kläger das behaupten? Wie häufig zerstört nach einem Streit über einen mittleren Geldbetrag unter Freunden der eine absichtlich das Eigentum des anderen? Usw. Wer Beweise rational würdigt, der generalisiert. Erkenntnis ist ohne Generalisierung ausgeschlossen. Erkenntnis ist Induktion und In-

ungewollte und damit ungerechte Verteilungsfolgen haben, Britz 2008, S. 113. Ferner erschwerten Generalisierungen dem generalisierend Behandelten die ihm durch das APR gewährleistete freie Selbstdarstellung, Britz 2008, S. 179–209. Beide in der Einzelfallgerechtigkeit zum Ausdruck kommenden Wertungen gegen Generalisierung spielen im Kontext des formal angelegten Beweisrechts, das hauptsächlich epistemische Ziele verfolgt, nicht materiell verteilt und den Parteien keine Darstellungsrechte einräumt, sondern sie bei der Wahrheit hält, eine untergeordnete Rolle.

23 Zusammenfassend Britz 2008, S. 212.

24 Ausdrücklich Britz 2008, S. 167–168.

25 Britz 2008, S. 158.

26 Klingt bei Britz immerhin an: „Ermöglichung (…) von Informationsgewinnung“, Britz 2008, S. 158.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

duktion erweitert bekanntes Wissen auf unbekannte Fälle. Durch Generalisierung. Es ist bekannt, dass man sich bei diesen Schlüssen vertun kann. Das ist kein „Generalisierungsunrecht". Es ist schlicht ein Fehler und der Schließende muss sich Rechenschaft darüber ablegen, wie schlimm ein Fehler wäre und ob es sich angesichts dieser Folgen lohnen könnte, Weiteres gegen Fehler zu unternehmen.

Erkenntnis erfordert also Abstraktion. Aber auf einem ausreichend hohen Abstraktionsniveau sind alle Fälle gleich. Da bedarf es eines Stoppschilds. Der Anspruch, Gleiches gleich und Ungleiches ungleich zu behandeln, kann nur dann relevant bleiben, wenn er nicht durch Abstraktion umgangen wird. Der Gedanke der Einzelfallgerechtigkeit scheint mir die Funktion dieses Stoppschilds zu erfüllen. Aber wann abstrahiert man „zu hoch"? In Konkretisierung des Fehlerkostenansatzes von Gabriele Britz: Wenn der Nutzen zusätzlicher Abstraktion kleiner ist als deren Kosten.

D. Einzelfallgerechtigkeit als Nutzung der größten homogenen Referenzklasse

Auch hier hilft das Denken in Referenzklassen. Für die Auswahl der Referenzklasse, an der man die Anfangswahrscheinlichkeit (oder auch die Beweiswerte von Indizien) bestimmt, gibt es keine allgemeingültigen, aber immerhin akzeptierte Regeln.[27] Vieles scheint dafür zu sprechen, auf die engste Referenzklasse abzustellen. Die engste Referenzklasse ist die Klasse von Elementen, die die meisten Merkmale mit dem betrachteten Fall teilt. Dieses Kriterium entspricht allerdings der Einzelfallgerechtigkeit im Sinne des Oxymorons. Wenn man im Kartellschadensrecht die Overcharge eines dreijährigen Gebietskartells von sechs deutschen Produzenten von Nutzfahrzeugen bestimmen will, die vor allem in Süddeutschland verkaufen, müsste man schauen, zu welchem Schaden in der Vergangenheit Gebietskartelle über Nutzfahrzeuge in Süddeutschland im Mittel geführt haben, die zwischen zwei und vier Jahren aktiv waren und in einem weiten Oligopol von sechs Kartellanten praktiziert wurden. Ein solches Kartell existiert wahrscheinlich nur einmal – im zu entscheidenden Fall. Ein zweites zum Vergleich wird man in den Metastudien nicht finden, geschweige denn eine Anzahl gleichartiger, über die man in Ausnutzung des

27 Morell 2022, S. 123–128.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Gesetzes der großen Zahl einen verlässlichen Durchschnittswert bilden könnte. Die engste einschlägige Klasse ist der Fall selbst. Aber dieser Fall ist keine Referenz. Abstraktion ist unumgänglich, um sinnvolle Anfangswahrscheinlichkeiten zu bilden.

Die Eigenschaften zur weiteren Eingrenzung der Referenzklasse müssen daher auf Relevanz für das gesuchte Merkmal (im Beispiel: den Mindestschaden) untersucht werden. Eigenschaften, die für den Schluss auf das gesuchte Merkmal irrelevant sind, sollten die Referenzklasse nicht verkleinern. Die wohl überzeugendste Regel zur Auswahl der Referenzklasse ist daher die Anweisung, auf die größte homogene Referenzklasse abzustellen. Homogen sind Elemente einer Referenzklasse, wenn sie sich nicht durch Eigenschaften unterscheiden, die für den Schluss auf das gesuchte Merkmal relevant sind. Diese Regel ist mit dem Gedanken der Einzelfallgerechtigkeit im Sinne des notwendigen Stoppschilds kompatibel und scheint mir die recht verstandene Einzelfallgerechtigkeit zu erhellen. Es kommt nicht auf den Blick auf den einen Fall an. Es muss abstrahiert werden. Aber grundsätzlich nur so weit, wie dadurch keine relevante Information über das gesuchte Merkmal verloren geht.[28]

E. Einzelfallgerechtigkeit bei Unerreichbarkeit einer homogenen Referenzklasse: Ein Kosten-Nutzen-Ansatz

Unglücklicherweise sind die Häufigkeiten der gesuchten Merkmale in homogenen Referenzklassen oft nicht bekannt und alle Häufigkeiten, mit denen man arbeiten kann, entstammen bekanntermaßen heterogenen Referenzklassen. Hier erst wird die Auswahl der Referenzklasse zum wirklichen Problem – für das es keine eindeutige konzeptionelle Lösung gibt.[29] Das Beispiel vom schwedischen Lourdes-Pilger illustriert es anschaulich, wenn auch vielleicht nicht scharf genug: Nach Lourdes pilgern gemeinhin Katholiken. In Schweden wohnen ganz überwiegend Protestanten. Wo liegt nun die Anfangswahrscheinlichkeit, dass der schwedische Lourdespilger katholisch ist? Stellt man auf die Schweden ab oder auf die Lourdespilger? Die (relativ) größte (wenn auch absolut wohl kleine) Referenzklasse wäre die Klasse aller schwedischen Lourdespilger, denn sowohl

28 Morell 2022, S. 123–128.
29 Morell 2022, S. 123–128.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

die schwedische Herkunft als auch die Pilgerschaft tragen Information über die Religionszugehörigkeit. Aber die Häufigkeit des Merkmals „katholisch" in dieser Klasse ist nicht bekannt.[30]

Ähnlich liegen die Dinge auch bei der Bestimmung des Kartellschadens. Der Medianschaden für alle bekannten Kartelle liegt bei 16 %. Die Elemente dieser großen Referenzklasse unterscheiden sich klar in Eigenschaften, die für das gesuchte Merkmal relevant sind (Weite des Oligopols, Marktabdeckung des Kartells, Art der Absprache, …). Die große Referenzklasse, für welche die Häufigkeit des Merkmals bekannt ist, ist heterogen. Aber für viele kleinere Referenzklassen (wie die süddeutscher Nutzfahrzeugsgebietskartelle mit sechs Mitgliedern) ist die Häufigkeit des gesuchten Merkmals „Mindestschaden in Höhe von x" unbekannt oder basiert auf so wenigen Beobachtungen, dass man sich auf den Zusammenhang zwischen mittlerer Tendenz des Schadens und den geteilten Eigenschaften der Kartelle nicht verlassen kann. Das Referenzklassenproblem wird üblicherweise als Frage beschrieben, welche Wahrscheinlichkeit einer Tatsachenbehauptung prima facie zuzuschreiben ist, d. h. welcher von vielen verfügbaren Referenzklassen die Häufigkeit zu entnehmen ist, die man als Anfangswahrscheinlichkeit behandelt. Es ist in dieser Diskussion meist klar, dass man eine der Referenzklassen wählen wird. Die Schärfe gerade des praktischen Kartellrechtsschadensproblems ergibt sich daraus, dass die Gerichte nun gar keine Aussage über einen Mindestschaden treffen wollen. Damit machen sie faktisch (aus Gründen der Beweislast) die Aussage, der Schaden liege (prima facie) bei null, obwohl diese Aussage nun von gar keiner der verfügbaren Referenzklassen gestützt wird. Ist das von der Einzelfallgerechtigkeit geboten? Welche Orientierung bietet die Einzelfallgerechtigkeit in einem solchen Kontext?

Die Einzelfallgerechtigkeit schützt nicht vor jeder, sondern nur vor übertriebener Abstraktion. Wann eine „Übertreibung" vorliegt, bemisst sich an Gründen, die dem Recht entnommen werden müssen. Im Beweisrecht müssen es angesichts des Beibringungsgrundsatzes, welcher der zutreffenden Tatsachenermittlung dient,[31] zunächst epistemische Gründe sein. Die Homogenität ist das einfache Kriterium: Abstraktion, welche die Erkenntnis nicht verwässert, ist immer zulässig. Im Problem um die Nutzung von Meta-Studien zum Kartellschaden wird aber deutlich, dass

30 Rüßmann 1987, Rn. 9.

31 Morell 2022, S. 325–326.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Abstraktion notwendig ist, um Wissen überhaupt verfügbar zu machen. Wenn genaueres, fallspezifisches Wissen nicht vorliegt, muss man sich fragen, ob man aufgrund seines allgemeinen Wissens überhaupt etwas über den zu entscheidenden Fall weiß. Gebietet die Einzelfallgerechtigkeit also, gar nicht zu wissen, wenn man nicht genau weiß? Meines Erachtens ist hier eine Kosten-Nutzen-Abwägung ausschlaggebend.

Die Einzelfallgerechtigkeit verbietet es jedenfalls, sich auf vermeidbarer Unwissenheit auszuruhen. Wenn man gerade nicht weiß, aber ohne Umstände wissen könnte, muss man forschen. Andererseits sollte die Einzelfallgerechtigkeit den Zufall, dass über bestimmte Rechtsgenossen gerade nicht mehr in Erfahrung zu bringen ist, nicht in das Privileg ummünzen, dass diese Rechtsgenossen vom Recht nicht belangt werden können, obwohl man etwas (Ungenaues) über sie weiß. Wenn also Wissen mit vertretbarem Aufwand zu erwerben ist, gebietet es die Einzelfallgerechtigkeit, die Referenzklassen weiter zu verengen. Wenn aber mit vertretbarem Aufwand kein weiteres Wissen zu erlangen ist, muss die kleinste bekanntermaßen heterogene Referenzklasse als „praktisch homogen“ gelten – man hat nichts Besseres.[32] Die Häufigkeit in der heterogenen Referenzklasse repräsentiert alles, was man weiß und gerade wissen kann. Dieses Wissen darf man nutzen. Die Entscheidung wird durch freiwillige Ignoranz ja nicht besser.

Bei der Frage, ob nun zu forschen ist, oder die kleinste heterogene Referenzklasse, für die man noch Häufigkeiten kennt, als „praktisch homogen“ zu behandeln ist, sind die Kosten der Wissensbeschaffung (der Referenzklassenverengung) mit den erwarteten Kosten der fälschlichen Belastung des Normadressaten abzuwägen.[33]

F. Einzelfallgerechtigkeit und Beibringungsgrundsatz

Im Falle des Kartellschadens wird anhand der Kosten-Nutzen-Betrachtung erkennbar, dass der Beibringungsgrundsatz das Problem in Wohlgefallen auflöst. Der Beibringungsgrundsatz ist „gebaut“ für die sukzessive Erschließung von Wissen. Seine Konstruktion erlaubt es, in jeder Situation das

32 Schweizer 2015, S. 363.

33 Für die Wahrscheinlichkeit der fälschlichen Belastung ist auf die Häufigkeit der heterogenen Referenzklasse abzustellen, die das Wissen repräsentiert, das vor der weiteren Wissenssuche besteht. Anderes Wissen steht nicht zur Verfügung.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

gerade verfügbare (vorläufige) Wissen zu unterstellen. Dann gewährleistet er „automatisch" eine zutreffende und „einzelfallgerechte" Entscheidung für oder gegen weitere Aufklärung von dieser Stufe an. Der Beibringungsgrundsatz lässt, wenn der Richter prima facie den Medianschaden der heterogenen Referenzklasse aller Kartelle unterstellt, die subjektive Beweislast für einen geringeren Schaden auf den Beklagten Kartellanten wechseln und bürdet damit die weiteren Aufklärungskosten demjenigen auf, der den Vorteil der weiteren Aufklärung hat. Er wird aufklären, wenn die Aufklärung weniger kostet als ihr Vorteil dem Belasteten wert ist. Aufklärung erfolgt also genau dann, wenn die Abwägung von Kosten der Aufklärung und der durch sie vermeidbaren (erwarteten) unberechtigten Belastung des Kartellenten eben diese Aufklärung gebietet. Wenn die Kosten der Aufklärung höher sind als die durch zusätzliche Aufklärung vermeidbare (erwartete) unberechtigte Belastung des Kartellanten, wird er nicht aufklären. In dieser Situation wäre die weitere Aufklärung auch nicht geboten und die Einzelfallgerechtigkeit lässt es zu, dass auf die heterogene Referenzklasse abgestellt wird, die alles verfügbare Wissen repräsentiert.

Etwas schwieriger liegen die Dinge, wenn der Kartellant höhere Kosten der Sachverhaltsaufklärung hat als andere am Verfahren beteiligte. Ein solcher Umstand implizierte, dass die Kosten der Sachverhaltsaufklärung günstiger sein könnten, so dass sie die Nachteile des Kartellanten nicht überwögen, müsste ein anderer Akteur die Sachlage aufklären. Hier können auf der Darlegungsebene die sekundäre Darlegungslast und auf der Beweisebene Vorlagepflichten (etwa die recht verstandene Anwendung von § 142 ZPO) oder richterliche Würdigung mit Hilfe des Grossman'schen Schließens[34] helfen, die Kostenvorteile des Prozessgegners zu realisieren.

Der Zivilprozess institutionalisiert mithin die erforderliche Kosten-Nutzen-Abwägung, welche die Einzelfallgerechtigkeit ausfüllt. Er tut dies mit Hilfe des Beibringungsgrundsatzes.

G. Ausblick

Ich glaube die hier angestellten Überlegungen sind über das Kartellschadensrecht hinaus relevant. Wie die Ergebnisse in weiteren Rechtsfragen nach dem hier vorgeschlagenen Ansatz lauten, wäre eigens auszuarbeiten.

34 Morell 2022, 189–279.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Daher seien hier nur Denkanstöße zur Illustration vorgetragen. Credit Scoring ist in Frankfurt jüngst ausführlich diskutiert worden. Gebietet die Einzelfallgerechtigkeit, ggf. ausgefüllt durch Grundrechte, dass die Entscheidung über die Kreditvergabe letztlich von Menschen getroffen wird, weil der „Einzelfall“ ja möglicherweise nicht zur „Statistik“ passt? Dem menschlichen Kreditvergeber werden apriori auch keine homogeneren Referenzklassen zur Verfügung stehen als dem Vergabealgorithmus. Aber ein menschlicher Kreditvergeber kann vielleicht eher zusätzliche Nachforschungen anstellen, als ein vorgegebener Algorithmus, um seine Referenzklasse aktiv zu verkleinern. Ob die Einzelfallgerechtigkeit die Bank dazu verpflichtet, diese menschliche Nachforschung zu ermöglichen, sollte davon abhängen, ob die Kosten dieser Aufklärung für die Bank die Nachteile der Abstraktion für Kreditnehmer übersteigen. Auch hier steht ein Mechanismus zur Verfügung, der zumindest grundsätzlich davon dispensieren kann, dass ein Dritter diese Kosten-Nutzen-Abwägung explizit durchführen muss: Funktionierender Wettbewerb am Markt. Automatisiertes credit scoring löst die Befürchtung aus, dass Mitgliedern bestimmter Gruppen aufgrund der Generalisierung Kredite verwehrt werden, die sie nach ihrer tatsächlichen, aber nicht beobachteten Kreditwürdigkeit eigentlich (ggf. günstiger) erhalten sollten. Sind die Kosten der Verweigerung für die Betroffenen größer als die Vorteile der Generalisierung, sollte sich ein anderer Wettbewerber finden, der mit Hilfe von Menschen diese Differenzierung vornimmt und sich den Vorteil mit der betroffenen Gruppe teilt.

Vielleicht könnte dieses Vorgehen auch Diskussionen über Diskriminierung im öffentlichen Recht informieren: Der Staat hat durch Sachverhaltsaufklärung kleine, möglichst homogene Referenzklassen zu bilden, aber nur solange die Kosten den Nutzen nicht übersteigen. Hier verlasse ich als Zivilrechtler nun wirklich mein Gebiet und will daher die Ergebnisse zusammentragen.

H. Schluss

Das Problem der Heterogenität als Problem für Steuerung zeigt sich auch in der Bestimmung der Höhe des Kartellschadens. Hier ist der Medianschaden aller Kartelle bekannt, aber unklar, wie diese Information für den „Einzelfall“ zu nutzen ist. Das Problem der Heterogenität als Problem für Steuerung ist eng verwandt mit dem Referenzklassenproblem.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Letztlich „löst" man das Problem durch Abstraktion. Der zulässigen Abstraktion sind allerdings Grenzen durch das Ziel der „Einzelfallgerechtigkeit" gesetzt. Sie verbietet nicht Abstraktion, sondern nur deren Übermaß. Grundsätzlich liegt die Grenze zwischen zulässiger und übermäßiger Abstraktion an der größten (noch) homogenen Referenzklasse. Anfangswahrscheinlichkeiten aus ihr zu bestimmen, entspricht immer dem Gebot der Einzelfallgerechtigkeit, aber weiter als bis zu ihr sollte grundsätzlich nicht abstrahiert werden. In den schwierigen Fällen sind die Häufigkeiten des gesuchten Merkmals in der homogenen Referenzklasse aber nicht bekannt, sondern es stehen nur bekanntermaßen heterogene Referenzklassen zur Information der Anfangswahrscheinlichkeit zur Verfügung. In diesen Fällen muss das Gebot des Strebens nach Einzelfallgerechtigkeit durch eine Kosten-Nutzen-Analyse ausgefüllt werden. Liegen die Kosten der Informationssuche, welche die Referenzklasse verengen würde, unter der (erwarteten) unberechtigten Belastung des Adressaten der Maßnahme, ist die Referenzklasse durch Erforschung weiterer Zusammenhänge zu verengen. Liegen die Kosten weiterer Forschung über denen der (erwarteten) Belastung des Adressaten der Maßnahme, ist die verfügbare heterogene Referenzklasse als epistemisch homogen zu behandeln. Diese Kosten-Nutzen-Abwägung ist im Beibringungsgrundsatz institutionalisiert. Der Richter kann daher zu jeder Zeit das verfügbare Wissen zu Grunde legen, auch wenn er es aus einer heterogenen Referenzklasse schöpft.

Ich schließe mit der Hoffnung auf viele weitere anregende Diskussionen mit dem Jubilar über Heterogenität als Problem für Steuerung und die vielen anderen faszinierenden Probleme zwischen Recht, Wahrscheinlichkeit und sozialer Realität.

Literatur

Aristotles (1911): Nikomachische Ethik. Übersetzt und mit einer Einleitung und erklärenden Anmerkunden versehen von Eugen Rolfes. 2. Aufl. Leipzig: Verlag von Felix Meiner (Philosophische Bibliothek, 5).

BGH, Urteil vom 11.12.2018, Aktenzeichen KZR 26/17. WuW 2019, 91–97.

BGH, Urteil vom 28.01.2020, Aktenzeichen KZR 24/17. NZKart 2020, 136–141.

BGH, Urteil vom 09.07.2024, Aktenzeichen KZR 98/20.

Britz, Gabriele (2008): Einzelfallgerechtigkeit versus Generalisierung. Verfassungsrechtliche Grenzen statistischer Diskriminierung. Tübingen: Mohr Siebeck.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Coppik, Jürgen; Heimeshoff, Ulrich (2021): Praxis der Kartellschadensermittlung. Ökonomische Evidenz zur Effektivität von Kartellen. Düsseldorf: Handelsblatt Fachmedien.

Franck, Jens-Uwe (2024): § 33a. In: Torsten Körber, Heike Schweitzer und Daniel Zimmer: GWB. Kommentar zum Deutschen Kartellrecht. 7. Aufl. München: C.H. Beck (Wettbewerbsrecht, Band 2).

Galle, René (2016): Der Anscheinsbeweis in Schadensersatzfolgeklagen – Stand und Perspektiven. In: NZKart, S. 214–220.

Heusel, Nicola; Hildebrand, Thomas; Mattes, Anselm (2024): Empirie schlägt Vermutung. Vorschläge für eine bessere Nutzung ökonomischer Gutachten in Kartellschadensersatzverfahren. In: WuW, S. 379–387.

Hornkohl, Lena (2020): Freie Schätzung der Kartellschadenshöhe nach § 287 ZPO. Eine Reaktion auf jüngste Entwicklungen. In: NZKart, 661.

Inderst, Roman; Thomas, Stefan (2021): Zum Umgang mit Regressionsanalysen in Kartellschadensersatzfällen. In: ZWeR, S. 432–459.

Mallmann, Roman A.; Leinemann, Franziska (2023): Sonderzivil(prozess)recht für kartellrechtliche Schadensersatzansprüche? In: ZWeR, S. 44–56.

Martinek, Michael (2023): Der „kleine Anscheinsbeweis“ im Kartellschadensersatzrecht. Anm. zu BGH, v. 29.11.2022, KZR, 42/20 - Schlecker. In: jM, S. 144–148.

Morell, Alexander (2022): Der Beibringungsgrundsatz. Tübingen: Mohr Siebeck.

Morell, Alexander (2025a): Schätzung von Kartellschäden: Beweismaß, inkrementelles Vorgehen und die Rolle von Erfahrungssätzen. Teil 1. In: Wirtschaft und Wettbewerb, S. 129–137.

Morell, Alexander (2025b): Schätzung von Kartellschäden: Beweismaß, inkrementelles Vorgehen und die Rolle von Erfahrungssätzen. Teil 2. In: Wirtschaft und Wettbewerb, S. 194–200.

Oxera; Komninos, Assimakis P. (2009): Quantifying antitrust damages. Towards non-binding guidance for courts : study prepared for the European Commission. Luxembourg: EUR-OP.

Roth, Herbert (2016): Private Rechtsdurchsetzung im Zivilprozess. In: JuristenZeitung 71 (23), S. 1134–1140. DOI: https://doi.org/10.1628/002268816X14712722914313.

Roth, Wulf-Henning (2018): Vorb. §§ 33-33h GWB. In: Wolfgang Jaeger, Juliane Kokott, Petra Pohlmann und Dirk Schroeder: Frankfurter Kommentar zum Kartellrecht. Mit Kommentierung des GWB, des EG-Kartellrechts und einer Darstellung ausländischer Kartellrechtsordnungen, Stand: 9/2024, 109. Köln: Verlag Dr. Otto Schmidt.

Rüßmann, Helmut (1987): § 286 ZPO. In: Rudolf Wassermann: Alternativkommentar zur Zivilprozessordnung. Neuwied.

Schweitzer, Heike; Woeste, Kai (2022): Zum Umgang mit ökonomischer Unsicherheit bei der Schätzung von Kartellschäden. Eckpfeiler eines kartellschadensersatzspezifischen Beweisrechts. In: ZWeR, S. 46–79.

Schweizer, Mark (2015): Beweiswürdigung und Beweismaß. Rationalität und Intuition. Tübingen: Mohr Siebeck (Jus privatum, v.189).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Stancke, Fabian (2018): Rechtliche Rahmenbedingungen kartellrechtlicher Massenklagen. In: WuW (2), S. 59–66.
Weitbrecht, Andreas (2017): Eine neue Ära im Kartellschadensersatzrecht – Die 9. GWB-Novelle. In: NJW, S. 1574–1578.
Weitbrecht, Andreas (2024): Kartellschadensersatz 2023. In: NZKart, S. 170–177.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Rechtsinnovationen über Grenzen
Der lange Weg zum Europäischen Verbund für Territoriale Zusammenarbeit

Stefan Okruch

A. Einleitung

Innovationsfreude ist ein Phänomen im Wandel. Während über weite Teile der Menschheitsgeschichte die Bewahrung eines stabilen Kanons von Gütern, Regeln, Riten, Gesellschafts- und Herrschaftsformen als anzustrebendes Ziel galt, scheint es inzwischen als ganz selbstverständlich, für Kreativität, Agilität oder gar Disruption, kurzum: für Innovation zu sein. Auch die Wirtschaftspolitik gibt sich innovationsfreudig, auch wenn die normative Grundlage durchaus voraussetzungsvoll ist, denn fundamentale Neuigkeit versperrt sich ebenso einer wohlfahrtsökonomischen Beurteilung wie einem gesamtgesellschaftlichen Konsens über die positiven Wirkungen von Innovationen (Witt/Schubert 2010).

Neben der gesteigerten gesellschaftlichen und wirtschaftspolitischen Innovationsfreude hat sich auch die ökonomische Sicht auf deren Träger (und Trägerinnen) verändert. Während bei Schumpeter (1911) der „Unternehmer" als Innovator weniger beschrieben als heroisch überhöht wurde, agieren in der modernen Ökonomik Innovatoren mit verschiedenen Rollen in Nationalen Innovationssystemen (NIS; OECD 1997). Prototyp ist somit nicht mehr der einsam vor sich hinbrütende Tüftler, vielmehr wird Innovationsfreude in einer komplexen Interaktion unterschiedlicher Akteure („Helix") gefördert, werden Innovationen in smart eingerichteten Akzeleratoren beschleunigt oder in Inkubatoren bis zur Marktreife gepflegt.

Mit dem Hinweis auf solche Einrichtungen wird auch klar, dass die Förderung von Produkt- und Prozessinnovation vorgängige organisationale und institutionelle Innovationen erfordern kann. Womit man zur Innovationsfreude im Recht (und auch in der Rechtswissenschaft) gelangt, die aus einer interdisziplinären Perspektive ebenfalls erstaunlich ausgeprägt ist und die im Folgenden, wenn auch nur schlaglichtartig und am

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Beispiel einer supranationalen, für interregionale und transnationale Zusammenarbeit gedachten – und (auch) in diesem Sinne grenzwertigen – Rechtsinnovation beleuchtet werden soll.

B. Rechtsinnovationen: Phänomen und Bewertung

Engel und Weinshall (2022, 141) bemerken zur Ubiquität von Rechtsinnovationen lakonisch: „The law is permanently under construction. Hence, innovation is not a rare event. It is the standard modus operandi of the law". Auch wenn sich diese Aussage mit Bedacht auf gesetzgeberische Innovationen bezieht, wirft der empirische Befund Fragen aus der Perspektive einer normativen Ökonomik auf: Bedeutet eine allzu hektische Gesetzgebung (auch und gerade in wirtschaftlich relevanten Bereichen) unheilvollen wirtschaftspolitischen Punktualismus? Oder ist gar Ausdruck einer sich immer schneller drehenden Interventionsspirale? Auch abseits solcher inzwischen vielleicht etwas altbacken wirkenden ordnungspolitischen Bedenken kann befürchtet werden, dass zu rasche Gesetzesänderungen die Funktion des Rechts als (relativ) stabile Umwelt für die wirtschaftliche Dynamik gefährden. Denn nicht nur für legislativen Neuerungen wird typischerweise angenommen, dass die Innovationsrate und Evolutionsgeschwindigkeit des Rechts nicht höher sein darf als die Innovations- oder Anpassungsgeschwindigkeit von denjenigen Systemen, die das Recht regeln soll. Dies gilt demnach auch für Innovationen in der (professionellen) Rechtsanwendung. Und noch allgemeiner: auch die Innovationfreude jedweder Rechtsunterworfenen mag eine empirische Tatsache sein, freilich bedeuten solche innovativen Erweiterungen von Devianz ganz offensichtlich ein normatives Problem.

Aber vielleicht ist es auch an der Zeit, den normativen Standard relativ stabilen Rechts zu hinterfragen oder nachgerade gegen den Strich zu bürsten. Wenn Recht als ein Bündel von Standards und als Netzwerkgut aus einer ökonomischen Sicht von einem Zuviel an Wandel bedroht sein kann (*excess momentum*), so ist *a priori* auch die Gefahr übermäßigen Beharrens (*excess inertia*) nicht ausgeschlossen. Und auch juristisch wurde diese Gefahr in jüngerer Zeit gesehen und gefordert, den rechtlichen Innovationsprozess grundlegend zu verändern und zu dynamisieren (Hadfield 2017). Dieser aktivistische Ansatz begegnet freilich sogleich der schon beschriebenen Komplexität der Innovationsförderung. Rechtinnovationen können, anders gewendet, immer nur ein Teil des NIS sein,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

setzen jedoch auch ein eigenes Innovationssystem voraus, in dem das komplexe Geschehen rund um juristische Invention und rechtliche Innovation stattfinden und beschrieben werden kann.

Im Folgenden wird die EU-Rechtsinnovation des Europäischen Verbundes für Territoriale Zusammenarbeit (EVTZ) näher beleuchtet, die organisationale Neuerungen ermöglichen und anregen soll, welche schließlich im grenzüberschreitenden europäischen Verbund von Akteuren innovative Angebote von Gemeinschaftsgütern ermöglichen. Dabei geht es weder um eine detaillierte Wirkungsanalyse noch um die Untersuchung der Diffusion (Eckardt/Okruch 2021). Auch die Regelungsdetails können außen vor bleiben[1]. Vielmehr soll das spezifische Innovationssystem beschreiben werden, also die Akteure und ihre Anreize, vor allem aber die Komplexität der Produktion von rechtlichen Neuerungen – und ihre Langwierigkeit.

C. Grenzen als rechtliche Innovationsquelle

I. Geographien der Grenze

Das grundlegende institutionelle Problem grenzüberschreitender Kooperationen besteht in der fehlenden Kongruenz zwischen politisch-administrativen Territorien – einschließlich der territorial gebundenen Rechtsordnungen – und der „Geographie der Probleme" (Frey 1997, 13). Diese strukturellen Herausforderungen von Grenzregionen verändern sich nicht allein dadurch, dass eine supranationale Ebene wie die EU etabliert wird, solange supranationale (Wirtschafts-)Politik primär über die Mitgliedsstaaten kanalisiert wird. Allerdings hat die supranationale Ebene, unabhängig von der Staatsorganisation der Mitgliedstaaten, Anreize, subnationale Akteure möglichst unmittelbar anzusprechen. Dadurch erweitert sich ihr Handlungsspielraum im Mehrebenensystem und sie gewinnt potenziell größere Unabhängigkeit gegenüber den Mitgliedstaaten. Grenzregionen stellen in diesem Kontext einen besonders geeigneten Ansatzpunkt dar, da sich hier sowohl die polit-ökonomische Interessenlage des supranationalen Zentrums als auch reale strukturelle Benachteiligungen der grenznahen Peripherie bündeln. Eine supranationale Förderung von grenzüberschreitender Zusammenarbeit kann für die

1 Vgl. dazu Eckardt/Okruch (2017), worauf auch der folgende Textteil basiert.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Internalisierung interregionaler Externalitäten sorgen und die lokale und grenzüberschreitende Bereitstellung öffentlicher Güter ermöglichen. Eine solche dezentrale Bereitstellung über Grenzen kann im Sinne der klassischen Theorie des Fiskalföderalismus angebots- und nachfrageseitig optimal sein – oder jedenfalls besser als die getrennte Produktion (oder Nicht-Produktion) durch die Nationalstaaten.

Nicht überraschend also fungieren Grenzregionen in der EU seit Langem als „Mikromodelle europäischer Integration" (Pasi 2007). In ihnen wurden früh Versuche unternommen, die Diskrepanz zwischen politisch-administrativer und funktionaler Raumgliederung zu reduzieren. Dies wirft jedoch zugleich die Frage nach geeigneten, innovativen Governance-Strukturen für solche Kooperationsformen auf:

> „European integration has had a dual impact on border regions. On the one hand, borders were physically dismantled across most of the EU's internal territory (…). On the other hand, border regions have become a fertile ground for territorial cooperation and institutional innovation" (de Sousa 2013, 669).

Vor diesem Hintergrund entstanden bereits früh vielfältige Initiativen zur institutionellen Bewältigung der Herausforderungen grenzüberschreitender Zusammenarbeit. Diese waren zunächst überwiegend bottom-up initiiert, d. h. sie erhielten weder institutionelle noch substanzielle finanzielle Unterstützung durch nationale oder supranationale Akteure.

Eine zentrale Rolle als Dachverband und „institutional entrepreneur" (Perkmann 2003, 168) nahm seit 1971 die Arbeitsgemeinschaft Europäischer Grenzregionen (AGEG) ein, mit der sich die verschiedenen Bottom-up-Initiativen eine gemeinsame Stimme gaben. Ab den 1980er Jahren intensivierten dann auch die zwei Top-down-Akteure europäischer Integration ihre Unterstützung: Der Europarat fokussierte auf die (völker-)rechtliche Fundierung grenzüberschreitender Kooperationen, während die Europäische Union ab 1990 zuerst auf eine massive finanzielle Förderung territorialer Zusammenarbeit setzte. Trotz der unterschiedlichen Ansätze richteten sich beide Strategien schlussendlich auf dasselbe institutionelle Kernproblem – die rechtliche Verfasstheit der entstehenden Kooperationsstrukturen – und führten letztlich zu ähnlichen Lösungsmodellen.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Formen und Verfassungen grenzüberschreitender Zusammenarbeit

Die deutsch-niederländische Euregio, gegründet 1958, gilt als Vorreiterin zahlreicher in der Folgezeit entstandener grenzüberschreitender Kooperationen. Deren institutionelles Grundmodell bestand häufig aus vertraglichen Kooperationen ohne eigene Rechtspersönlichkeit oder aus Vereinen nach nationalem Recht (Zapletal 2010, 18). Die Gründung und langfristige Tragfähigkeit solcher Kooperationen waren somit abhängig von den rechtlichen Rahmenbedingungen der beteiligten Jurisdiktionen. Fehlten geeignete Rechtsformen, hing ihre Fortführung oftmals vom „good will" der beteiligten Akteure (Perkmann 2003, 155) oder von gemeinsamen normativen Orientierungen (Svensson 2014) ab.

Die Europäische Union griff diese institutionellen Hindernisse erstmals 1973 auf, doch der Vorschlag der Kommission zur „Europäischen Kooperationsvereinigung" (EKV) blieb erfolglos. Auch die spätere „Europäische wirtschaftliche Interessenvereinigung" (EWIV, Verordnung Nr. 2137/85) erwies sich für grenzüberschreitende Kooperationen als ungeeignet.

Ein wesentlicher Schritt zur rechtlichen Vereinheitlichung war das 1980 im Europarat verabschiedete „Europäische Rahmenübereinkommen über die grenzüberschreitende Zusammenarbeit zwischen Gebietskörperschaften" (Madrider Abkommen). Kooperationen, die darauf basierten, werden unter dem – bewusst offenen – Begriff „Eurodistrikt" zusammengefasst. Ihre Zahl wuchs jedoch nur langsam, da trotz Rahmenübereinkommen jeweils bi- oder multilaterale völkerrechtliche Verträge notwendig blieben. Obwohl dies eigentlich nur für begründete Einzelfälle vorgesehen war, wurde diese Voraussetzung häufig als Regelfall interpretiert. Die Zusatzprotokolle von 1995 und 1998 reduzierten diese Hürden nur bedingt, zumal die Zahl der Ratifizierungen gering blieb. Einzelne Abkommen ermöglichten die Gründung grenzüberschreitender Zweckverbände (Local Grouping of Cross-Border Cooperation, LGCC), jedoch ausschließlich durch Gebietskörperschaften und nur in den Grenzregionen der Vertragsstaaten. Damit war dieses Modell als generelle Rechtsform territorialer Kooperation wenig geeignet (Engl 2016, 146).

Seit 1990 veränderten sich die Anreize zur Kooperation durch die finanzielle Ausstattung im Rahmen der INTERREG-Programme der EU grundlegend. Die Zahl der Euregios stieg rasch auf rund 150 (Svensson 2014, 409); nahezu alle EU-Grenzregionen waren bald in mindestens einer Form grenzüberschreitender Zusammenarbeit organisiert (Perkmann

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2007, 861). Gleichzeitig blieb das institutionelle Grundproblem bestehen und trat im Umgang mit EU-Fördermitteln besonders deutlich zutage: Die gemeinsame Verwaltung ohne eine einheitliche Rechtsstruktur führte aufgrund divergierender nationaler Rechts- und Verwaltungsverfahren zu Ineffizienzen und unklaren Verantwortlichkeiten (de Sousa 2013, 670).

Die EU-Kommission legte schließlich 2004 einen Vorschlag für eine europäische Rechtsform grenzüberschreitender Zusammenarbeit vor. Dieser enthielt bereits die wesentlichen Strukturmerkmale des späteren „Europäischen Verbunds für territoriale Zusammenarbeit" (EVTZ). Die Verordnung (EG) Nr. 1082/2006 implementierte sodann den rechtlichen Kern des EVTZ, also eine neuartige europäische Rechtsform, die die grenzüberschreitende, transnationale und interregionale Zusammenarbeit erleichtern und fördern soll. Der EVTZ ist eine juristische Person, die grundsätzlich von Gebietskörperschaften aus mindestens zwei Mitgliedstaaten gegründet wird. Mit einem EVTZ ist es möglich, ohne weitere zwischenstaatliche Verträge gemeinsam öffentliche Dienstleistungen zu erbringen.

Mit der Änderungsverordnung (EU) Nr. 1302/2013 wurden u. a. die möglichen Aufgabenbereiche von EVTZ erweitert, die Mitgliedschaft privatrechtlicher, jedoch öffentlich kontrollierter Akteure ermöglicht, die Beteiligung von Partnern aus Drittstaaten präzisiert und die Verfahrensvorschriften für Gründung und Änderung eines EVTZ konkretisiert. Auch Regelungen zur arbeits – und sozialrechtlichen Zuordnung von Beschäftigten des Verbunds wurden vereinheitlicht.

Einen weiteren Mosaikstein zum Gesamtbild des Innovationsprozesses fügt 2009 auch der Europarat hinzu und schafft eine eigene Rechtsform: Das dritte Zusatzprotokoll zum Madrider Abkommen führte den „Verbund für euroregionale Zusammenarbeit" (VEZ) ein. Inhaltlich orientiert sich der VEZ eng am EVTZ und bietet demnach für Akteure aus EU-Mitgliedstaaten keinen Mehrwert. Seine praktische Bedeutung ist gering.

III. Der EVTZ als institutionelle Disruption?

Vor dem Hintergrund der traditionellen Kohäsionspolitik wurde die Einführung des EVTZ durchaus als disruptiv wahrgenommen:

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

> „The Regulation triggered a lively debate, since the EU was for the first time ‚legislating' on the governance and legal structures of regional policy, rather than on usual (and important) business such as the provision of a multi-annual plan and financial framework" (Spinaci/Vara-Arribas 2009, 6).

Die intensive Debatte resultierte auch daraus, dass mit der EVTZ-Verordnung Zuständigkeitsfragen neu verhandelt wurden und die faktische Regionalisierung – und damit eine potenzielle Umgehung der Nationalstaaten – erstmals eine normative Grundlage erhielt. Schließlich wurde mit dem Vertrag von Lissabon 2007 das Ziel des territorialen Zusammenhalts auch als neues Unionsziel verankert (Art. 3 III EUV, Art. 174 I AEUV).

Unabhängig von den polit-ökonomischen Aspekten stellt die Rechtsform des EVTZ ein zusätzliches Instrument im vertikalen Wettbewerb der Rechtssysteme dar. Dies ist insofern von erheblicher Bedeutung, als die (horizontale) Mobilität rechtlicher Organisationsformen für die grenzüberschreitende Zusammenarbeit öffentlicher Körperschaften – und ebenso öffentlicher Unternehmen – bislang nur eingeschränkt gegeben ist. Privatrechtliche Rechtsformen, bei denen die Mobilität des Gesellschaftsrechts grundsätzlich gewährleistet ist, erweisen sich für diesen Zweck als ungeeignet. Die bestehenden spezifischen Rechtsformen zur territorialen Zusammenarbeit sind demgegenüber nur begrenzt mobil, da sie ausschließlich einem engen Nutzerkreis offenstehen und in der Regel erst nach Erfüllung zusätzlicher rechtlicher Voraussetzungen in Anspruch genommen werden können.

Obwohl der EVTZ somit einen wesentlichen Beitrag zur Weiterentwicklung der rechtlichen Rahmenbedingungen territorialer Kooperation leistet, bleiben zahlreiche Fragestellungen im Zusammenhang mit seinem praktischen Betrieb ungeklärt. Dies betrifft insbesondere arbeits- und sozialrechtliche Aspekte, deren Lösung weder durch eine Harmonisierung der einschlägigen Normen noch über das Prinzip der gegenseitigen Anerkennung realistisch erscheint. Der Innovationsprozess wird also weitergehen.

D. Fazit: Kooperation im Systemwettbewerb

Die Entwicklung grenzüberschreitender Zusammenarbeit in Europa zeigt, dass das zentrale institutionelle Problem – die fehlende Kongruenz zwischen politisch-administrativen Grenzen und funktionalen Problem-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lagen – bis in die Gegenwart ein strukturprägendes Moment und einen Treiber des institutionellen Innovationsprozesses darstellt. Die Vielzahl an Kooperationsformen, von frühen Euregios bis hin zu formalisierten Verbünden wie dem EVTZ, lässt sich vor diesem Hintergrund als sukzessive Antwort auf ein dauerhaft ungelöstes Koordinationsproblem interpretieren.

Die Analyse der institutionellen Entwicklung verdeutlicht, dass zunächst vor allem bottom-up getragene Initiativen und flexible, häufig rechtlich prekäre Arrangements im Vordergrund standen. Diese frühen Formen grenzüberschreitender Kooperation waren stark von politischem Wohlwollen, persönlichen Netzwerken und gemeinsamen normativen Orientierungen abhängig und blieben damit institutionell fragil. Erst mit der schrittweisen Einbindung supranationaler Organisationen – der Europäischen Union und des Europarats – wurde versucht, diese Kooperationen durch rechtliche und finanzielle Instrumente zu stabilisieren und zu verstetigen. Gleichwohl zeigte sich, dass finanzielle Förderung (z. B. durch INTERREG) institutionelle Defizite nicht automatisch behebt, sondern diese im Zuge der komplexen Mittelverwaltung oftmals noch sichtbarer macht.

Mit der Einführung des EVTZ ist eine qualitativ neue Stufe der Institutionalisierung erreicht worden. Erstmals steht eine spezifische europäische Rechtsform zur Verfügung, die auf die besonderen Anforderungen territorialer Kooperation und grenzüberschreitender Governance zugeschnitten ist. Die Verordnung über den EVTZ sowie ihre spätere Reform adressieren zentrale Fragen der Rechtspersönlichkeit, der Verantwortlichkeit, der Mitgliedschaft und der arbeits- und sozialrechtlichen Zuordnung. Zugleich markiert der EVTZ eine politisch bedeutsame Aufwertung territorialer Zusammenarbeit: Mit der Aufnahme des territorialen Zusammenhalts als eigenständiges Ziel der EU wird die bislang häufig funktional begründete Regionalisierung auch normativ verankert.

Die parallele Entwicklung des VEZ durch den Europarat unterstreicht, dass das Bedürfnis nach geeigneten Rechtsformen für grenzüberschreitende Kooperation nicht auf den EU-Kontext beschränkt ist. Allerdings zeigt die geringe praktische Bedeutung des VEZ, dass sich im Binnenraum der EU der EVTZ als de facto-Standard etabliert hat. Für EU-Mitgliedstaaten bietet der VEZ keinen erkennbaren zusätzlichen Mehrwert, sodass er überwiegend eine ergänzende Option an den Rändern des Unionsrechtsrahmens bleibt.

Insgesamt lässt sich festhalten, dass die institutionelle Innovation durch EVTZ und verwandte Instrumente zwar wesentliche Hemmnisse

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

grenzüberschreitender Zusammenarbeit reduziert, das grundlegende Spannungsverhältnis zwischen nationalstaatlicher Souveränität, supranationaler Steuerung und regionaler Handlungskapazität jedoch nicht auflöst.

Literatur

De Sousa, Luis (2013), Understanding European Cross-border Cooperation: A Framework for Analysis, in: Journal of European Integration 35, S. 669–687.

Eckardt, Martina /Okruch, Stefan (2021), The Legal Innovation of the European Grouping of Territorial Cooperation and its Impact on Regulatory Competition, in: Masson, Antoine/Robinson, Gavin (Hrsg.), Mapping Legal Innovation, Cham, S. 331–355.

Eckardt, Martina /Okruch, Stefan (2017), Kooperation im Systemwettbewerb: der Europäische Verbund für Territoriale Zusammenarbeit, in: List Forum für Wirtschafts- und Finanzpolitik, 43, S. 251–275.

Engel, Christoph /Weinshall, Keren. (2022), Diffusion of Legal Innovations, in: Annual Review of Law and Social Science.

Engl, Alice (2016), Bridging Borders through Institution Building: the EGTC as a Facilitator of Institutional Integration in Cross Border Regions, in: Regional & Federal Studies 26, S. 143–169.

Frey, Bruno S. (1997), Ein neuer Föderalismus für Europa: Die Idee der FOCJ, Tübingen.

Hadfield, G. K. (2017). Rules for a flat world. Why humans invented law and how to reinvent if for a complex global economy, New York.

Moss, Timothy (2012), Geographien von Gemeinschaftsgütern: Erklärungsansätze aus der neueren institutionen-,handlungs- und raumtheoretischen Forschung, in: Geographische Zeitschrift 100, S. 208–227.

OECD (1997), National Innovation Systems. http://www.oecd.org/science/inno/2101733.pdf.

Pasi, Paolo (2007), Euroregions as Micro-models of European Integration, in: Langer, Josef (Hrsg.), Euroregions – the Alps–Adriatic Context, Frankfurt am Main, S. 73–79.

Perkmann, Markus (2003), Cross-border Regions in Europe: Significance and Drivers of Regional Cross-border Co-operation, in: European Urban and Regional Studies 10, S. 153–171.

Perkmann, Markus (2007), Policy Entrepreneurship and Multilevel Governance: a Comparative Study of European Cross-border Regions, in: Environment and Planning C: Government and Policy 25, S. 861–879.

Schumpeter, Joseph A. (1911), Theorie der wirtschaftlichen Entwicklung, Berlin.

Spinaci, Gianluca /Vara-Arribas, Gracia (2009), The European Grouping of Territorial Cooperation (EGTC): New Spaces and Contracts for European Integration?, in: EIPASCOPE 2009/2, S. 5–13.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Svensson, Sara (2014), Forget the Policy Gap: why Local Governments Really Decide to Take Part in Cross-border Cooperation Initiatives in Europe, in: Eurasian Geography and Economics 54, S. 409–422.

Witt, Ulrich /Schubert, Christian (2010), *Constitutional interests in the face of innovations: how much do we need to know about risk preferences?* ,in: Papers on economics and evolution # 803. Verfügbar unter: https://nbn-resolving.org/urn:nbn:de:gbv:27-20100603-135030-4.

Zapletal, Jirka (2010), The European Grouping of Territorial Cooperation (EGTC): A New Tool Facilitating Cross-Border Cooperation and Governance, in: Quaestiones Geographicae 29, S. 16–26.

Primärquellen

Europäische Kommission (1973), Vorschlag einer Verordnung (EWG) des Rates über die Europäische Kooperationsvereinigung (EKV) vom 21. Dezember 1973; verfügbar unter http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1478961876064&uri=CELEX:51973PC2046 (letzter Zugriff: 31.10.2025).

Europäische Kommission (2004), Vorschlag für eine Verordnung des Europäischen Parlament und des Rates bezüglich der Schaffung eines Europäischen Verbunds für grenzüberschreitende Zusammenarbeit vom 17. April 2004; https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52004PC0496 (letzter Zugriff: 31.10.2025).

Europäisches Rahmenübereinkommen über die grenzüberschreitende Zusammenarbeit zwischen Gebietskörperschaften [Europarat] vom 21. Mai 1980 [„Madrider Abkommen"]; verfügbar unter https://www.coe.int/de/web/conventions/full-list/-/conventions/treaty/106 (letzter Zugriff: 31.10.2025).

Protokoll Nr. 2 zum Europäischen Rahmenübereinkommen über die grenzüberschreitende Zusammenarbeit zwischen Gebietskörperschaften in Bezug auf die interterritoriale Zusammenarbeit vom 5. Mai 1998; verfügbar unter https://www.coe.int/de/web/conventions/full-list/-/conventions/treaty/169 (letzter Zugriff: 31.10.2025).

Protokoll Nr. 3 zum Europäischen Rahmenübereinkommen über die grenzüberschreitende Zusammenarbeit zwischen Gebietskörperschaften betreffend Verbünde für euroregionale Zusammenarbeit (VEZ) vom 16. November 2009; verfügbar unter https://www.coe.int/de/web/conventions/full-list/-/conventions/treaty/206 (letzter Zugriff: 31.10.2025).

Verordnung (EG) Nr. 1082/2006 des Europäischen Parlaments und des Rates vom 5. Juli 2006 über den Europäischen Verbund für Territoriale Zusammenarbeit (EVTZ), Amtsblatt der Europäischen Union L 210, S. 19-24.

Verordnung (EU) Nr. 1302/2013 des Europäischen Parlaments und des Rates vom 17. Dezember 2013 zur Änderung der Verordnung (EG) Nr. 1082/2006 über den Europäischen Verbund für Territoriale Zusammenarbeit (EVTZ) im Hinblick auf Präzisierungen, Vereinfachungen und Verbesserungen im Zusammenhang mit

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

der Gründung und Arbeitsweise solcher Verbünde, Amtsblatt der Europäischen Union L 347, S. 303-319; verfügbar unter http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri= CELEX:32013R1302&from=DE (letzter Zugriff: 31. 10. 2025).

Verordnung (EWG) Nr. 2137/85 des Rates vom 25. Juli 1985 über die Schaffung einer Europäischen wirtschaftlichen Interessenvereinigung (EWIV); verfügbar unter http://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A31985R2137 (letzter Zugriff: 31. 10. 2025).

Vorschlag für eine Verordnung des Europäischen Parlament und des Rates bezüglich der Schaffung eines Europäischen Verbunds für grenzüberschreitende Zusammenarbeit vom 17. April 2004; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52004PC0496 (letzter Zugriff: 31. 10. 2025).

Zusatzprotokoll zum Europäischen Rahmenübereinkommen über die grenzüberschreitende Zusammenarbeit zwischen Gebietskörperschaften vom 9. November 1995; verfügbar unter https://www.coe.int/de/web/conventions/full-list/-/conventions/treaty/159 (letzter Zugriff: 31. 10. 2025).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Der Rundfunkbeitrag
Rechts- und verhaltenswissenschaftliche Perspektiven auf ein umstrittenes Finanzierungsmodell

Jörn Lüdemann

I. Inspirationen

Die Pflege eines beschaulichen wissenschaftlichen Gärtchens ist *Christoph Engels* Sache nicht. Hat er neue Forschungsfelder angelegt und intellektuell zur Blüte gebracht, gibt er sich nicht damit zufrieden, die gewachsenen Strukturen anschließend nur noch mit der Nagelschere in Form zu halten. Lieber macht er sich an die kreative Konzeption und Ausgestaltung neuer und größerer Gartenanlagen. Zwei bedeutende Areale seiner über die Jahrzehnte gewachsenen wissenschaftlichen Parklandschaft bilden das Medienrecht, das einen thematischen Schwerpunkt seines ersten Lehrstuhls bildete, sowie die verhaltenswissenschaftliche Analyse des Rechts, die er am Bonner Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern national wie international vorangetrieben hat. Die nachfolgenden Überlegungen zum Rundfunkbeitrag sind von beiden Forschungsbereichen maßgeblich inspiriert. Sie ergründen, warum der schon rechtssystematisch unpassende Rundfunkbeitrag (II.) insbesondere auch aus einer verhaltenswissenschaftlichen Perspektive nicht als Finanzierungsmodell für den öffentlich-rechtlichen Rundfunk überzeugen kann (III.). Weil sich die strukturellen Probleme nicht durch Detailkorrekturen beheben lassen, spricht rechtspolitisch viel für eine grundlegende Reform der öffentlichen Medienfinanzierung (IV.).

II. Öffentliche Medienfinanzierung im unpassenden Rechtskleid

1. Finanzierung durch Vorzugslasten

Das deutsche Rundfunkrecht ist in besonderem Maße pfadabhängig. Das gilt nicht nur für die Grundkonzeption des öffentlich-rechtlichen Rundfunks, dessen Auftrag vom Bundesverfassungsgericht in zahlreichen Ur-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

teilen[1] mit immer neuen Begründungen unterlegt und verfassungsrechtlich abgesichert worden ist. Auch die Finanzierung der öffentlichen Medienangebote bewegt sich auf gut gespurten Pfaden.

Die Länder als Mediengesetzgeber haben zum 1. 1. 2013 zwar die frühere gerätebezogene Rundfunkgebühr[2] durch den raumbezogenen Rundfunkbeitrag ersetzt, der seither von den Inhabern von Wohnungen, Betriebsstätten und gewerblich genutzten Kraftfahrzeugen zu entrichten ist;[3] zentrale Beweggründe für die Abkehr von der Rundfunkgebühr waren neben der schwierigen Kontrolle der Abgabenehrlichkeit insbesondere die technische Konvergenz, die den Konsum von Medieninhalten auch auf neueren, multifunktionalen Geräten wie Laptops oder Smartphones ermöglichte. Abgabentypologisch ist die Medienpolitik mit dem Rundfunkbeitrag aber dem Finanzierungspfad durch eine Vorzugslast treu geblieben. Alternative Wege einer Finanzierung aus Haushaltsmitteln, die bei der Reform ebenfalls diskutiert wurden, sind hingegen bewusst nicht eingeschlagen worden.[4] Es waren nicht zuletzt die Rundfunkanstalten selbst, die sich während des Reformprozesses für die Beitragsfinanzierung stark gemacht und sie mit verfassungsrechtlichen Gutachten[5] unterstützt hatten.

2. Probleme der Passform

Auch wenn das Bundesverfassungsgericht ungeachtet vielfach geäußerter Kritik[6] dem Rundfunkbeitrag seine grundsätzliche Vereinbarkeit mit dem

1 Grundsätzliche Kritik an dieser Judikatur früh bei *C. Engel*, Rundfunk in Freiheit, AfP 1994, 185.

2 Die streng genommen abgabensystematisch auch ein Beitrag war, vgl. *C. Engel*, Medienordnungsrecht, 1996, 122.

3 Eingehend dazu *E. Wagner*, Abkehr von der geräteabhängigen Rundfunkgebühr, 2011; *S. Röß*, Die Neuordnung der Finanzierung des öffentlich-rechtlichen Rundfunks durch den Rundfunkbeitragsstaatsvertrag, 2015.

4 Zu verworfenen Alternativmodellen *S. Eggerath*, Der Rundfunkbeitragsstaatsvertrag der Länder, 2016, 145 ff.

5 *P. Kirchhof*, Die Finanzierung des öffentlich-rechtlichen Rundfunks, 2010; nach der Reform noch einmal bekräftigend *H. Kube*, Der Rundfunkbeitrag, 2014.

6 Z. B. *C. Degenhart*, Verfassungsrechtliche Zweifelsfragen des Rundfunkbeitragsstaatsvertrags, ZUM 2011, 193; *S. Korioth/M. Koemm*, Gut gemeint, doch schlecht gemacht, DStR 2013, 833.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Grundgesetz bescheinigt hat,[7] zwängt dieses Abgabenmodell die Medienfinanzierung in ein ausgesprochen schlecht sitzendes Rechtskleid. Am Beispiel der Beitragspflicht im privaten Bereich, auf die sich auch die nachfolgende verhaltenswissenschaftliche Analyse konzentrieren wird, lässt sich illustrieren, warum der Rundfunkbeitrag die juristische Passform des Beitrags in jeder Hinsicht überdehnt.

a) Sondervorteil

Bereits der gruppenbezogene Sondervorteil, der den Beitrag als Vorzugslast gegenüber den Steuern als Gemeinlasten prägt, gerät beim Rundfunkbeitrag aus der Form. Gemeinhin wird die Erhebung von Beiträgen verfassungsrechtlich damit legitimiert, dass eine Personengruppe durch die Inanspruchnahme einer öffentlichen Einrichtung oder Leistung einen Sondervorteil erlangt. Deshalb erscheint es gerechtfertigt, die begünstigte Gruppe über die steuerlichen Allgemeinlasten hinaus mit einer Sonderlast an den Kosten der Einrichtung zu beteiligen.[8] Anders als bei der Gebühr kommt es beim Beitrag zwar nicht darauf an, ob dieser Sondervorteil tatsächlich in Anspruch genommen wird; es reicht die bloße Möglichkeit einer Nutzung.[9] Aber gerade wegen dieser gelockerten Vorteilsbeziehung (Nutzungsmöglichkeit statt Nutzung) ist es für die Erhebung einer Sonderlast erforderlich, den Vorteil zumindest einer besonderen Gruppe zuordnen zu können, die der öffentlichen Einrichtung oder Leistung näher steht als die Allgemeinheit. An einer abgrenzbaren Gruppe fehlt es beim Rundfunkbeitrag allerdings, weil die Gesamtheit der beitragspflichtigen Wohnungsinhaber nahezu deckungsgleich mit der Allgemeinheit ist; Ausnahmen bestehen aus sozialen Gründen lediglich an den Rändern des Beitragssystems.[10] Wo aber nahezu alle Bürgerinnen und Bürger Vorteilsempfänger sind, ist ein Sondervorteil für eine bestimmte Gruppe nicht mehr erkennbar. Der Rundfunkbeitrag überdehnt diese Vorzugslast vielmehr in Richtung einer steuerähnlichen Allgemeinlast.

7 BVerfGE 149, 222.
8 BVerfGE 149, 222 (250).
9 BVerfGE 38, 281 (311); 137, 1 (18).
10 § 4 Abs. 1 und 6 RBStV.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

b) Individualvorteil

Ein zweites Passformproblem ist damit eng verbunden. Selbst wenn man es mit dem Bundesverfassungsgericht für möglich hält, dass ein Beitrag von (fast) allen Bürgerinnen und Bürgern erhoben werden kann,[11] setzt die Abgrenzung zur Steuer zumindest einen individualisierbaren Vorteil für den einzelnen Beitragszahler voraus.[12] Doch auch dieser ist beim Rundfunkbeitrag ausgesprochen konturenschwach.

Die Blässe des Individualvorteils, den man durch die Nutzung der öffentlichen Medienangebote ziehen kann, beruht nicht allein darauf, dass das Angebot auch dem Gemeinwohl dienen soll, indem es zu einer breiten und vielfältigen öffentlichen Meinungsbildung beiträgt. Dass sich Individual- und Gemeinschaftsnutzen ergänzen, ist angesichts des staatlichen Gemeinwohlauftrags kein Spezifikum des Rundfunkbeitrags. Gleichwohl liegen die Dinge hier anders als bei klassischen Beiträgen, weil der kollektive Nutzen den Individualvorteil beim öffentlich-rechtlichen Rundfunk deutlich überstrahlt. Wie sehr sich die Situation von anderen Beiträgen unterscheidet, lässt sich bereits daran erkennen, dass sich Individual- und Kollektivvorteil kaum sauber voneinander trennen lassen. Denn die Nutzung der Angebote durch den Einzelnen ist zwingende Voraussetzung für das Gemeinwohl, weil der öffentlich-rechtliche Rundfunk seine gesamtgesellschaftlichen Wirkungen nur entfalten kann, wenn er vom Einzelnen rezipiert wird.

Das leitet zum eigentlichen Grund über, warum die Vorstellung der Abschöpfung eines Individualvorteils nicht passt. Die Beitragspflicht lässt sich verfassungsrechtlich am ehesten mit der Bereitstellung meritorischer Güter rechtfertigen.[13] Darunter versteht man in der Volkswirtschaftslehre Güter, die aus normativen Gründen als wünschenswert angesehen werden, die bei ausschließlich marktwirtschaftlicher Koordination aber nicht im erwünschten Maß bereitgestellt würden, weil die durchschnittliche Zahlungsbereitschaft der Konsumenten nicht ausreicht.[14] Unabhängig davon, wie berechtigt die Marktskepsis im Rundfunkbereich tatsächlich ist und wie weit das Angebot nur aus meritorischen Gütern besteht, liegt

11 BVerfGE 149, 222 (252).

12 So auch BVerfGE 149, 222 (252).

13 Für anderes wird er richtigerweise nicht gebraucht, vgl. bereits *C. Engel*, Medienordnungsrecht, 1996, 63.

14 Zum Konzept *R. A. Musgrave*, The Theory of Public Finance, 1959, 13 ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

darin nach wie vor der Kerngedanke der rundfunkrechtlichen Sonderdogmatik des Bundesverfassungsgerichts: Weil die marktwirtschaftlich agierenden privaten Medien nach Ansicht des Gerichts die verfassungsrechtlich gebotene mediale Versorgung der Gesellschaft alleine nicht erbringen können, sollen die Rundfunkanstalten dies sicherstellen.[15] Wenn *Public Value*-Inhalte aber aus Gründen des gemeinen Wohls bereitgestellt werden und nicht, weil die Bürgerinnen und Bürger in ihnen durchschnittlich selbst einen entsprechenden Vorteil sehen, dann ist es wenig überzeugend, den Rundfunkbeitrag gerade mit der Abschöpfung eines Individualvorteils zu rechtfertigen.[16]

Bei anderen meritorischen Gütern macht es der Staat nicht ohne Grund ganz anders. Entweder er finanziert das Angebot vollständig aus Steuern. Oder er beteiligt im Sinne einer Mischfinanzierung auch diejenigen durch eine Gebühr an den Kosten, die etwa eine Theatervorstellung besuchen und damit einen über die Kulturförderung hinausgehenden individuellen Sondervorteil genießen. Dass beim öffentlich-rechtlichen Rundfunk aus normativen Gründen keine von der tatsächlichen Nutzung abhängigen Gebühr im Sinne eines Pay-TV erhoben werden soll,[17] macht den Beitrag nicht plausibler. Dass möglichst viele die Angebote nutzen sollen, zeigt im Gegenteil noch deutlicher, dass der Gemeinschaftsnutzen ganz im Vordergrund steht.

c) Vorteilsanker

Vor diesem Hintergrund nimmt es denn auch nicht wunder, dass sich kein überzeugender Vorteilsanker finden lässt, an den sich die Beitragszahlung tatbestandlich knüpfen lässt. Während etwa beim Kurbeitrag der Aufenthalt im Kurort oder beim Erschließungsbeitrag das erschlossene Grundstück sachgerechte Vorteilsanker für eine Beitragspflicht sind, hat die Anknüpfung an die Wohnung beim Rundfunkbeitrag mit einer beitragsrechtlichen Zurechnung nichts gemein.[18] Weder gewinnt die Wohnung als solche durch die Existenz öffentlich-rechtlichen Rundfunks an Wert

15 BVerfGE 149, 222 (260 ff.).

16 Anders BVerfGE 149, 222 (259 ff.).

17 BVerwGE 108, 108 (113 f.).

18 Kritisch etwa auch *M. Cornils*, ZJS 2018, 627 (636); *J. A. Kämmerer*, Verfassungsmäßigkeit des Rundfunkbeitrags in Zeiten von Digitalisierung, Mobilität und „Fake News", NJW 2018, 3209 (3210 f.).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

(wie das bei einem Grundstück der Fall ist, für das ein Erschließungsbeitrag zu entrichten ist), noch ist sie als räumliche Einheit faktisch oder rechtlich Voraussetzung der Rundfunknutzung. Im Unterschied zu früheren Jahrzehnten sind die Medieninhalte angesichts der Digitalisierung längst auch außerhalb der Wohnung überall nutzbar. Und die Wohnung ist auch keine rechtliche Voraussetzung für die Nutzung, wie dies etwa bei einer Kammermitgliedschaft der Fall ist. Wenn eine Leistung aber an jedem beliebigen Ort von jedem voraussetzungslos genutzt werden kann, dann kann es keinen überzeugenden Zurechnungsanker geben außer den Menschen selbst.[19] Alle alternativen Anknüpfungspunkte sind im Grunde beliebig. Die Wohnung ist bestenfalls eine administrativ leicht handhabbare und stabile Erhebungsadresse, um Kontroll- und Vollzugsprobleme bei der Rundfunkfinanzierung zu vermeiden.

d) Lastenverteilung

Dass sich kein überzeugender Zurechnungsanker finden lässt, ist schließlich auch der Grund dafür, warum der Rundfunkbeitrag mit Blick auf die Lastenverteilung schlecht zum Beitragsmodell passt. Während die Vorzugslast die verfassungsrechtlich geforderte Lastengleichheit eigentlich dadurch befördern soll, dass sie diejenigen an der Finanzierung beteiligt, die von einem Gruppenvorteil besonders profitieren, konterkariert der Rundfunkbeitrag eine gleichheitsgemäße Lastenverteilung eher. Die Wohnungspauschale verteilt die Kosten des öffentlich-rechtlichen Rundfunks im Vergleich zur Steuer nämlich deutlich gröber. So werden etwa Menschen in Wohngemeinschaften finanziell weniger in Anspruch genommen, als wenn sie in Singlewohnungen lebten. Vor allem aber bemisst sich der Beitrag im Unterschied zur Erhebung progressiver Steuern auch unabhängig von der finanziellen Leistungsfähigkeit des Einzelnen, sodass etwa einkommensstarke Doppelverdiener denselben Beitrag zahlen wie eine Einzelperson mit geringem Einkommen.

19 Sieht man von den Endgeräten einmal ab, die nach der Abkehr von der geräteabhängigen Rundfunkgebühr gerade nicht mehr ausschlaggebend sein sollten.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. Programmierte Reaktanz: Die verhaltenswissenschaftliche Perspektive

1. Die Akzeptanzkrise des Rundfunkbeitrags

Das Beitragsmodell löst aber nicht nur aus einer rechtsdogmatischen Perspektive ein erhebliches Störgefühl aus. Es gibt wohl auch keine andere Abgabe, die in der Bevölkerung auf so viel Unverständnis und Ablehnung stößt wie der Rundfunkbeitrag, und zwar auch über den Kreis der eingefleischten Gegner des öffentlich-rechtlichen Rundfunks hinaus.[20] So hat der Beitragsservice allein im Jahr 2024 rund 1,59 Millionen Vollstreckungsersuchen stellen müssen.[21]

Das Ausmaß des Widerstands mag auf den ersten Blick vielleicht erstaunen, nicht zuletzt angesichts der zwar nicht geringen, im Vergleich zu anderen Abgaben aber auch nicht außergewöhnlichen Beitragshöhe.[22] Der beobachtbare Widerstand hat seine Wurzeln aber auch nicht ausschließlich in der finanziellen Belastung oder der Unzufriedenheit mit dem Programmangebot, sondern es ist – so die zentrale These dieser Studie – vor allem auch das Beitragsmodell selbst, das für die Akzeptanzkrise verantwortlich ist. Aus einer verhaltenswissenschaftlichen Perspektive, die nach den Wirkungen des Rechts auf die Adressaten fragt,[23] provoziert der Rundfunkbeitrag nämlich geradezu systematisch Widerstand.

2. Psychologische Reaktanz

Verhaltenswissenschaftlich lässt sich ein großer Teil der Gegenwehr, die etwa in Zahlungsverweigerungen, Klagen oder öffentlichen Protesten gegen den Rundfunkbeitrag zum Ausdruck kommt, als Ergebnis psychologischer Reaktanz deuten. Als psychologische Reaktanz bezeichnet man einen motivationalen Widerstand, der entsteht, wenn Menschen ihre Ent-

20 Zu deren Beweggründen *V. Stamer/R. Hohlfeld*, Öffentlich-rechtlicher Rundfunk in den Augen seiner Kritiker, Communicatio Socialis 2025, 17.

21 *ARD ZDF Deutschlandradio Beitragsservice* (Hrsg.), Jahresbericht 2024, 20.

22 *M. Cornils*, Anmerkung, ZJS 2018, 627 (636).

23 Zum Verhaltensparadigma *C. Engel/L. O'Hara/S. Egidy/Y. Hermstrüwer/L. Hoeft/P. Langenbach*, Öffentliches Recht als Verhaltensordnung, 2025; C. Engel/M. Englerth/J. Lüdemann/I. Spiecker gen. Döhmann (Hrsg.), Recht und Verhalten, 2007.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

scheidungsfreiheit als bedroht empfinden.[24] Dieser wahrgenommene Angriff auf ihre Autonomie verletzt das psychologische Grundbedürfnis, Entscheidungen als selbstbestimmtes Ergebnis des eigenen Willens zu erleben und nicht als bloßes Resultat von Zwang.[25] Je stärker die eigene Autonomie bedroht wird und je ungerechtfertigter dies dem Einzelnen erscheint, desto stärker wird tendenziell der Drang, die Selbstbestimmung wiederherzustellen. Das kann bei staatlichem Zwang etwa geschehen, indem man sich ihm zu entziehen versucht, oder indem man Gegenargumente formuliert, Protest und Widerstand organisiert, eine Klage anstrengt oder die Institution, von der der Zwang ausgeht, öffentlich in Frage stellt.

Selbstverständlich gibt es staatliche Beschränkungen der Entscheidungsfreiheit auch in anderen Bereichen des Rechts, gerade auch des Abgabenrechts. Doch das ungewöhnliche Ausmaß, in dem sich gerade beim Rundfunkbeitrag die psychologische Reaktanz Bahn bricht, deutet darauf hin, dass der staatliche Zwang gerade bei dieser Abgabe von vielen Bürgerinnen und Bürgern als besonders stark und/oder besonders ungerechtfertigt wahrgenommen wird.

3. Auslöser und Verstärker von Reaktanz beim Rundfunkbeitrag

Warum das der Fall ist, lässt sich mit Hilfe psychologischer und verhaltensökonomischer Theorien erklären. Sie erlauben es, strukturelle Auslöser und Verstärker von Reaktanz beim Rundfunkbeitrag zu identifizieren.

a) Ausgestaltung als Gegenleistungsabgabe

Ein zentraler Auslöser liegt bereits in der gesetzgeberischen Entscheidung, die Finanzierung des öffentlich-rechtlichen Rundfunks als Gegenleistungsabgabe für jedermann auszugestalten. Die allermeisten Menschen vollziehen Rechtspflichten nicht in jedem Detail nach. Sie orientieren sich

24 *J. W. Brehm*, A Theory of Psychological Reactance, 1966; *R. A. Wicklund*, Freedom and Reactance, 1974.

25 *R. M. Ryan/E. L. Deci*, Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being, American Psychologist 55 (2000), 227.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

vielmehr etwa an grundlegenden Fairnessintuitionen und Heuristiken,[26] mit denen sie die Anforderungen des Rechts im Alltag ohne großen kognitiven Aufwand wahrnehmen und bewerten können.[27] Wenn die Rechtsordnung eine Abgabe wie hier als Beitrag bezeichnet und entsprechend ausgestaltet, betrachten die Bürgerinnen und Bürger sie etwa häufig wie eine Art Vertragsverhältnis und bewerten es danach, ob sie eine angemessene Gegenleistung für den von ihnen zu zahlenden Preis erhalten.[28]

Geht man gedanklich von einem vertragsähnlichen Verhältnis aus, kann es beim Rundfunkbeitrag nicht nur als unangemessener Zwang wahrgenommen werden, wenn man dem Austauschverhältnis persönlich gar nicht entgehen kann (außer durch die lebensfremde Aufgabe der eigenen Wohnung). Diejenigen, die das öffentliche Medienangebot nicht oder nur sehr selten in Anspruch nehmen – sei es aus Ablehnung, Desinteresse oder einer veränderten Mediennutzung in jüngeren Generationen – können den Beitrag insbesondere auch als Zahlungspflicht ohne entsprechende Gegenleistung empfinden. Da dieses gestörte Reziprozitätsempfinden die Vorstellung eines gerechten Vertrags und damit das grundlegende menschliche Bedürfnis nach Fairness verletzt,[29] kann die Beitragspflicht als ungerechtfertigter Zwang wahrgenommen werden, der Reaktanz auslöst.

Dass juristisch betrachtet allein schon die potentielle Nutzungsmöglichkeit der Angebote eine beitragsauslösende Leistung sein soll, dürften die meisten hingegen weder wahrnehmen noch durch die Brille einer vertraglichen Austauschbeziehung als angemessene Gegenleistung bewerten. Dazu ist die an die Wohnung gekoppelte Möglichkeitskonstruktion beim Rundfunkbeitrag – im Unterschied zu klassischen Beiträgen wie dem Kur- oder Erschließungsbeitrag – viel zu kontraintuitiv.

---

26 *A. Tversky/D. Kahneman*, Judgment under Uncertainty: Heuristics and Biases, Science 183 (1974), 1124.

27 Allgemein *H. A. Simon*, A Behavioral Model of Rational Choice, The Quarterly Journal of Economics 69 (1955), 99.

28 Allg. dazu *S. Leder/L. Mannetti/E. Hölzl, E. Kirchler*, Regulatory Fit Effects on Perceived Fiscal Exchange and Tax Compliance, The Journal of Socio-Economics 39 (2010), 271; *L. P. Feld/B. S. Frey*, Tax Compliance as a Result of Psychological Tax Contract, Law & Policy 29 (2007), 102.

29 *J. H. Ellard/A. Harvey/M. J. Callan*, The Justice Motive: History, Theory, and Research, in: S. Sabbagh/M. Schmitt (Eds.), Handbook of Social Justice Theory and Research, 127 (127).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Als unfairer Zwang kann die Rundfunkbeitragspflicht schließlich auch von denjenigen empfunden werden, die das öffentliche Medienangebot zwar nutzen, den Preis für die erbrachte Gegenleistung aber als überzogen erachten.

b) Beitragspflicht neben privaten Medienausgaben

Im Unterschied zu den meisten anderen Gegenleistungsabgaben kommt beim Rundfunkbeitrag hinzu, dass dieser in der dualen Medienordnung häufig zusätzlich zu weiteren Zahlungen für private Medien entrichtet wird. Diese finanzielle Zusatzbelastung kann einen sog. Budgeteffekt auslösen. Nach der Theorie des *mental accounting* ordnen viele Menschen Zahlungen gedanklich unterschiedlichen Konten etwa für Lebensmittel, Urlaub oder Miete zu und behandeln diese Konten wie getrennte Kassen, auch wenn Geld objektiv fungibel ist.[30] Wer bereits aus eigener Entscheidung für private Streamingdienste oder andere Medien bezahlt, kann das eigene Medienbudget bereits als ausgeschöpft erleben. Der Rundfunkbeitrag wirkt in diesem Fall wie eine aufgezwungene Budgetüberschreitung, die als besonderer Verlust und zugleich als Bruch mit der eigenen Budgetautonomie wahrgenommen werden kann.

c) Wohnungspauschale

Selbst wenn man den Rundfunkbeitrag trotz der Gegenleistungskonstruktion nicht ausschließlich als Preis für einen individuellen Vorteil begreift, sondern auch den Kollektivnutzen des Medienangebots in Rechnung stellt, kann die Abgabenpflicht als ungerechtfertigter Zwang empfunden werden. Geht es um den Beitrag aller zur Finanzierung von Staatsaufgaben, orientieren sich die meisten Menschen insbesondere am Leistungsfähigkeitsprinzip:[31] Wer mehr Einkommen hat, soll größere Lasten tragen; wer weniger hat, soll weniger belastet werden. Dieser aus dem progressiven Steuersystem vertrauten Fairnesserwartung entspricht der Rundfunkbeitrag aber gerade auch nicht. Weil er als Wohnungspauschale erhoben wird, zahlt ein einkommensschwacher Singlehaushalt, wie wir gesehen

30 *R. H. Thaler*, Mental Accounting Matters, Journal of Behavioral Decision Making 12 (1999), 183.

31 *R. A. Musgrave/P. B. Musgrave*, Public Finance in Theory and Practice, 5. Aufl., 1989, 218 (228).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

haben, denselben Beitrag wie ein vermögender Mehrpersonenhaushalt. Anders als bei der Einkommensteuer gibt es keine Progression und keine Freigrenzen.

Dieser Konflikt mit den distributiven Gerechtigkeitsvorstellungen dürfte beim Rundfunkbeitrag auch deshalb reaktanzauslösend oder -verstärkend wirken, weil das Angebot der Rundfunkanstalten häufiger von Menschen mit hohem Einkommen genutzt wird als von Geringverdienern.[32] Für Letztere kann der Rundfunkbeitrag deshalb gleich eine doppelte Zumutung bedeuten: Obwohl sie die Angebote nicht oder nur wenig nutzen, sollen sie gemessen am Einkommen überproportional viel zur Finanzierung beitragen; eine mit Fairnessintuitionen schwer zu vereinbarende Umverteilung von unten nach oben.

d) Finanzierung werterelevanter Güter

Von herausragender Bedeutung dürfte für die psychologische Reaktanz darüber hinaus sein, dass mit dem Rundfunkbeitrag ein öffentliches Angebot finanziert wird, das im Unterschied zu vielen anderen Gegenleistungsabgaben für die eigene Selbstwahrnehmung besonders relevant ist.

Bei den meisten Menschen steht der Medienkonsum in einer mehr oder minder engen Verbindung zu den eigenen Wertvorstellungen oder politischen Überzeugungen.[33] Das gilt namentlich für Inhalte, die über reine Unterhaltungsformate hinausgehen. Die wenigsten Menschen abonnieren etwa eine Tageszeitung, deren Ausrichtung sie ablehnen. Weil der Rundfunkbeitrag aber von allen zu entrichten ist, müssen ihn auch diejenigen zahlen, die sich an der inhaltlichen Ausrichtung bestimmter Formate stören oder die sogar eine politische Tendenz des Programms wahrnehmen, die sie ablehnen oder für unausgewogen halten.[34] Und diese Wahrnehmung ist nicht nur in einem bestimmten politischen Lager verbreitet, sondern nach dem sog. *Hostile Media*-Effekt besteht bei starker persönlicher Involviertheit allgemein die Tendenz, dass man selbst In-

32 Für Großbritannien etwa *Ofcom*, Exploring D and E Socio-Economic Groups' Relationship with the BBC, 2023, 3.

33 Vgl. etwa *S. Iyengar/K-S. Hahn*, Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use, Journal of Communication 59 (2009), 19.

34 Um Letzteres geht es auch im jüngsten Beitragsurteil des BVerwG 6 C 5.24v. 15.10.2025.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

halte, die sich um Neutralität bemühen,[35] als voreingenommen empfindet – weshalb auch Gruppen mit gegensätzlicher Überzeugung dieselben Inhalte gleichzeitig als jeweils gegen sie gerichtet wahrnehmen können.[36]

Wer einen Konflikt mit seinen eigenen Wertvorstellungen erlebt, kann daher eine doppelte und besonders gravierende Beschneidung seiner Autonomie verspüren: Er muss nicht nur für ein Angebot bezahlen, das er nicht bestellt hat und nutzt, sondern er wird auch noch gezwungen, Inhalte mitzufinanzieren, die in seiner Wahrnehmung mit der eigenen sozialen oder politischen Identität konfligieren. Da das Ausmaß psychologischer Reaktanz insbesondere auch vom Stellenwert der jeweils bedrohten Entscheidungsfreiheit abhängig ist,[37] dürfte die gesetzgeberische Entscheidung, gerade ein so identitätssensibles Gut wie ein Medienangebot durch eine Pflichtabgabe für jedermann zu finanzieren, einer der zentralen Reaktanz-Verstärker sein und erklären, warum manche Menschen geradezu erbitterten Widerstand gegen den Rundfunkbeitrag leisten.

e) Hohe Salienz

Zusätzlich getriggert werden kann die psychologische Reaktanz durch die hohe Salienz des Rundfunkbeitrags. In Experimenten hat sich gezeigt, dass hoheitliche Abgaben, die einzeln und prominent ausgewiesen werden, deutlich weniger akzeptiert werden als identische Abgabenforderungen, die nicht gleichermaßen ins Auge stechen.[38] Beim Rundfunkbeitrag, der als eigener Buchungsposten regelmäßig auf dem Kontoauszug erscheint bzw. an den Beitragsservice überwiesen wird, ist die Sichtbarkeit im Vergleich zu anderen Abgaben besonders hoch. Zugleich kann die wiederkehrende Zahlung die psychologische Reaktanz immer wieder aufs Neue befeuern.

35 Welche Hindernisse diesen Anspruch begrenzen, zeigt die Forschung zu Media Bias, z. B. *D. D'Alessio/M. Allen*, Media Bias in Presidential Elections, Journal of Communication 50 (2000), 133.

36 *R. P. Vallone/L. Ross/M. Lepper*, The Hostile Media Phenomenon, Journal of Personality and Social Psychology 49 (1985), 577.

37 *S.-M. Hong/S. Faedda*, Refinement of the Hong Psychological Reactance Scale, Educational and Psychological Measurement 56 (1996), 173.

38 *R. Chetty/A. Looney/K. Kroft*, Salience and Taxation: Theory and Evidence, American Economic Review 99 (2009), 1145.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

f) Verfahrenslogik

Ein letzter Grund liegt schließlich in der Verfahrenslogik des Rundfunkbeitrags, die mit den Erwartungen an prozedurale Fairness[39] konfligieren und auch deshalb das Gefühl ungerechtfertigten Zwangs verstärken kann. Obwohl der Rundfunkbeitrag viele Fragen aufwirft, liegt es in der Natur des Massenverfahrens, dass Einwände der Beitragszahler mit vorgestanzten Standardbriefen beantwortet werden. Das gibt dem Einzelnen regelmäßig nicht das für die prozedurale Fairness vorteilhafte Gefühl, einer ernsthaften und wertschätzende Auseinandersetzung mit seinen Einwänden.

Vor allem aber wird beim Rundfunkbeitrag auch die staatliche Anwendung von Zwang durch den breiten Einsatz des Verwaltungsvollstreckungsrechts bis hin zur Erzwingungshaft besonders deutlich wahrnehmbar. Das kann die Reaktanz noch einmal deutlich erhöhen, weil der Staat die als ungerechtfertigt wahrgenommene Beitragspflicht mit aller Macht durchsetzt. Und für diese Wahrnehmung ist es noch nicht einmal erforderlich, dass der Einzelne den Zwang persönlich erfährt. Durch mediale Berichte über die konsequente Anwendung der Verwaltungsvollstreckung kann die Zwangskulisse nach der sog. Verfügbarkeitsheuristik[40] allein schon durch die bloße Präsenz entsprechender Nachrichten für die eigene Wahrnehmung von besonderer Bedeutung sein.

g) Dominante Zwangswahrnehmung

Weil der Zwang aus all diesen Gründen als besonders ungerechtfertigt und massiv erlebt werden kann, ist es nicht überraschend, dass gerade hier die psychologische Reaktanz bei vielen Bürgerinnen und Bürgern ungewöhnlich stark ausgeprägt ist. Sie bündelt sich wie in einem Brennglas in der Rede vom „Zwangsbeitrag". Weil dieses Framing den emotional aufgeladenen Widerstand pointiert einfängt, ist der Begriff in der öffentlichen Debatte entsprechend wirkmächtig.[41]

---

39 Zur Bedeutung prozeduraler Fairness *E. A. Lind/T. R. Tyler*, *The Social Psychology of Procedural Justice*, 1988; *S. Sunshine/T. R. Tyler*, The Role of Procedural Justice and Legitimacy in Shaping Support for Policing, Law & Society Review 37 (2003), 513.

40 *A. Tversky/D. Kahneman*, Availability: A Heuristic for Judging Frequency and Probability, Cognitive Psychology 1973, 207.

41 Zum Mechanismus *A. Tversky/D. Kahneman*, The Framing of Decisions and the Psychology of Choice, Science 1981, 453.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Im Sinne des in der Steuerpsychologie etablierten *Slippery Slope Framework* erzeugt die Kombination aus mangelnder Akzeptanz der Pflicht und staatlichem Zwang insgesamt ein antagonistisches Klima, in dem der Staat eher als Gegner denn als Garant des Gemeinwohls erlebt wird.[42] Insoweit erweist sich die Entscheidung für den Rundfunkbeitrag auch über die Medienfinanzierung hinaus als ein rechtspolitischer Problemfall.

IV. Für einen Pfadwechsel in der Medienfinanzierung

Der im Modell des Rundfunkbeitrags angelegte strukturelle Widerstand lässt sich nicht mit Detailkorrekturen beheben. Und ebenso wenig kann er mit politischer Kommunikation aufgelöst werden. Im Gegenteil kann etwa der bisweilen unternommene Versuch, die Rede vom „Zwangsbeitrag" mit einem Begriff wie „Demokratieabgabe" zu überschreiben, die Reaktanz aus verschiedenen Gründen sogar noch verstärken. Erstens kann bereits die Verbindung eines Fundamentalwerts wie der Demokratie mit einer Geldzahlungspflicht als unpassend empfunden werden; in der Psychologie spricht man insoweit von einem *taboo trade-off*.[43] Zweitens kann das Framing als Demokratieabgabe bei einer kritischen Haltung zur Beitragspflicht als unfaire Unterstellung mangelnder Demokratietreue wahrgenommen werden. Und drittens schließlich: Wenn der Rundfunkbeitrag als Demokratieabgabe gelabelt wird, erzeugt das die subjektive Erwartung, dass mit dem Beitrag vor allem demokratierelevante Inhalte finanziert werden. Wenn Rezipientinnen und Rezipienten im Programm dann jedoch einen großen Anteil von Krimis, Kochshows oder Volksmusik wahrnehmen, kann dies im Sinne der *Expectancy Violations Theory* als Erwartungsverletzung erlebt werden[44] und den Eindruck eines bewussten Etikettenschwindels hervorrufen (mag der umfassende Funktionsauftrag der Rundfunkanstalten vom Bundesverfassungsgericht auch demokratiefunktional begründet sein).

Es spricht insgesamt viel dafür, den Finanzierungspfad der Gegenleistungsabgaben rechtspolitisch zu verlassen. Eine alternative Finanzierung aus Steuer-/Haushaltsmitteln wäre nicht nur mit Blick auf die Ver-

42 *E. Kirchler*, The Economic Psychology of Tax Behaviour, 2007, 21 ff.

43 *A. P. Fiske/P. E. Tetlock*, Taboo Trade-offs, Political Psychology 18 (1997), 255.

44 *J. K. Burgoon*, Interpersonal Expectations, Expectancy Violations, and Emotional Communication, Journal of Language and Social Psychology 12 (1993), 30.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

teilungsgerechtigkeit angemessener, mit geringeren Vollzugskosten verbunden und bei hinreichender Absicherung der Staatsferne auch verfassungsrechtlich zulässig.[45] Sie wäre auch deutlich verhaltenssensibler, weil die Finanzierung der öffentlichen Medien wie bei vielen anderen Gütern an der normativen Selbstverständlichkeit der Steuer als akzeptierter, kulturell verfestigter und solidarischer Finanzierungsform für staatliche Aufgaben teilhätte.

Das bedeutet nicht, um nicht missverstanden zu werden, dass man über die öffentlichen Medienangebote und ihre Kosten nicht streiten kann und sollte. Wie notwendig eine solche Debatte ist, hat *Christoph Engel* bereits vor vielen Jahren deutlich gemacht.[46] Kein freiheitlicher Staat sollte es sich aber leisten, allein durch die Wahl eines Finanzierungsmodells massiven Widerstand bei den Bürgerinnen und Bürgern zu provozieren. Das gilt in besonderer Weise auf einem so demokratierelevanten Feld wie der Medienpolitik.

45 Zur verfassungsrechtlichen Zulässigkeit überzeugend *C. Waldhoff*, Verfassungsrechtliche Fragen einer Steuer-/Haushaltsfinanzierung des öffentlich-rechtlichen Rundfunks, AfP 2011, 1.; vgl. auch *M. Cornils*, Anmerkung, ZJS 2018, 627 (638).

46 Z. B. *C. Engel*, Rundfunk in Freiheit, AfP 1994, 185; *ders.*, Medienordnungsrecht, 1996.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Using Behavioural Insights When Designing Regulatory Measures

Morten Broberg

A. The behavioural turn – and the law

Over the past half-century, leading scholars such as Engel, Thaler, Sunstein, Kahneman, Tversky, and Jolls have shaped what is now termed 'the behavioural turn' – a paradigm shift in the social sciences that revises classical assumptions about human rationality through insights from psychology and behavioural research. Rather than viewing individuals as fully rational decision-makers, it is acknowledged that behaviour is often influenced by cognitive limitations, social norms and emotional factors. This understanding has enabled new approaches to analysing and guiding behaviour, notably via behaviour-based policy instruments such as nudging.

In legal scholarship, the application of behavioural insights is known as 'behavioural law and economics' (BLE) where Professor Engel is a shining beacon. This Festschrift contribution contends that legislators may enhance the efficacy of traditional regulatory measures by integrating behavioural insights into the design process.

Section B outlines BLE and regulatory design. Section C demonstrates how BLE may support the development of 'smarter regulation'. Section D considers the broader implications of integrating BLE into legislative processes.

B. Behavioural Law and Economics – and Regulatory Design

Within the legal domain, it is widely acknowledged that humans often act irrationally – indeed, many regulatory measures are adopted precisely to address this. This stands in contrast to neoclassical economic theory, which assumes that economic actors (consumers and businesses) behave rationally, seeking to maximise utility (consumers) or profit (businesses).

When a public authority enacts a regulatory measure, the explicit aim is often to influence behaviour. Even when not stated, regulatory mea-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

sures typically have such impact. It is therefore advisable to incorporate behavioural insights into the drafting process. In this respect, two key questions arise: what behaviour is desired, and how can it best be promoted.

A pertinent example is organ donation legislation. Its overaching goal is to ensure organ availability by allowing individuals to decide whether they wish to donate upon death. The desired behaviour is that all willing donors register accordingly. Legislators typically choose between two models: automatic registration with an opt-out option, or no registration with an opt-in option. Although the decision to donate is far from trivial, studies show that opt-out systems yield significantly higher donor rates.[1] Given the general shortage of organs, countries with opt-in systems have increasingly moved towards opt-out models. The United Kingdom, for instance, has made this transition.[2]

Regulatory measures that influence behaviour may have unintended consequences. This risk is especially pronounced when behavioural insights are explicitly embedded in the design. Before the measure enters into force, it is crucial to clarify how it will function in practice. This is typically done through trials to gauge citizens' response – though such trials require behavioural insights expertise as well as time, both of which are scarce in fast-paced legislative environments.

There are numerous examples of regulation producing unexpected behavioural effects. For instance, financial service providers have been required to include warnings about product risks. Yet some consumers interpret these warnings as signs of official approval, assuming the products are safer – thus undermining the intended cautionary effect.[3]

In the present author's view, behavioural impact should always be considered during the design phase of regulatory measures. This is a complex task. A simplified framework is presented below, based on the British Behavioural Insights Team's (BIT) EAST model.

1 Albertsen, 60.

2 England implemented an opt-out model on 20 May 2020. See https://www.odt.nhs.uk/statistics-and-reports/opt-out-legislation-reports/.

3 Halpern (2019), 177.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. The EAST model

I. Why use the EAST model?

Modern behavioural science rests on a range of complementary theories. Researchers in the field often work within a specific theoretical framework, whereas practitioners tend to focus on how these theories can be applied in real-world contexts. The latters' interest lies in translating theory into a practical 'toolbox'.

Several such toolboxes have been developed to help civil servants and other practitioners apply behavioural insights when designing regulatory measures. Among the most well-known are Pelle Guldborg Hansen's BASIC model.[4] Paul Dolan and Ivo Vlaev's SNAP model.[5] The MINDSPACE model developed by the British Institute for Government.[6] And the EAST model developed by the BIT.[7]

Each model has its strengths and limitations. For the purposes of this contribution, the EAST model is the most suitable. The BIT describes it in the following terms:

> EAST is a simple, pragmatic framework for practitioners, and that simplicity seems to have been a strength. It continues to be a touchstone for our projects; it has spawned a set of cards, in several languages, that help people generate ideas in new ways; it can communicate the basics of behavioural insights in a few minutes.[8]

EAST comprises four partly overlapping principles: Make it Easy, Attractive, Social, and Timely. It is deliberately simple and designed for practical use. Widely applied and revised in 2024, the model highlights the behavioural factors most relevant in regulatory design. It also serves as an effective educational tool for demonstrating the value of behavioural insights in this context. However, its simplicity comes at a cost: it captures only a limited subset of the behavioural insights relevant to regulatory design.

4 OECD.
5 Vlaev and Dolan.
6 Institute for Government.
7 BIT. Halpern (2019), 60, explains the origin of the (peculiar) acronym.
8 BIT, 2.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Easy

1. It's all about friction

Citizens are more inclined to "do something" that they perceive as easy than something they perceive as difficult. Even where what needs to be done is solely in their own interests. 'Friction' – any obstacle or effort required – thus affects decision-making, and even minimal friction can have a surprisingly strong impact.

Thaler and Sunstein have illustrated this with the example of pension scheme enrolment. When employees must actively sign up, participation rates drop – even when no financial cost is involved. The two authors write:

> To be sure, there are situations, say for young workers with other pressing financial needs, in which it could be sensible not to join [a pension scheme] even with an employer match. But in many cases, the failure to join is simply a blunder. One extreme example comes from the United Kingdom, where some defined-benefit plans do not require any employee contributions and are fully paid for by the employer. They do require employees to take action to join the plan. Data on twenty-five such plans reveal that scarcely half of the eligible employees (51 per cent) signed up! This is equivalent to not bothering to cash your paycheque.[9]

Of the four EAST principles, 'Make it Easy' is arguably the most crucial. When regulation requires citizens to make a choice, drafters should adopt the citizen's perspective, ensuring that the simplest option aligns with both individual and societal interests.

"Make it Easy" may be divided into three components; (1) default settings, (2) practical and cognitive barriers, and (3) simplifying the message.

2. Defaults

Defaults are ubiquitous. For instance, when sending a document to a printer, settings such as single- or double-sided printing are pre-selected. While users may change these settings, most accept the default.[10]

Defaults also play a significant role in legislation. The above example of organ donation illustrates this well: some countries adopt opt-in

9 Thaler and Sunstein, 117. At 263 the two authors note: "We thank David Blake and the U.K. Department of Work and Pensions for providing us with the data".

10 Rutgers, Office of Information Technology, and Egebark and Ekström.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

systems, others opt-out.[11] Where the default is 'organ donor' (opt-out), registration rates are markedly higher than in opt-in countries.

Organ donation is perhaps the most cited example of legislative defaults, though its ethical sensitivity necessarily influences policy choices.[12] However, many less sensitive cases also demonstrate the importance of default settings in regulatory design.

One such example comes from Switzerland, where researchers studied the impact of making renewable energy the default for existing customers. Prior to the change, uptake ranged from 0.7 % to 3 %. Afterwards, it rose dramatically to between 77 % and 84.7 %, with the effect persisting over several years.[13]

A key point in using defaults is that when regulation offers a choice, the *presentation* of that choice matters. For instance, if public institutions are required to offer meals with or without meat to support climate goals (since meals without meat have a lower climate impact), the default option – meat or meat-free – will significantly influence behaviour.[14]

3. Practical and cognitive barriers

Firms offering subscriptions – such as gyms and streaming services – know that once automatic payments are set up, subscriptions often continue for long periods, even if unused. The need for the subscriber to *do something* to cancel creates a barrier, often with significant effect.[15]

When drafting regulatory measures aimed at promoting specific behaviours, drafters should identify all practical and cognitive barriers, including those that may seem minor. This requires adopting the citizen's perspective. Often the citizens are made up of several heterogeneous groups that are faced with different barriers – meaning that the drafters must consider a variety of barriers.

A small experiment at Rutgers University in the USA illustrates the importance of even minor barriers well. In many countries, free flu vaccinations are offered, yet uptake remains low. To test whether removing

11 In addition to opt-in and opt-out, other schemes also exist.

12 Albertsen.

13 Liebe *et al.*

14 Other aspects of communication are addressed in section C.II.4 below.

15 Default options – as presented in section C.II.2 above – illustrate this point, since choosing the default requires no action whereas choosing an alternative does.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a minor barrier could improve participation, university employees were split into two groups. One group received an email inviting them to book a vaccination appointment online. The other group received an email with a pre-scheduled time and location; they only needed to act if they wished to reschedule or decline. The second group showed significantly higher participation, demonstrating that even small frictions – like booking an appointment – can deter action.[16]

Regulatory drafters should therefore seek out such frictions, as they may undermine intended outcomes. However, friction can also be used constructively. In the UK and Denmark, for example, supermarkets may only sell one packet of painkillers per transaction. This restriction was introduced because the painkiller paracetamol is frequently used in suicide attempts. Remarkably, this small barrier has led to a measurable decline in such attempts, showing that friction can, in some cases, serve a protective function.[17]

4. Make communication simple and easily accessible

Prohibitions and injunctions – the classic legal instruments – are not always suitable. In such cases, drafters may turn to economic incentives or to simply provide information. Crucially, this information must be communicated in a way that is *easily accessible* to citizens.

Public authorities often produce detailed reports containing all relevant information, yet these reports frequently go unread. The key to effective communication lies in simplicity and accessibility.

A good example of a legislator using easily accessible communication as an efficient regulatory means is the EU's energy labelling scheme for household appliances – aimed at the consumer. Instead of lengthy technical reports, the scheme uses a straightforward scale from A to G, enabling consumers to easily compare energy efficiency when choosing products.[18]

16 Chapman *et al*. The study was limited to a university workplace sample, and vaccination records were limited to vaccinations received at the occupational health department. Still, the authors conclude that the results suggest that automatic scheduling of flu shot appointments may be an effective way to increase vaccination rates.

17 Limiting customers' access to certain products (such as tobacco) to the sales counter is a similar way of creating 'friction'.

18 Regulation 2017/1369 of 4 July 2017 setting a framework for energy labelling, OJ 2017, L198/1. See similarly Regulation 2020/740 of 25 May 2020 on the labelling of tyres with respect to fuel efficiency and other parameters, OJ 2020, L177/1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. Attract

When designing regulatory measures, drafters should consider how to make the desired behaviour more attractive. This can be achieved by increasing its salience and reducing that of undesired behaviour, or by introducing targeted incentives.

For example, the Scottish Parliament has restricted the display of alcohol in retail outlets to reduce impulse purchases by making alcohol less visible.[19]

Subsidies and taxes are among the most common regulatory incentives, but their design is critical. The UK's 2018 tax on sugary soft drinks is a notable example.[20] Rather than targeting consumers, the tax was structured to influence producers: the higher the sugar content, the higher the tax. This encouraged reformulation, resulting in a 46% reduction in sugar content in soft drinks between 2015 and 2020 – allowing consumers to maintain consumption while lowering intake.[21]

Not all incentives are financial. Blood donors, volunteers supporting the elderly, or those running homework cafés are typically motivated by altruism. Introducing monetary rewards may undermine this motivation. For instance, offering €8 to blood donors – who previously gave blood freely – can be perceived as devaluing their contribution. Instead of feeling they have done something meaningful, donors may compare the modest compensation with the inconvenience and physical effects of giving blood.

IV. Social

The fear of 'standing out' or being seen 'doing something wrong' can be a powerful behavioural driver. For instance, many roads feature digital speed displays: if a driver complies with the limit, their speed appears in plain numbers; if they exceed it, flashing lights are added – creating the impression with the driver that others can see their transgression.

---

19 Licensing (Scotland) Act 2005, Schedule 3, section 13. For another example, see Directive 2003/33 of 26 May 2003 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the advertising and sponsorship of tobacco products, OJ L152/16.

20 See, in particular, The Soft Drinks Industry Levy Regulations 2018.

21 BIT, 23, with further references.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Most people prefer not to deviate from the norm. Convincing citizens that a particular behaviour is typical can be highly effective – provided the behaviour aligns with the one the regulators want to promotoe. If the norm is undesirable, however, highlighting it may be counterproductive. For example, a campaign to combat violence in the night life stating that many young people carry knives at night may inadvertently normalise the undesired behaviour. Similarly, pointing out widespread bribery among civil servants as part of a campaign to combat corruption carries the risk of reinforcing such corrupt practices.[22]

Moreover, appeals to social norms only work if the target group identifies with the reference group. Telling elderly men what is typical among young women – or *vice versa* – is unlikely to produce meaningful behavioural change.

V. Timely

The fourth element of the EAST model – Timely – encompasses two dimensions:

(1) The timing of behavioural interventions, including time of day,[23] week, or year, as well as major life events such as marriage, parenthood, relocation, or job changes.[24]
(2) The way individuals respond to events depending on whether they occur in the present or the future.

When designing regulatory measures to influence behaviour, careful attention should be paid to the timing of communication. The EU's Denied Boarding Regulation illustrates this well.[25] It grants air passengers rights

22 The social psychologist Robert Cialdini has referred to the risk of legitimising an unwanted behaviour in this way as the 'big mistake', cf Halpern (2015) and (2019), 35.

23 An examination of using text messages to blood donors to keep a constant flow of new donations showed that messages sent in the evening increased donations by 6.5 %, cf Fosgaard *et al.*

24 Less significant disruptions may also be relevant. For example, when car owners switch from a fossil fuel car to an electric vehicle, they are likely to be particularly open to changing their habits with regards to their cars; such as switching to a different car insurer, cf Wendt Jensen, 73.

25 Regulation 261/2004 of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, OJ2004 L46/1 and Broberg.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

in cases of cancellation, delay, or denied boarding, and specifies when and how carriers must inform passengers. Timely notification is essential for the regulation to be effective.

The adace 'an ounce of prevention is worth a pound of cure' should also be considered by drafters of regulatory measures. Preventing bad habits is often more effective than correcting them later. For example, start-up entrepreneurs frequently neglect compliance with tax and VAT rules while focusing on business development. Once non-compliance becomes habitual, it is harder to reverse. Authorities should therefore intervene early, when habits are still forming.

Timing also matters where the state promotes a certain behaviour by offering subsidies to citizens. Programmes tend to be more successful when a substantial portion of the subsidy is paid only after the desired behaviour has been completed.[26]

The second dimension of timing reflects the human tendency to prioritise immediate outcomes. People are more motivated by short-term rewards than long-term benefits, and more willing to accept future obligations than immediate ones. Thus, regulatory measures may be more effective if they offer immediate rewards for desired behaviour or impose immediate costs on undesired actions.

Finally, when regulation requires citizens to follow specific procedures, success rates improve if barriers to following the procedures are identified and guidance to overcome the barriers is provided. For instance, unemployed citizens receiving benefits are often required to apply for jobs and attend training. If job centres help them develop concrete plans to meet these requirements, the success prospects increase considerably.[27]

D. The future of behavioural insights in the design of regulation

The above overview covers only a fraction of the behavioural insights relevant to regulatory design. Nonetheless, there is little doubt that such insights can significantly enhance the effectiveness of new measures.

26 See further Halpern (2019), 135–137.

27 Halpern (2019), 9. A similar situation is where patients are discharged from hospital and are thereupon required to perform certain treatment tasks themselves.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Incorporating behavioural insights into the process of drafting regulatory measures presents three key practical challenges. *First*, most civil services rely primarily on prohibitions, mandatory orders, and economic incentives (e.g. subsidies and taxes). Recognising that factors such as timing may be equally important requires a shift in mindset.[28] *Second*, applying behavioural insights is time-consuming and may extend the drafting process – often an impractical option. *Third*, successful integration depends on access to behavioural expertise, which is currently limited in many public administrations.

Moreover, ethical concerns may arise, as illustrated by the example of organ donation defaults.[29]

Despite these challenges, the benefits of integrating behavioural insights into regulatory design are compelling. It therefore appears likely that future regulation will increasingly draw on these insights during the drafting phase.

Bibliography

Albertsen, Andreas, *Organdonation – og behovet for en ny model*, Aarhus Universitetsforlag 2020

BIT (Behavioural Insights Team), *EAST – Four simple ways to apply behavioural insights – revised and updated edition*, 2024, https://www.bi.team/wp-content/uploads/2014/04/BIT-EAST-1.pdf

Broberg, Morten, 'Air Passengers' Rights in the European Union: The air carriers' obligations *vis-à-vis* their passengers under Regulation 261/2004', *Journal of Business Law*, 2009, 727–742

Chapman, Gretchen & Li, Meng & Colby, Helen & Yoon, Haewon, 'Opting In vs Opting Out of Influenza Vaccination', *JAMA: the journal of the American Medical Association*. 304. 43–4. https://doi.org/10.1001/jama.2010.892

Duflo, Esther & Kremer, Michael & Robinson, Jonathan, 'Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya', *American Economic Review* 2011: 2350–90.DOI: https://doi.org/10.1257/aer.101.6.2350

Egebark, Johan & Ekström, Mathias, 'Can indifference make the world greener?', *IFN Working Paper*, 2013 No. 975, Research Institute of Industrial Economics (IFN), Stockholm

Fosgaard Toke & Hansen Lars & Jacobsen Catrine & Sørensen Erik & Romose Merle & Ullum Henrik, 'Can text messages save lives? A field experiment on blood donor

28 Duflo *et al.*

29 See also Hansen & Jespersen as well as Hansen.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

motivation', *Transfusion*. 2020 Mar;60(3):460-465. doi: https://doi.org/10.1111/trf.15633. Epub 2019 Dec 14. PMID: 31837032

Halpern, David, 'How Can Governments and Businesses Avoid the 'Big Mistake?', *Behavioral Scientist*, 5 December 2015, https://behavioralscientist.org/how-can-governments-and-businesses-avoid-the-big-mistake/

Halpern, David, *Inside the Nudge Unit – How small changes can make a big difference*, WH allen 2019

Hansen, Pelle Guldborg, 'The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?', *European Journal of Risk Regulation*, 2016 155–174

Hansen, Pelle Guldborg & Jespersen, Andreas Maaløe, 'Nudge and the Manipulation of Choice – A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy', *European Journal of Risk Regulation*, 2013 3–28

Institute for Government, *MINDSPACE – Influencing behaviour through public policy*, https://www.instituteforgovernment.org.uk/sites/default/files/publications/MINDSPACE.pdf

Liebe, Ulf & Gewinner, Jennifer & Diekmann, Andreas, 'Large and persistent effects of green energy defaults in the household and business sectors', *Nat Hum Behav* 5, 576–585 (2021). https://doi.org/10.1038/s41562-021-01070-3.

OECD, *Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit*, OECD Publishing 2019, https://doi.org/10.1787/9ea76a8f-en

Rutgers, Office of Information Technology, *Print Management Information*, 2013, https://perma.cc/03BNSKSY4oZ

Thaler, Richard & Sunstein, Cass, *Nudge – Improving decisions about health, wealth and happiness',* Penguin Books, revised edition 2009

Vlaev, Ivo & Dolan, Paul, *From changing cognitions to changing the context: a dual-route model of behavour change*, Imperial College, Business School, Discussion Paper 2009/04, https://core.ac.uk/download/pdf/297061.pdf

Wendt Jensen, Anders, 'Derfor stiger din bilforsikring', *Motor* 2025, 72–73

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Datenökonomie
Zwischen ordnungspolitischer Inversion und staatlicher Willkür

Klaus Heine

A. Einleitung

In diesem Beitrag soll aus rechtsökonomischer Perspektive der Verfasstheit des Dateneigentums nachgegangen werden. Dies ist kein ganz neues Thema, es taucht immer wieder in zum Teil hochkomplexen Diskussionen auf (mit einem Überblick siehe Kerber 2023). Zum Beispiel wenn es um das Eigentum von humanen Bio-Daten geht, dem Schutz der Privatsphäre, den Zugang zu technischen Daten für kleinere und mittlere Unternehmen, oder beim Urheberrecht an Texten, das zum Training von großen Sprachmodellen genutzt wird. Viele weitere Beispiele könnten genannt werden, in denen Eigentumsrechte den Handlungsspielraum zur Datennutzung bestimmen, und damit sowohl die ökonomischen Möglichkeiten für Marktteilnehmer als auch den normativ-ethischen Bezugsrahmen der Datenökonomie abstecken. Es geht bei einer Diskussion von Dateneigentum also immer um beides: Die Ausschöpfung des ökonomischen Potentials und die Passung in die Werteordnung einer Gesellschaft. Das ist zwar sicherlich kein Spezifikum von Daten allein, sondern gilt allgemein für das Eigentum an Dingen und den Ordnungsrahmen einer Gesellschaft.[1] Es kann jedoch gefragt werden, ob die historisch vorgefundene Eigentumsordung und deren rechtliche Verfasstheit noch den gesellschaftlichen Bedürfnissen der Datenökonomie gerecht wird.[2]

1 Hier sei nur an die ordoliberale Tradition in den Rechtswissenschaften erinnert. Franz Böhm, Ernst-Joachim Mestmäcker, Christoph Engel und Heike Schweitzer sind jeweils zeitgemäß und mit adäquatem Instrumentarium der Frage nachgegangen, wie eine humane Rechtsordnung sinnvoll mit der Wirtschaftsordnung verkoppelt werden kann.

2 Dies ist eine alte Frage der Rechtsökonomik, die erstmals von Demsetz (1967) ausbuchstabiert wurde und seitdem immer wieder aufgegriffen wird (z. B. Hardin 1968; Ostrom 1990; Ellickson 1991).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Daher soll die Frage dieses Beitrags sein, welche institutionellen Alternativen zur Ordnung des Dateneigentums grundsätzlich bestehen und ob es möglicherweise eine Ordnungsform gibt, die sowohl den ökonomischen als auch den ethisch-normativen Ansprüchen besser gerecht wird als die derzeitigen Ordnungsentwürfe. Dazu wird im nächsten Schritt rekapituliert, was Eigentum aus rechtlicher und ökonomischer Sicht bedeutet. Dann werden zugespitzt die beiden realen Eigentumsordnungen bezüglich Daten mit ihren generischen Governanceproblemen skizziert, bevor dann im letzten Schritt ein alternativer Vorschlag präsentiert wird, der der gesellschaftlichen Einbettung von Digitaltechnologie möglicherweise besser entspricht.

B. Eigentum an Daten – einige Grundüberlegungen

Eigentumsrechte beschreiben die Beziehung zwischen einem Gegenstand, Personen und der Gesellschaft. Eigentum definiert, wann ein Gegenstand unter die Kontrolle von jemandem fällt und wann ein Wechsel der Kontrolle stattfindet (Cooter und Ulen 2008, S. 77). Daraus folgen eine Reihe von Kernfragen des Eigentumsrechts: Wann liegt eine Verletzung des Eigentums vor, und wann erfordert dies eine Wiederherstellung des Ursprungszustands, oder wann ist eine Kompensation des Schadens angezeigt? Dies spiegelt sich auch in den bekannten Diskussionen rund um Gefährdungs- und Verschuldenshaftung wider (siehe bereits Coase 1960 oder Calabresi und Melamed 1972). Es geht also darum, wie „ausschließlich" Eigentum ist. Eigentumsrecht beschäftigt sich ebenfalls mit der Frage, welche Gegenstände welchen Eigentumsschutz haben. Dies ist besonders wichtig bei der Unterscheidung von tangiblen und intangiblen Sachen, wie intellektuellem Eigentum (IPR).

Geht man von diesen eigentumsrechtlichen Grundüberlegungen einen kleinen Schritt weiter, dann kommt man zu der Feststellung, dass Eigentumsrechte Verfügungs- bzw. Handlungsrechte für Akteure kreieren, und dass man diese so abstufen kann, wie man es als gesellschaftlich für notwendig erachtet.[3] Was ist jedoch gesellschaftlich wünschenswert? Dies ist eine normative Frage, für deren Beantwortung man sich mit ökonomischen

3 Bereits im Römischen Recht findet sich die Unterscheidung von Usus, Usus Fructus, Abusus und Ius Abutendi, um verschiedene Handlungsmöglichkeiten mit und an Sachen zu normieren.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Überlegungen nähern kann. So kann man zum Beispiel die Leistungsfähigkeit ganzer Eigentumsordnungen auf die Wirtschaft analysieren (Alchian 1950; Furubotn und Pejovich 1972). Man kann aber auch kleinteiliger untersuchen, wer das letztendliche Verfügungsrecht an einer Sache haben sollte, wenn ein Vertrag aufgrund von Unvollständigkeit schweigt (Hart 2017). Diese Überlegungen führen zu der Erkenntnis, dass es die Ausgestaltung der (residualen) Verfügungsrechte ist, die die Anreize von Akteuren steuert. Das Design von Verfügungsrechten kann freilich nicht unabhängig von den Charakteristika des Gutes betrachtet werden, um das es geht. Grundstücke, Häuser, Maschinen, Wissen von Mitarbeitern, Patente, natürliche Ressourcen usw. haben ihre jeweiligen technischen Unterschiede, die sich auf die Gestaltung der Verfügungsrechte auswirken.

Daten haben die Eigenschaft von Clubgütern (Buchanan 1965). Es besteht Nichtrivalität im Konsum, gleichzeitig können andere von der Nutzung ausgeschlossen werden. Das impliziert, dass eine Marktpreisbildung nach dem Marginalprinzip nicht möglich ist, sondern dass für die Nutzung der Leistungen des Clubs eine festzulegende Nutzungsgebühr anfällt. Wenn man die Clubmitgliedschaft bezahlt hat, kann man dem Club betreten und dessen Dienstleistungen frei nutzen. Sind Daten erstmal erstellt, können sie beliebig oft verwendet werden ohne sich abzunutzen. Gleichzeitig können Nutzer vom Gebrauch von Daten ausgeschlossen werden. Beispielsweise sind elektronische Zeitschriften von beliebig vielen Lesern nutzbar, aber der Ausschluss von Lesern, die nicht die Nutzungsgebühr bezahlen, ist möglich.

Neben der Clubguteigenschaft zeichnen sich Daten und die damit verbundenen Informationen durch Skalenerträge und Netzwerkeffekte aus. Der (Gebrauchs-) Wert von Daten steigt stark mit ihrer Kumulation an; zum Beispiel ist ein großes Sprachmodell umso besser, je grösser der Trainingsdatensatz ist. Die Eigenschaft als Clubgut und die Skalenerträge führen dazu, dass Daten auf Plattformen zurechtkommen. Dort werden sie mit einer Technologie zu ihrer Erschließung – typischerweise ein Algorithmus – zu einem Geschäftsmodell vereint.

Ein letztes Merkmal, das lange im Schatten der Diskussion stand, ist, dass Daten zwar intangibel sind, ihre Aufbewahrung und Übermittelung aber an eine physische Infrastruktur gekoppelt ist. Es bedarf riesiger Rechenzentren, Satelliten, Unterwasserkabel und Kraftwerke zur Stromerzeugung, um die Datenökonomie zu betreiben (Greenstein 2021).

Das Ergebnis all dieser Eigenschaften von Daten ist eine von wenigen marktmächtigen Unternehmen und ihren Plattformen dominierte Ökono-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

mie, wobei eine große Bandbreite an regulierungsbedürftigen Problemen entsteht. Dabei ist es fraglich, ob die traditionelle Werkzeugkiste von Gesetzen und Regulierungen geeignet ist, um mit diesen Problemen umzugehen.

C. Die zwei idealtypischen Ordnungsformen der Datenökonomie

Entsprechend der älteren ordnungsökonomischen Literatur gibt die Art der Koordination der Wirtschaftspläne Auskunft darüber, welches Wirtschaftssystem vorliegt. Wenn die Koordination dezentral über Marktpreise geschieht, liegt eine Marktwirtschaft vor. Erfolgt die Koordination über einen zentralen Plan, handelt es sich um eine Planwirtschaft. Daneben tritt die Eigentumsordnung als ein quasi subsidiäres Element, das etwas über die Akzeptanz der Wirtschaftsordnung in der Gesellschaft aussagt sowie die Stärke der Anreize für Akteure setzt, um zu einer Wirtschaftsordnung beizutragen (Leipold 1988). Mit dieser Unterscheidung konnte nach dem zweiten Weltkrieg der Systemvergleich zwischen West und Ost betrieben werden.

Anhand des traditionellen Bezugspaars Planung und Eigentum lässt sich der heutige wirtschaftliche Erfolg Chinas allerdings nicht gut erklären (Heine 2010). Denn offenbar ist China wirtschaftlich sehr erfolgreich, obwohl private Eigentumsrechte nicht garantiert sind und die Wirtschaftsplanung stark zentralisiert ist. Im Prinzip bestimmt die Kommunistische Partei über das Eigentum und die Wirtschaftspläne. Dies gilt insbesondere für die Datenökonomie, in der der chinesische Staat spezifische Vorgaben für die einzelnen Unternehmen macht (Global Times 2025).

Spiegelbildlich kann man fragen, wie stark der Gesetzgeber in den Vereinigten Staaten gegenüber den großen Technologieunternehmen ist und einen Wettbewerb auf offenen Märkten durchsetzen kann. Ein konstitutives Merkmal der Ordnungsökonomik ist ja, dass der Staat den Rahmen für einen Wettbewerb setzt, in dem alle Unternehmen den gleichen Spielregeln unterliegen. Das kann aber hinsichtlich der großen Technologieunternehmen bezweifelt werden, da der amerikanische Staat stark von den Leistungen dieser Unternehmen selbst abhängig ist. Sei es hinsichtlich digitaler Infrastruktur, Weltraummissionen oder Sicherheitstechnologie. Trotz dieser ordnungspolitischen Inversion erscheint die amerikanische (Digital-) Wirtschaft sehr vital.

Es ist daher sinnvoll, ausgehend von den im vorherigen Abschnitt beschriebenen Charakteristika, die Unterschiede von Datenökonomien

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

grundlegender anhand des residualen Kontrollrechts am Dateneigentum zu analysieren. Das residuale Kontrollrecht an Daten gibt Auskunft über die spezifischen Governanceprobleme, die eine Datenökonomie hat.

I. Die Vereinigten Staaten und ihr digitales Governance-Modell

Beim Blick auf die Vereinigten Staaten fällt als erstes auf, dass die fünf bekannten Unternehmen Apple, Amazon, Meta, Alphabet und Microsoft mit ihren digitalen Plattformen den US-amerikanischen Markt und darüber hinaus weite Teile der Welt dominieren. Die Technologieunternehmen sind mit ihren Produkten und Dienstleistungen dabei in die Rechtsordnung eingebunden, genauer gesagt in die Privatrechtsordnung, die allen Bürgern gleichermaßen die Vertragsfreiheit innerhalb der Privatautonomie zusichert. Das bedeutet, die Abwesenheit von Zwang durch vom Staat vorgegebene Ziele für Bürger. Es bedeutet aber auch, dass die autonomen Beziehungen der Privaten untereinander in eine Rechtsordnung eingepasst werden, die ein faires Miteinander ermöglichen soll. Sittenwidrige Verträge, Betrug oder einseitige Ausbeutung werden ausgeschlossen. Es bedarf somit eines Vertrags- und Wettbewerbsrechts, das um Konsumentenschutz und Sektorregulierungen ergänzt wird.

In der Digitalökonomie sind die Skalen- und Netzwerkeffekte jedoch so stark, dass sich in den Vereinigten Staaten Digitalunternehmen herausgebildet haben, bei denen der alte Werkzeugkasten der Regulierung nicht mehr gut funktioniert. Es kann vermutet werden, dass das nicht daran liegt, dass das Wettbewerbsrecht und Spezialregulierungen nicht bereits komplex genug wären, sondern dass ein grundlegenderes Governanceproblem vorliegt, das die Weiterverwendung der traditionellen Wettbewerbsinstrumente einschränkt (Kerber 2023).

Das Governanceproblem liegt darin begründet, dass die Digitalunternehmen das *faktische* residuale Eigentumsrecht über Daten haben. Also nicht nur das Kontroll*recht* über Daten haben, wenn Verträge schweigen, sondern die Kontrolle trotz Verträgen oder sonstigem Recht, wie Wettbewerbsrecht und Sektorregulierungen, besitzen. Die faktische Macht zur Datennutzung liegt an den starken Komplementaritäten (Stieglitz und Heine 2007) zwischen Daten, physischer Infrastruktur und Wissensbasiertheit der Plattform-Geschäftsmodelle. Es macht diese Unternehmen quasi immun gegen den traditionellen Instrumentenkasten von Wettbewerbspolitik und Regulierung. Denn regulative Eingriffe in die Unter-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

nehmen könnten wohl die Ausnutzung von Netzwerkeffekten behindern, würden aber keineswegs automatisch die Wettbewerbsintensität erhöhen oder die Innovationskraft zu neuen digitalen Anwendungen antreiben. Staatliche Eingriffe in die Geschäftsmodelle bleiben steril oder sind sogar kontraproduktiv. Und auch der Besteuerung entkommen diese Geschäftsmodelle leicht, so dass eine Abschöpfung von Gewinnen und Umverteilung anstelle von Regulierung weitgehend ausgeschlossen ist.

Das faktische residuale Kontrollrecht an Daten bei Unternehmen der Datenökonomie stellt die Funktionsfähigkeit einer Wirtschaftsverfassung in Frage, bei der der Staat die „rules of the game" liefern soll, unter der dann die privatautonomen Spielzüge zum gegenseitigen Vorteil gemacht werden (Buchanan und Tullock 1962). Im Gegenteil, die mächtigen Digitalunternehmen stellen dem Staat die Ressourcen und Infrastruktur als Plattform zur Verfügung, auf der er seine Leistungen für die Bürger erbringt. Demokratische Abstimmungen durch die Bürger können den „protective state" (Buchanan 1975) auf der Verfassungsebene so nur noch teilweise bestimmen.[4] Denn das faktische residuale Kontrollrecht über Daten und die damit verbundenen Geschäftsmodelle liegt bei den Unternehmen der Datenökonomie. Die Wirtschaftsordnung mit ihren Verfügungsrechten wird nicht mehr ursächlich von den Bürgern bestimmt, weil diese das faktische Eigentumsrecht an und über Daten verloren haben. Der „protective state" scheint ausgehebelt und es tritt eine ordnungspolitische Inversion ein.

II. China und sein digitales Governance-Modell

Die chinesische Digitalökonomie unterliegt denselben technologischen Bedingungen wie bereits im vorigen Abschnitt beschrieben. Die Governance der chinesischen Digitalökonomie unterscheidet sich jedoch grundsätzlich von der der Vereinigten Staaten. Zwar gibt es formal private Unternehmen in der Digitalwirtschaft, diese werden jedoch maßgeblich durch den Staat gelenkt und instrumentalisiert. Nicht jede unternehmerische Entscheidung muss staatlich autorisiert werden, aber das residuale Kontrollrecht liegt beim Staat, der zum Beispiel über Investitionen oder Fusionen bestimmt (siehe z. B. Ioannidou und Xiong 2025, S. 391 ff.). Vor allem aber hat der

4 Zum grundsätzlichen Zusammenhang von residualen Kontrollrechten, Verfassung und Regulierungswettbewerb siehe Heine (2006).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Staat Zugriff auf die gesammelten Daten. Das kommt oberflächlich besonders zum Ausdruck, wenn der Staat seine Bürger mit Hilfe von Daten einschätzt (social scoring) oder Meinungen auf Sozialen Medien zensiert. Aus wirtschaftlicher Sicht ist bedeutsam, dass der chinesische Staat durch Ausübung seines residualen Kontrollrechts über Daten das Konglomerat von datenintensiven Unternehmen, wie Baidu, Alibaba, Tencent, ByteDance oder Xiaomi steuern und dabei Netzwerkeffekte von Daten effektiv nutzen kann. Anders als im Fall der Vereinigten Staaten gerät der chinesische Staat nicht in eine Abhängigkeit von datenintensiven Unternehmen, sondern bestimmt deren Handlungen. Es sind aber nicht die „rules of the game" die allgemein festgelegt werden und dann im Rahmen der Privatautonomie von Unternehmen ausgefüllt werden, sondern es sind die Spielzüge der Unternehmen selbst, die bestimmt werden. Von einem „protective state", der eine Privatrechtsordnung aufspannt, kann nicht die Rede sein.

Das sich auftuende Governanceproblem muss allerdings noch ein wenig mehr herausgearbeitet werden, um allgemeingültig zu sein. Es könnte argumentiert werden, dass der chinesische Staat nur zufällig den „protective state" verletzt und in die Spielzüge der Privaten eingreift. Eine zukünftige Regierung könnte ihre politische Macht nämlich ohne weiteres dazu gebrauchen, um Eigentum und Vertragsfreiheit zu garantieren und Eingriffe in den (digitalen) Markt auf ein Minimum beschränken. Das könnte die Regierung auch in eine Verfassung schreiben. Dem steht jedoch entgegen, was man als Paradoxon der Macht bezeichnet und das Weingast (1993, S. 287) so beschreibt: „A government strong enough to protect property rights is also strong enough to confiscate the wealth of its citizens." Es bedarf demnach faktischer konstitutioneller Mechanismen, die eine Regierung das Privateigentum schützen lässt und gleichzeitig die Regierung daran hindert, ihre Macht zur Ausbeutung zu gebrauchen. Solche „checks and balances" existieren in China nicht bzw. haben sich bislang nicht herausbilden können.

Das chinesische Governancemodell ermächtigt die Regierung, die Datenökonomie zum Erreichen beliebiger eigener Ziele zu instrumentalisieren. Dabei steht neben dem Gewinnerzielungsziel besonders das Absichern der politischen Macht von etablierten politischen Eliten. Die Datenökonomie liefert hierzu zwei wichtige Bestandteile. Zum einen schafft sie Profite, die mit den Bürgern so geteilt werden können, dass sie am zunehmenden Wohlstand teilhaben. Zum anderen ermöglicht die Datenökonomie den Einsatz von wirksamer Überwachungstechnologie, die gegen die Bürger eingesetzt werden kann.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Man findet in China eine Governance der Datenökonomie vor, in der die Digitalunternehmen zwar dem politischen Willen untergeordnet sind. Allerdings ist der politische Wille nicht an einen Ordnungsrahmen gebunden, der eine Privatrechtsordnung mit gleichen Rechten für alle Bürger aufspannt. Denn die Macht über die Datenökonomie bleibt diskretionär in den Händen einzelner, die das residuale Eigentumsrecht ausüben und den Wettbewerb auf digitalen Märkten bestimmen können.

D. Residuale Kontrollrechte und die Ordnung der Datenökonomie – Herstellung von Wettbewerb durch unkonventionelle Maßnahmen

Die technischen Eigenschaften von Daten (Clubgüter und Skalenerträge) führen scheinbar entweder in eine ordnungspolitische Inversion (Vereinigte Staaten) oder zu politischer Willkür (China). Im Folgenden soll ein Vorschlag gemacht werden, wie mit Hilfe der Zuteilung des residualen Eigentumsrechts an Daten ein Ordnungsrahmen für die Digitalwirtschaft geschaffen wird, der eine funktionsfähige Privatrechtsordnung aufrecht erhält. Interessanterweise ist es ein Rahmen, der sich weder auf den alten (ausgetretenen) wettbewerbsrechtlichen Pfaden des deutschen und europäischen Kartellrechts oder des amerikanischen Antitrust bewegt, noch dem europäischem Aktivismus (DA, DSA, DMA, AI Act usw.) der Sektorregulierung der letzten Jahre folgt.

Ausgangspunkt der Überlegung sind die digitalen Geschäftsmodelle, die quasi als Treibstoff große Mengen an Daten benötigen. Damit die Daten nutzbar für eine Anwendung in einem Geschäftsmodell werden, ist ein Algorithmus nötig, der die Daten erschließt. Sowohl die Beschaffung und technische Bereitstellung der Daten als auch die Entwicklung und Pflege des Algorithmus ist äußerst kapitalintensiv. Es bedarf nicht nur enormer Geldsummen, um die Hardware und Daten zu beschaffen, sondern auch sehr qualifizierter Mitarbeiter, die an dem Algorithmus arbeiten. Zudem stellen sich neben dem Problem der Beschaffung des notwendigen Sach- und Humankapitals grundlegende ethische und Sicherheitsfragen. Wie ist mit Patientendaten umzugehen, was passiert mit den Daten von Kindern auf Sozialen Medien oder welche Datenverknüpfungen zwischen Plattformen sind vereinbar mit dem Anspruch auf Privatheit? Daneben tauchen Fragen der Ausbeutung und Diskriminierung von Nutzern durch marktmächtige Plattformen auf und des möglicherweise gehemmten Innovationswettbewerbs durch etablierte Anbieter.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Es geht um die Organisation eines Marktes, auf dem die Technologie strukturell zu Marktversagen führt und sehr hohe spezifische Investitionen benötigt werden. Zudem ist es ein Markt, der erhebliche Sicherheitsrisiken für das einzelne Individuum birgt als auch für die Bürger in einer freiheitlichen Demokratie insgesamt. Es ist nicht nur ein resistentes natürliches Monopol, das es einzuhegen gilt, sondern es ist auch ein grundsätzliches Ordnungsproblem, wobei die Technologie neben den beschriebenen Risiken enorme Vorteile für den Einzelnen und die Gemeinschaft bereithält.

Dieses Szenario ist nicht einmalig in der Technikgeschichte und es kann daher aus der historischen Betrachtung gelernt werden. Das Szenario, das es zu betrachten gilt, ist die friedliche Nutzung der Kernenergie nach dem zweiten Weltkrieg. Dabei gibt der europäische Fall ein differenziertes Bild, das Parallelen zur Ordnung der Datenökonomie aufscheinen lässt.

Kernenergie hat einige bemerkenswerte Gemeinsamkeiten mit der Datenökonomie. Beide Anwendungen haben ein sehr großes Potential zur Steigerung der allgemeinen Wohlfahrt. Beide Technologien können aber auch großen Schaden anrichten, der sich nicht oder nur schwer kompensieren lässt. Beide Technologien bedürfen Erkenntnisse fortgeschrittener Wissenschaft, die erst durch Kumulation den Einsatz der Technologie ermöglicht. Für die Kernenergie sind das die Erkenntnisse der Atomphysik, erweitert um das Ingenieurwissen für den Bau von Kernkraftwerken. Für die Datenökonomie ist es das theoretische Wissen über Algorithmen (z. B. Neuronale Netze) im Zusammenspiel mit sehr leistungsfähiger Hardware. Beide Technologien haben einen speziellen Betriebsstoff nötig, den sie in ein anderes Produkt oder Service transformieren. Im Falle der Kernenergie sind das Spaltstoffe (z. B. Uran oder Plutonium) und in der Datenökonomie sind es Daten (technische oder persönliche). Für beide Technologien bedarf es zudem großer Kapitalmengen, um die Technologie bereitzustellen. Schließlich müssen in beiden Technologien umfangreiche Wettbewerbs- und Sicherheitsprobleme gelöst werden.

In den 1950er Jahren entfaltete sich eine breite Diskussion in der Bundesrepublik, wie die Probleme im Bereich der friedlichen Nutzung der Kernenergie gelöst werden könnten.[5] Der Bau und Betrieb von Kernkraftwerken erfordert zwar große – vor allem einmalige – Investitionen, aber

5 Einen guten Überblick der kontroversen Diskussion geben die Bundestagsprotokolle ab dem Jahr 1955. Zudem ist die Debatte über die Kernenergie auch in den damaligen Tageszeitungen gut abgebildet.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

private Betreiber können diese Kapitalmengen durchaus aufbringen. Zudem scheidet ein rein staatlicher Bau und Betrieb eigentlich aus, weil zum einen eine privatwirtschaftliche Betriebsführung Effizienzvorteile bietet, aber auch weil andere nichtnukleare privatwirtschaftliche Energieunternehmen parallel im Markt sind. Ein rein staatlicher Kernenergieproduzent wäre ein systemfremdes Element, das Wettbewerbsverzerrungen auf dem Energiesektor erzeugt.[6]

Was in der damaligen Diskussion eine zusätzliche Rolle spielte und heute weitgehend in Vergessenheit geraten ist, sind die antizipierten Wettbewerbsprobleme durch wenige Kernkraftwerksbetreiber, deren Position – einmal eingenommen – von Konkurrenten nicht bestreitbar ist. Es können nicht beliebig viele Kernkraftwerke hinzu- oder abgebaut werden. Diese Restriktion ergibt sich aus dem hohen Kapitalbedarf einerseits und aus den nötigen Sicherheitsmaßnahmen zum Betrieb andererseits.[7]

Die intensiv geführte Diskussion wurde ab 1955 durch einen französischen Vorschlag unter der Federführung von Jean Monnet bereichert, der Deutschlands Ambitionen bezüglich Kernenergie europäisch einbinden und von einer stärkeren Kooperation mit den Vereinigten Staaten und Großbritannien abhalten sollte.[8] Dieser Vorschlag wurde 1957 mit der Europäischen Atomgemeinschaft (EURATOM) durch die Römischen Verträge umgesetzt.

Das aus eigentumsrechtlicher Perspektive besondere an der Ordnung des europäischen Kernenergiemarktes durch EURATOM ist die Zuteilung des residualen Kontrollrechtes an den sogenannten besonderen Spaltstoffen (vor allem für Kraftwerke, Nuklearmedizin und Forschung). Denn in

6 Bereits der EWG-Vertrag von 1957 verlangt, dass öffentliche Unternehmen dem Beihilfeverbot und dem Wettbewerbsgebot unterliegen. Diese Vorschrift wurde freilich von den Gründungsmitgliedern der EWG unterschiedlich interpretiert und ernst genommen. In den Vereinigten Staaten wurden Wettbewerbsverzerrungen mit Blick auf die Kernenergie auch als problematisch gesehen und es kam 1974 zur Gründung der NRC (Nuclear Regulatory Commission), um Verzerrungen auf dem Energiemarkt durch verschiedenen Energieträger zu verhindern.

7 Hier kann nicht die verwickelte politische Diskussion nachgezeichnet werden, wie der Betrieb von Kernkraftwerken in der Bundesrepublik ab den 1950er Jahren rechtlich organisiert werden sollte. Mit dem in Kraft treten des ersten Atomgesetzes und dem darin verankerten Konzessionsmodell war jedoch 1960 ein Endpunkt der Debatte erreicht.

8 Note by Jean Monnet in the French position at the Brussels Conference on nuclear energy (26 July 1955), Fondation Jean Monnet pour l'Europe, [s.l.]. Archives Jean Monnet. Fonds AMK. 38/1/15.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

den Vertragsstaaten von EURATOM ist der öffentliche Eigentümer der besonderen Spaltstoffe EURATOM (Art. 86 des Vertrags zur Gründung der Europäischen Atomgemeinschaft),[9] die eine eigenständige und unabhängige internationale Organisation mit Sitz in Brüssel ist. EURATOM überlässt den Betreibern von Kernkraftwerken nukleare Brennstoffe auf unbestimmte Dauer für die kommerzielle Nutzung. EURATOM definiert zudem die Sicherheitsstandards für Spaltstoffe. Das hat zur Konsequenz, dass EURATOM das Nutzungsrecht den Kernkraftwerksbetreibern entziehen kann, wenn die Situation es erfordert. Es bedeutet freilich nicht, dass EURATOM das Nutzungsrecht dann selbst ausübt, es könnte jedoch auf einen anderen Betreiber übertragen werden. Genauso kontrolliert EURATOM sein Eigentumsrecht, wenn ein Kernkraftwerksbetreiber das Nutzungsrecht auf einen anderen übertragen will. Das Eigentumsrecht an Spaltstoffen ist damit ein Eigentumsrecht besonderer Natur, weil es zwar das formale Verfügungsrecht bestimmt, aber nicht unmittelbar den Besitz und die wirtschaftliche Nutzung (Schnorr 1961; Böhm 1962; Herpers 1967).

Die Gesichtspunkte, unter denen EURATOM die Ausübung des Nutzungsrechts betrachtet, sind in erster Linie an der Sicherheit der Kernkraft orientiert. Wenn also der Kraftwerksbetreiber seine Garantie zur Sicherheit nicht hält oder halten kann. Damit verwoben sind wettbewerbspolitische Aspekte. Denn das Ausnutzen einer marktmächtigen Stellung eines Kernkraftwerksbetreibers stößt an seine Grenzen, wenn EURATOM das Nutzungsrecht entzieht, eine Übertragung nicht zulässt oder einem Kraftwerksbetreiber die Zuteilung weiterer Spaltstoffe verweigert. Offensichtlich kann eine sehr wirkungsvolle und vor allem zeitnahe Wettbewerbspolitik mittels residualer Kontrollrechte betrieben werden. Schon die Möglichkeit der Ausübung des residualen Kontrollrechts hat eine disziplinierende Wirkung auf das Marktverhalten.

Die Ausübung des residualen Kontrollrechts zur Durchsetzung von Sicherheits- und wettbewerbspolitischen Zielen ist jedoch nur dadurch möglich, dass es sich um einen einzigen, klar definierten Stoff – nämlich Spaltstoffe – handelt. Dadurch wird zwar auf der einen Seite der allgemeine Einsatz von residualen Eigentumsrechten zum Erreichen von wettbewerbs- und sicherheitspolitischen Zielen durch den Staat stark eingeschränkt, auf der anderen Seite wird ihr möglicher Einsatz bei abgrenz-

9 In den Vereinigten Staaten und China finden sich ebenfalls Modelle, die Spaltstoffe in öffentlichem Eigentum halten.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

baren Inputgütern, die mit einer spezifischen Technologie kombiniert werden, aber umso vielversprechender.

Daten sind, wie Spaltstoffe, ein gut abgrenzbares Gut. Es gibt zwar verschiedene Datenkategorien (genauso wie es unterschiedliche Spaltstoffe gibt), die von allgemein technisch bis zu hochsensibel und personalisiert reichen. Zudem ist es eine Frage, mit welchem Algorithmus und zu welchem Zweck Daten genutzt werden. Aber was Daten sind und wo und wie sie gesammelt werden ist eindeutig. Daher scheint die Datenökonomie ein Fall zu sein, in dem der Markt mit Hilfe von Überlegungen aus der Theorie residualer Kontrollrechte geordnet werden kann. Die sich ergebende digitale Ordnung vermeidet sowohl eine ordnungspolitische Inversion wie in den Vereinigten Staaten als auch eine willkürliche Machtstellung der Regierung wie in China.

Die Ordnung der Datenökonomie anhand von residualen Kontrollrechten lässt sich analog zum europäischen Markt für Spaltstoffe konstruieren. Der Eigentümer aller Daten ist die öffentliche Hand, die durch eine unabhängige Institution (Datenamt) vertreten wird. Eine solche Institution erhält den selben rechtlichen Status wie EURATOM oder die Europäische Zentralbank als internationale Organisation. Damit ist sie dem direkten politischen Einfluss und der Willkür einer Regierung entzogen.

Daten werden privaten Unternehmen zur dauerhaften wirtschaftlichen Nutzung überlassen. Genauer gesagt, es wird Unternehmen erlaubt, Daten für die kommerzielle Nutzung zu erheben, zu speichern, auszuwerten oder zu übertragen. Unternehmen der Datenökonomie „schürfen" Daten, die „aufbereitet" werden und dann quasi „verbraucht" werden. Diese drei Schritte können integriert in einem Unternehmen stattfinden oder getrennt voneinander. Wichtig ist, dass diese Schritte unterschieden werden und je nach Art der Daten können auf jeder Stufe Sicherheitsregeln, zum Beispiel zum Schutz der Privatsphäre, erlassen werden. Unternehmen, die den jeweiligen Sicherheitsstandard nicht belegen können, wird das Nutzungsrecht an den Daten untersagt bzw. das Datenamt übt sein Eigentumsrecht aus. Genauso können Wettbewerbsprobleme angegangen werden. Das Recht zur Datennutzung kann ganz oder teilweise (auf den drei Stufen) entzogen werden, oder die weitere Datennutzung wird vom Teilen der Daten mit anderen Unternehmen abhängig gemacht.[10] Mit anderen Worten, die Organisation des Wettbewerbs der

10 Hier ergeben sich beispielsweise interessante neue Möglichkeiten zur Gewährung von Marktzugang. Während sich in der Bronner-Rechtsprechung (C7/97) das Ei-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Datenökonomie erfolgt nicht über das allgemeine Wettbewerbsrecht und flankierenden Sonderregulierungen, sondern über ein unabhängiges Datenamt, das das residuale Eigentumsrecht über Daten ausübt. Eine regulative Zersplitterung, wie sie derzeit mit nationalen und europäischen Wettbewerbsrechten und parallelen Sektorregulierungen (DSGV, DSA, AI Act usw.) existiert, wird zurückgedrängt. Ebenso werden lange Gerichtsprozesse vermieden, die ex post nur wenig an einer Verminderung des Wettbewerbs oder des Datenschutzes ändern können. Grundsätzlich vermeidet ein eigentumsrechtlicher Ansatz regulative Dysfunktionalitäten, die sich daraus ergeben, dass Wettbewerbsrecht keinen direkten Zugriff auf die Daten erlaubt, mit denen Unternehmen ihre Geschäftsmodelle betreiben.

E. Schlussbetrachtung

In diesem Beitrag konnte die Idee der residualen Eigentumsrechte zur Herstellung einer freiheitlichen Ordnung der Datenökonomie nur angerissen werden. Es sollte jedoch klar geworden sein, dass Ordnungsentwürfe, die die Privatautonomie in der Datenökonomie herstellen, durchaus möglich sind. Auch wenn diese auf den ersten Blick unkonventionell klingen und nicht in die hergebrachten wirtschafts- und rechtspolitischen Schemata passen. Bei dem hier vorgestellten Entwurf geht es nicht horizontal um die Allokation von Eigentumsrechten zwischen Unternehmen, sondern vertikal zwischen Unternehmen und dem Staat. Damit können nicht nur Wettbewerbsprobleme auf digitalen Märkten gelöst werden, sondern auch effektiver Datenschutz betrieben werden.

---

gentumsrecht an tangiblen und intangiblen Gütern als Bollwerk zum Zugang von Wesentlichen Einrichtungen erweist, ist dies hier nicht der Fall, denn das Eigentum liegt formal beim Datenamt. Es wäre auch zu vermuten, dass das Ergebnis des jüngsten Gerichtsurteils zu Google vom 2. 9. 2025, in dem Google zwar den Chrome Browser nicht abspalten, aber einen Teil der Daten mit anderen Unternehmen teilen muss, zumindest sehr viel schneller erzielt worden wäre. Der Fall war bereits seit 2020 beim United District Court for the District of Columbia anhängig.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Literatur

Alchian, A.A. (1950): Uncertainty, Evolution and Economic Theory, in: Journal of Political Economy, Vol. 58, S. 211–221.

Böhm, P. (1962): Ownership of Nuclear Materials in Euratom, in: American Journal of Comparative Law, Vol. 11, S. 167–183.

Buchanan, J.M. (1965): An Economic Theory of Clubs, in: Economica, Vol. 32, S. 1–14.

Buchanan, J.M. (1975): The Limits of Liberty: Between Anarchy and Leviathan, Chicago.

Buchanan, J.M.; Tullock, G. (1962): The Calculus of Consent: Logical Foundations of Constitutional Democracy, Ann Arbor.

Calabresi, G.; Melamed, A.D. (1972): Property Rules, Liability Rules, and Inalienability: One View of the Cathedral, in: Harvard Law Review, Vol. 85, S. 1089–1128.

Coase, R.H. (1960): The Problem of Social Cost, in: Journal of Law & Economics, Vol. 3, S. 1–44.

Cooter, R.; Ulen, Th. (2008): Law & Economics, 5. Aufl., Boston.

Ellickson, R. (1991): Order without Law, Cambridge.

Furubotn, E.G.; Pejovich, S. (1972): Property Rights and Economic Theory: A Survey of Recent Literature, in: Journal of Economic Literature, Vol. 10, S. 1137–1162.

Global Times (2025): China's new regulations on government data sharing to boost service efficiency, 4. Juni.

Greenstein, S. (2021): Digital Infrastructure, in: Glaeser, E.L; Poterba, J.M. (Hrsg.), Economic Analysis and Infrastructure Investment, Chicago, S. 409–447.

Hardin, G. (1968): The Tragedy of the Commons, in: Science, Vol. 162, S. 1243–1248.

Hart, O. (2017): Incomplete Contracts and Control, in: American Economic Review, Vol. 107, S. 1731–1752.

Heine, K. (2006): Interjurisdictional Competition and the Allocation of Constitutional Rights: a Research Note, in: International Review of Law and Economics, Vol. 26, 2006, S. 33–41.

Heine, K. (2010): Comment on Barbara Krug and Hans Hendrischke – Market Design in China, in: Theurl, Th. (Hrsg.), Institutionelle Hintergründe und Konsequenzen von Wissen, Schriften des Vereins für Socialpolitik, Berlin, S. 171–178.

Herpers, H.H. (1967): Artikel 86 des Europäischen Atomgemeinschafts-Vertrages (EAGV) und Artikel 14, Absatz 3 des Grundgesetzes (GG), in: Zeitschrift für die gesamte Staatswissenschaft (JITE), Vol. 123, S. 339–346.

Ioannidou, M.; Xiong, B. (2025): Competition Law and Policy in Digital Markets, Oxford.

Kerber, W. (2023): Digital Revolution, Institutional Coevolution, and Legal Innovations, in: European Business Law Review, Vol. 34, S. 993–1016.

Leipold, H. (1988): Wirtschafts- und Gesellschaftssysteme im Vergleich, 5. Aufl., Stuttgart.

Ostrom, E. (1990): Governing the Commons, Cambridge.

Schnorr, G. (1961): Die Eigentumsordnung im Euratom-Vertrag, in: Wirtschaftsdienst, Vol. 41, S. 124–129.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Stieglitz, N.; Heine, K. (2007): Innovations and the Role of Complementarities in a Strategic Theory of the Firm, in: Strategic Management Journal, Vol. 28, S. 1–15.
Weingast, B.R. (1993): Constitutions as Governance Structures: The Political Foundations of Secure Markets, in: Journal of Institutional and Theoretical Economics (JITE), Vol. 149, S. 286–311.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Verantwortung in Politik und Beratung

Armin Steinbach

Im Eichensaal des Bundeswirtschaftsministeriums wurden einst Tanzstunden für die Militärärzte ausgerichtet und in den stuckgerahmten Emporen kann man sich mit etwas Fantasie noch heute eine Tanzkapelle vorstellen, die den Takt vorgibt, um den Ärzten gesellschaftlichen Schliff zu verleihen. Empfänge werden im Eichensaal noch heute gegeben. Tanz und Kammermusik sind inzwischen anderen Festlichkeiten gewichen. Zu besonderen Anlässen werden hier etwa die Mitglieder des wissenschaftlichen Beirats des Hauses bewirtet und unterhalten, jenem Gremium also, das der politischen Führung des Hauses seit 1948 auf die Finger schaut. Christoph Engel gehört ihm seit 1997 an. Er hat schon vielen solcher Anlässe beigewohnt, viele Minister im Eichensaal das Wort ergreifen hören. Um die Anwesenden bei solchen Veranstaltungen soll es hier gehen: Minister, die gewählt ins Amt kommen und die eine kleine Gruppe politischer Beamter auf exponierter Position berufen; Fachbeamte, die das Herz jedes Ministeriums und die unabhängig von der personellen Dynamik an der Hausspitze das stabile Fundament bilden; und Wissenschaftler, die über Beiräte und andere Gremien Politikberatung anbieten. Politische Beamte und Fachbeamte tragen in unterschiedlicher, aber komplementärer Handlungslogik zur Staatslenkung bei. In der Wissensverarbeitung bleiben sie (zumindest teilweise) angewiesen auf externe Wissensproduzenten. Der Beitrag fragt, wie sich die Verantwortungssphären der Akteure in ihrer Zurechnungs- und Sanktionsmechanik unterscheiden.

A. Vertrauenslegitimierte und Leistungslegitimierte in der Staatsführung

Das Leitbild bundesrepublikanischer Staatslenkung trägt demokratische wie meritokratische Züge. Einerseits zehrt sie von Legitimationsstiftung durch das Volk, das dem Politiker *Vertrauen* und ein freies Mandat schenkt und an dieses Recht und Pflicht knüpft, alle wesentlichen Entscheidungen der res publica im Parlament zu entscheiden – als Vertrauenslegitimierte, an die als einzige Eignungsvoraussetzung gestellt wird,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dass sie sich der Verfassung gegenüber loyal zeigen. Andererseits liegt Staatslenkung in der Hand der Regierung, genauer: eines zweigliedrigen Ministerialapparates, Kopf und Rumpf, d. h. eines Vertrauenslegitimierten an der Spitze eines voluminösen, arbeitsteilig und hierarchisch konstruierten Herrschaftsinstruments. Die Angehörigen dieses Rumpfes sind Leistungslegitimierte, die aufgrund der Selektionstrias Leistung, Eignung, Befähigung ins Amt kommen. Aus der hierarchischen Unterordnung der Leistungslegitimation gegenüber der Vertrauenslegitimation folgt aber keine verwaltungsübliche reine Exekutivfunktion der Fachbeamten gegenüber ihrer Hausleitung.[1] Damit tragen Parlamentarier und Ministerialspitze zwar das wahllegitimierte Mandat. Allerdings wird dem Vertrauenslegitimierten das Korsett des Leistungslegitimierten angelegt, welches politischen Impuls in die Bahnen des rechtlich, technisch, fachpolitisch Machbaren und Vernünftigen lenken soll.

Damit findet das Staatshandeln eben nicht allein in dem wahlgebundenen Vertrauenstransfer seinen Motor, sondern auch in Teilen eines meritokratischen Fachbeamtentums. Staatsleitung ist so verfassungsrechtlich auf zwei Schultern verteilt: die eine ist maßgeblich dem Volkswillen zugewandt, nimmt Impulse auf, stiftet Akzeptanz durch Versprechungen und dem Entwurf von Zukunftsbildern, hört zu und baut die Brücke zwischen den Sehnsüchten des Wahlvolks und dem Handeln des Staates. Die andere ruht auf dem Vertrauen in eine unpolitische Rationalität, die sich nicht nur in basalen Eigenschaften des Verwaltungshandelns erschöpft wie etwa der Rechtmäßigkeit des Staatshandelns, der Unbestechlichkeit und Neutralität der Staatsdiener. Sondern auch in dem Glauben an technische, wirkungsorientierte, wissenschaftsbasierte, fachspezifische Maßstäbe, die außerhalb des Politisch-Disponiblen liegen oder mit diesem konkurrieren und in einen Ausgleich zu bringen sind. Der Beamtenstaat ist damit auch ein Produkt der Aufklärung und des industriellen Fortschritts, der als treibende Feder nicht allein einer Wertrationalität verschrieben ist (grob verkörpert durch den Politiker), sondern in der Zweckrationalität den Bestand technischer und wissenschaftlicher Errungenschaften auf eine effiziente und effektive Professionalisierung der Staatsleitung überträgt.

Während die Ministeriumsleitung auf Reputations- und Profilierungsrenditen aus ist, nimmt der übrige Ministerialapparat nicht am par-

1 *Wißgott*, Parlamentarische Demokratie und Deutsches Berufsbeamtentum, Der Staat (2025), 499, 431.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

teilichen Wettstreit teil; er ist an der Macht und bleibt es auch, er muss sich darin nicht behaupten. Die Ministerialbürokratie ist seit der Einführung des preußischen Beamtentums im verfassungsrechtlichen Fundament einzementiert. In ihrem historischen Wuchs ging sie der Verfassungsgebung voraus, wurde vom Verfassungsgeber nahtlos übernommen; anders als das US-amerikanische System, in dem mehrere Schichten der Ministerialapparate bei einem politischen Wechsel ausgetauscht werden, überlebt die Ministerialbürokratie auch Wahlen in vergleichsweise großer Stabilität.[2] Hierzulande ist das unparteiliche Fundament breiter und entzieht dem System steuerbare Agilität. Aus dieser Distanz zum Politischen gehorcht der ministeriale Apparat einer am Sachzwang orientierten Eigenlogik, die den politischen Kopf an die bürokratische Handlungsmechanik stärker bindet als dass umgekehrt dieser seinen Apparat zu navigieren vermag.[3] Aus der hierarchischen Unterordnung der Leistungslegitimation hinter der Vertrauenslegitimation folgt nur scheinbar eine bloße Vollzugsrolle der Beamtenschaft, schon gar nicht eine unhinterfragte Gefolgschaft hinter den Vertrauenslegitimierten.[4] Als Anbieter und Herrscher über Wissen können aus Gelenkten Lenker werden, wie die BBC-Fernsehserie „Yes, Minister" anschaulich in Szene setzte. Der höchste Beamte, Sir Humphrey Alleby, hat schon manchen Minister kommen und gehen sehen und ihn auf Linie des Hauses gebracht, um ihn aus Sicht des Hauses „stubenrein" zu machen.[5] Gewiss: auch der restringierende Beamtenkorpus muss sich bewegen, je nachdem welcher Farbe die Hausspitze gerade angehörig ist, so wollen es Hierarchieprinzip und Weisungsrecht; er muss sich lenken lassen, jedoch nicht ohne der sich regelmäßig wechselnden Hausleitung erhebliche Retardierungs- oder gar Gegenkräfte entgegensetzen zu können.[6] Dieser in Unabhängigkeit, Neutralität und Amtsstabilität gebettete Korpus bewahrt Ruhe, wenn sein Haupt sich in alle Richtungen dreht. Sein Code basiert auf Leistung, Eignung, Befähigung, nicht auf Überzeugung, Gesinnung und Mehrheitsfähigkeit.

2 *Roßbach*, Das Personal der Republik, 2020.

3 *Luhmann*, Der neue Chef, 2016, 97.

4 *Luhmann*, Die Grenzen der Verwaltung, 2021, 146ff.

5 *Mertes*, Zyklen der Macht, 2021, 165f.

6 Zur Untauglichkeit klassischer Verwaltungsmodelle, die von einer befehlsbasierten Herrschaft durch das Hierarchieprinzip ausgeht *Luhmann*, Die Grenzen der Verwaltung, 2021, 146ff.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Diese Doppelstruktur der Staatsleitung als einer sich ergänzenden Polhaftigkeit zwischen statischer Rationalität und dynamischer Zukunftsgestaltung war lange Zeit alles andere als selbstverständlich. Bis zum Ersten Weltkrieg prägte die ministeriale Beamtenschaft ein hohes Maß an Homogenität. Sozialisten, Juden, Demokraten, Katholiken – ihnen wurde der Eintritt erschwert, statt ihrer war die Zugehörigkeit zur Oberschicht maßgebliches Eintritts- und Aufstiegsticket. Unter den Regierungsreferendaren waren 1913 noch 55% adlig.[7] Eine erste Demokratisierungswelle, die mit der homogenen Linie von Monarchie und Beamtentum brechen sollte, musste auf der republikanischen Gesinnung ihrer Beamten aufbauen. Diese Zäsur warf – wie auch später in der Erneuerungsphase nach Überwindung der Nazi-Diktatur – die Frage nach der Personalrekrutierung auf. Die eingeschliffenen Beamtenstrukturen beförderten ein hohes Maß an gleichförmiger, aber doch übungsintensiver Verwaltungstätigkeit und wenngleich die Gesinnung eine Zäsur geboten hätte, musste mit einem Qualitätsabfall in Folge einer personellen Tabula Rasa gerechnet werden. Erst mit der eingeleiteten Demokratisierung, forciert durch das Leistungskriterium als maßgebliches Kriterium und bei gleichzeitiger Irrelevanz anderer diskriminierender Merkmale wie der politischen Gesinnung oder des Geschlechts, konnten breitere Schichten in den Beamtendienst integriert, der soziologische Rekrutierungshintergrund und damit auch die politische Orientierung pluralisiert werden. In dem Umfang, in dem mit parteipolitischer und soziologischer Homogenität zwischen Beamten und Politikern gebrochen wurde, konnten sich legislative und ministerialbürokratische Staatsleitung entlang von Vertrauen und Leistung ausdifferenzieren.

Der Frage nach der Personalrekrutierung vorgelagert ist das normative Problem, welche Anteile Vertrauenslegitimierte und Leistungslegitimierte an der Staatsleitung innehaben sollen. Der verfassungsrechtliche und politikwissenschaftliche Mainstream besagt zunächst: wesentliche Entscheidungen der Gestaltung der res publica, insbesondere dort, wo Grundrechte betroffen sind, brauchen ein letztverbindliches Plazet des Parlaments. Damit ist aber nur eine Aussage über die Letztverantwortlichkeit, über das Parlament als höchste Instanz gestalterischen Wandels getroffen. Nichts besagt es zum Impuls gestalterischen Handelns. Wie weit kann ein Minister dem Wunsch seiner Bundestagsfraktion nach einem

7 *Pickart*, Vierteljahreshefte für Zeitgeschichte (1958).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

politisch gewollten Gesetz entsprechen? Schon die Dualität zwischen Amt und Amtsträger macht deutlich, dass der Minister nicht der verlängerte Arm einer Parlamentsmehrheit sein kann. Person und Amt sind getrennt, im Amt tritt die Person zurück. Sie verinnerlicht die Funktionalität des ihr nur auf Zeit untergebenen Steuerungsapparats.

Als Person will der Vertrauenslegitimierte an der Ministeriumsspitze das politische Programm seiner Fraktion und Partei umsetzen; als Amtsinhaber ist er durch das restringiert, was der freien politischen Gestaltbarkeit entgegengesetzt ist. Die Restriktion ist in ihren Dimensionen unterschiedlich eindeutig sichtbar: die Prüfung der Verfassungsmäßigkeit von einfachen Gesetzgebungsinitiativen durchläuft die Mühlen der Verfassungsressorts und ausgeflaggte Inkompatibilititätssorgen können je nach Höhe des jeweiligen Kassationsrisikos auch für die politische Leitung heikel werden, sollte sie sich darüber hinwegsetzen. Weniger sichtbar, doch ständig virulent ist die fachliche Kontrolle. Ob eine Maßnahme geeignet ist ein Ziel zu erreichen? Ist sie wirksam und zu welchen Kosten und mit welchen Nebenwirkungen? Welches alternative – bestehende oder zukünftige – Instrumentarium steht zur Verfügung? Wie korrespondieren Maßnahmen mit solchen anderer Politikbereiche, denen der Nachbarländer oder mit europäischen Politikvorgaben? Aus allen Quellen fachdisziplinärer, technischer, rechtlicher Erwägungen kann das Fachbeamtentum Argumentationsmaterial in die eine oder andere Richtung verdichten. Je zwingender es ist, je geschickter die Fachebene die Entscheidungsaufbereitung gestaltet, je überzeugender das Votum und der Sachstand einer Ministervorlage sind, die später aktenkundig die Auffassung der Fachebene dokumentiert, umso mehr muss der entgegengesetzte politische Impuls von Hierarchie und Weisung Gebrauch machen, um das Delta zwischen Leitung und Fachebene zu überbrücken. Mehr noch: sollte die Friktion zwischen politischer und fachlicher Ebene öffentlich werden, deckt die mediale Öffentlichkeit sie schonungslos auf und zwingt den Minister in die Defensive, weil er sich über die interne Expertise „seiner eigenen Leute" hinweggesetzt hat. Schlimmstenfalls verdichtet sich der Rechtfertigungszwang zu einem gerichtlichen Verfahren. Die Bremsklötze des Fachbeamtentums zum gestalterischen Drive des Vertrauenslegitimierten sind vielfältig, verfassungsrechtlich gewollt und ein Gegengewicht zum entwicklungsoffenen gesellschaftlichen Wandel, der sich im Diskurs des politischen Spektrums bilden soll.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

B. Verantwortungsdelta in Politik und Wissenschaft

In die ministeriumsinternen Mühlen geraten gelegentlich auch Impulse von außen. Der externe Wissenschaftler freut sich (meistens), wenn ihn Lockrufe der praktischen Welt erreichen und seiner Expertise schmeicheln. Dann geht es um die Bereitstellung von Wissen in Ergänzung zu ministeriumsintern produziertem Wissen. Im folgenden soll es darum gehen, dass innerhalb der Staatsgewalten produziertes Wissen und extern-wissenschaftlich bereitgestellte Expertise unterschiedlichen Zurechnungs-, Sanktions- und Verantwortungsmechanismen unterworfen sind. Als Expertise formuliertes Wissen kann wahr und falsch sein, darauf bauende Politik kann die intendierte Wirkung verfehlen oder sie kann schlicht dem gesellschaftlichen Gegenentwurf der Opposition widersprechen und wird von dieser öffentlich angegangen.

I. Ministeriumsinterne Verantwortungszuschreibung

Politische Verantwortung ist pyramidal. In ihr gipfelt bürokratisches Hierarchieprinzip, sie umfasst organisatorisches Delegationsprinzip und sie abstrahiert von eigener Handlung als verantwortungsbegründendes Momentum.[8] Organisatorisch strebt sie der Spitze zu, findet dort ihren verbindlichen Bezugspunkt für die Interaktion mit der Außenwelt. Alles unterhalb dieser Spitze steht in ihrem Dienst. Die Dichotomie zwischen Dienst und Dienstherrn im politischen Raum nimmt eine Divergenz in Kauf – zwischen der dezentral spezialisierten Kompetenz einerseits und dem zentralen Zurechnungsreservat andererseits. Expertise und Entscheidung werden ebenso getrennt wie Wissen und Wollen, Rat und Tat. Ein Minister versteht und weiß in der Regel nur wenig von dem, worüber er zu entscheiden hat und was ihm zugerechnet wird. Er verlässt sich auf das Wissen seines Apparates, der seinerseits seine Expertise bereitstellt, ohne dafür Autorenschaft in Anspruch zu nehmen. Der einzelne Ministerialbeamte offeriert das bestmögliche Produkt als Profilierungsgegenstand für Minister und Regierung, lehnt aber selbst jede Verantwortung ab. Er versteht sich als unselbständiges Element eines bürokratischen Korpus. Passiv-Formulierungen wie „Es wird vorgeschlagen, dass […]“ sind ebenso eine Art von Verantwortungsverschleierung wie breite Mitzeich-

8 *Nida-Rümelin*, Verantwortung, 2011, 150.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

nungsleisten bei einer Ministervorlage.[9] Der Beamte wird rechtlich durch restriktive Amtshaftung abgeschirmt; für Außenstehende ist er hinter Organigrammen und undurchsichtigen Zuständigkeiten nicht erkennbar. Damit wird nicht der dauerstabile Torso des Apparates, sondern nur der volatile Kopf des Apparates zur Rechenschaft gezogen.

Daraus entsteht die Metrik politischer Verantwortung – sie ist keine Handlungsverantwortung, die sich aus einem konkreten Handeln des Ministers ableitet. Sie birgt nicht einmal ein ethisches Unwerturteil über den Träger politischer Verantwortung. Auch Organisationsversagen trifft nicht den Kern politischer Verantwortung. Sie erschließt sich vielmehr aus der Eigenart eines entindividualisierten Berufsbeamtentums – dieses beherrscht das Dossier inhaltlich, stiftet Sinn und Kompetenz als unpersönlicher, anonymer Korpus. Ministeriale Gepflogenheiten verkörpern dies: die Ministervorlage transportiert Expertise und Entscheidungsvorschlag von unten nach oben, rückabwickelt das Delegationsprinzip und schafft Distanz zwischen Wissen und Entscheidung. Sie sichert fachliche Fundierung, indem sie die verästelten Wege des Organigramms von unten nach oben durchläuft. Der Weg aufwärts der Ministervorlage ist nötig, weil dezentral keine Verantwortung übernommen werden soll für das politische Handeln. Der Korpus wird hermetisch abgeschirmt von der Verantwortungsübernahme, seine Funktion besteht in der rein internen Bereitstellung von Handlungswissen. Der einzelne Beamte verschwindet hinter dem Gesetz, das er selbst geschrieben hat. Tritt er nach außen doch einmal in Erscheinung, dann eigentlich nur „i.A." – der nach außen kommunizierende Ministerialbeamte handelt „im Auftrag". Während der vertrauensernannte Staatssekretär noch „i.V.", also in Vertretung des Ministers handelt, begnügt sich das Personal unterhalb dieser Ebene damit, seine Unselbständigkeit als Auftragnehmer und Entscheidungsüberbringer zu betonen. Obwohl der Fachbeamte mit seiner Sachkunde der Entscheidung sachliche Legitimation vermittelt, bleibt er im Außenauftritt unsichtbar.[10] Damit läuft sämtliche Verantwor-

9 *Mertes*, Zyklen der Macht, 2021, 35.

10 Wir blenden hier einmal aus, welche psychologischen Implikationen das für einen Träger von Leistung und Befähigung hat, der sich sein Leistungsvermögen bis zu seinem Eintritt in den Ministerialdienst stets selbst zuschreiben durfte, womöglich in seiner Ausbildung und Vorkarriere durch Exzellenz bestach und sich jetzt nurmehr als schlichter Bote auszugeben hat. Besonders schmerzhaft (und nur durch die beamtenrechtlich zwingende Loyalitätspflicht zu rechtfertigen) sind die Fälle, in denen sein Entscheidungsvorschlag im politischen Entscheidungsprozess zur politisch-opportunen, aber sachlich-kruden Dezision mutiert.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tung auf die Spitze zu. Dort wird vollständiges Handlungswissen fingiert, um die Zurechnung der Verantwortung zu ermöglichen.

Wozu diese Fiktion und warum Entpersonalisierung von Wissen? Ohne sie wäre vollständige demokratische Kontrolle nicht möglich. Um Regierungshandeln kontrollierbar zu halten, muss es sich im politischen Raum rechtfertigen und dem Parlament gegenüber personell sanktionierbar sein. Öffentliche Kontrolle kann aber nur im selben Modus erfolgen, auf der politische Macht überhaupt erst zur Entstehung gekommen ist. Der durch Wahlen vermittelte Legitimationsgrund ist vertrauensbasiert, nicht leistungsbasiert. Parteien genießen das Vertrauen der Wähler; ein Kanzler das Vertrauen der Parlamentsmehrheit, seine Ministerriege das Vertrauen des Kanzlers; die politischen Beamten das Vertrauen des Ministers. Leistungsversagen des Fachbeamten muss sich der Minister also zurechnen lassen und führt zu Vertrauensentzug, obwohl kein eigenes Versagen ursächlich ist für Fehlleistung des Fachbeamtentums.[11] Die Sanktionen im Innenverhältnis (Dienst- und Disziplinarrecht) sind irrelevant für die Sanktion des Ministers im Außenverhältnis (Vertrauensentzug). Die unterschiedlichen Verantwortungssphären wirken in der politischen Verantwortung zum Nachteil des Ministers.

II. Verantwortung des Wissenschaftlers

Wenden wir uns als Kontrast dem Wissenschaftler zu, der wie der Ministerialbeamte Expertise als Entscheidungsgrundlage für politisches Handeln anbietet. Der Fachbeamte speist es direkt-intravenös im politischen Entscheidungskanal ein, der andere vermittelt es über Öffentlichkeit oder politiknahe Entscheidungsgremien, von wo es vom Ministerialbetrieb aufgegriffen wird. Die Art ihres Wissensangebots unterscheidet sich nicht grundlegend: das Angebot der Wissenschaftler ist professionstypisch erkenntnisgetrieben. Auch der leistungslegitimierte Ministerialbeamte pflegt ein Arbeitsethos, der durch den Modus der Sachgesetzlichkeit geprägt ist, wobei sich an die Stelle des Diktats strenger Wissenschaftlichkeit mit Blick auf das letzte Nadelöhr politischer Implementierung – Op-

11 Inhaltliche Qualitätskontrolle sichert im Grundsatz das ministeriumsinterne Abstimmungsgebot unterschiedlicher Abteilungen sowie das Gebot der Ressortabstimmung, aus dem ein Kabinettsbeschluss als gemeinschaftliche Entscheidung hervorgeht.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

portunität und Mehrheitsfähigkeit – eine Portion strategische Ingredienz beimischt. Sind externer und interner Experte inhaltsgeleitet, verlässt sich doch der eine auf eine Delegation seiner Verantwortung auf die politische Ebene (wie oben erläutert), während der andere dies nicht tut oder tun kann. Mit Folgen: der Wissenschaftler gerät an zwei Fronten ins Kreuzfeuer der Kritik, wenn er fehlbares Wissen transportiert. Wissenschaftsintern muss er sich im disziplinspezifischen Umfeld mit den Pflichten und Obliegenheiten guter wissenschaftlicher Praxis behaupten. Wissenschaftlicher Diskurs und das Urteil der *peers* schweben als Damoklesschwert über dem Ruf des Wissenschaftlers. Im Wissenschaftsideal entscheidet der Erkenntnisbeitrag über das Ansehen der Person. Die wissenschaftsinterne Review-Mechanik lockert sich jedoch im externen Verhältnis wissenschaftsinteressierter Öffentlichkeit. Zugespitzte Formulierung politscher Handlungsempfehlungen, die sich der wissenschaftstypischen Grautöne und Annahmekontingenz entledigen, stößt dort nur auf dünne Kontrolle. Herkömmliche Medien funktionieren nicht nach Rede und Gegenrede, eher schon rührt sich auf sozialen Medien der Widerspruch auf den kühnen Vortrag. Vor allem aber löst sich beim Wissenschaftler nicht der Autor von seinen Inhalten, bleibt die Herkunft des Ratschlags bei ihrem Vortragenden und diffundiert nicht in einem entpersonalisierten Korpus. Der Wissenschaftler ist in höchstem Maße individualisiert, seine ganz Karriere baut auf Eigenleistung und Selbstprofilierung auf. Zitierpflichten und geistiges Eigentum sind dort sanktionsbewehrte Absicherung. Wissenschaftliche Verantwortung kennt kein „i.A.".

Im Ministerium ankert Urheberschaft zwar noch in einem initialen ‚Ich-Vorschlag', sobald er die Ebene des Referats verlässt, wird er zu einem ‚Wir-Vorschlag', der dann zu einem abteilungsübergreifenden ‚Haus-Vorschlag' wird, der auf den Minister zuläuft, bei dem es schließlich mit einem grünen Haken zu dessen ‚Ich-Vorschlag' im politischen Raum und damit zu einem Profilierungsprodukt wird – geistige Produkte wechseln hier mehrfach ihren Urheber. Während in der Wissenschaft das Subjekt im Vordergrund steht und die Expertise eindeutig zurechenbar ist, funktioniert der Regierungsapparat nach der exklusiven Subjektivierung auf die Ministeriumsspitze hin. Anders als der Beitrag des Wissenschaftlers ist jener des ministerialen Ursprungs verborgen vor den Augen der Öffentlichkeit, kann im Innern Lob und Tadel hervorrufen und zur Grundlage für Beförderungen werden. Aber nur mit großem Unbehagen kann der federführende Referent von „meinem Gesetz" sprechen, auch wenn Idee und Umsetzung vollständig aus seiner Feder stammen. Glänzen darf nur

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

der oder die eine an der Spitze. Statt des Gleichlaufs aus Urheberschaft und Verantwortung wird ein Vakuum geschaffen, welches durch die Fiktion der politischen Verantwortung gefüllt wird. Das gute oder schlechte Gesetz, das falsche oder richtige Instrument – immer bezieht sich der öffentliche Raum, also Medien und parlamentarische Kontrolle, auf die Individualperformanz des Ministers. Das soll nicht heißen, dass dieser keinen individuellen Beitrag leisten kann. Er steuert, ändert, setzt Impulse, allein schon, um sich und seine Partei erkennbar werden zu lassen.

C. Fazit

In der Wissensproduktion der Staatslenkung steht der neutrale und wahrheitsaffine Ministerialbeamte als Wissensträger und -vermittler im Zentrum. Das Arbeitsethos in Amtsstuben hält zu Sachzwang an. Leistung, Eignung und Befähigung sind sowohl das Nadelöhr zur Aufnahme wie auch Richtschnur zur Ausübung des Amtes. Er heischt nicht nach dem kurzfristig möglichen politischen Akzeptanzerfolg, weil seine Arbeit über den Tag und die Wahlperiode hinausreicht. Ausbalanciert wird er durch den politisch legitimierten Amtsträger an der Spitze der Ministerien. Vertrauenslegitimierte Spitze und leistungslegitimierter Unterbau haben komplementäre Funktion. Der dynamischen, auf Wandel gepolten politischen Ministerialspitze liegt der Implementierungsfokus der ministerialen Fachebene häufig wie ein Klotz am Bein. Es ist diese Dichotomie zwischen Bürokraten und Entscheidern, die eine andauernde, produktive und unerlässliche Spannung erzeugt zwischen Wahrheit und Machtgewinn, Wissen und Glauben, Allgemeinem und Partikularem.

Die Zwitterfigur der Ministerialbürokratie spielt als Wahrheitsproduzent (und Gegengewicht zur politischen Spitze) wie auch als Wissensimportfilter eine wichtige Rolle gegenüber Expertiseangeboten der Wissenschaft. Die Beamten sind gegenüber der res publica im Allgemeinen und gegenüber dem jeweiligen Regierungsamt im Partikularen zu Loyalität verpflichtet. Sie sind kraft Leistung und Kompetenz legitimiert und aufgrund fachlicher Spezialisierung in der Lage, die Abhängigkeit vom externen Expertenrat zu minimieren, indem sie selbst externe Expertise einer inhaltlichen Kritik unterziehen oder sogar selbst produieren. Die Ministerialbürokratie als gate-keeper filtert, diversifiziert und demokratisiert den Expertiseimport für die Verwertung in politischem Handeln.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

V.
Machine Learning and the Law

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

AI in the Judiciary
Practice and Policy

Helena Cecilie Lyng-Olsen & Daniel E. Ho

While numerous efforts have developed and codified AI use case inventories for government in the executive branch, no such inventory exists for the judicial branch. Much focus has been on attorney utilization of AI, but no work has examined the policies courts have instituted to regulate AI use by court employees and judges themselves. We fill this gap in two ways. First, we provide an overview of publicly known AI use cases in the courts, categorizing them by risk and potential impact on judicial discretion (Table 1).[1] *Second, we compile and analyze the internal AI use policies of the 17 U.S. state courts that have promulgated such policies (Figure 1b).*[2] *Our paper proceeds as follows. Section A discusses the methodology for systematically canvassing AI use cases in the courts and state policies. Section B presents an overview of these use cases. Section C analyzes state court policies.*

A. Methodology

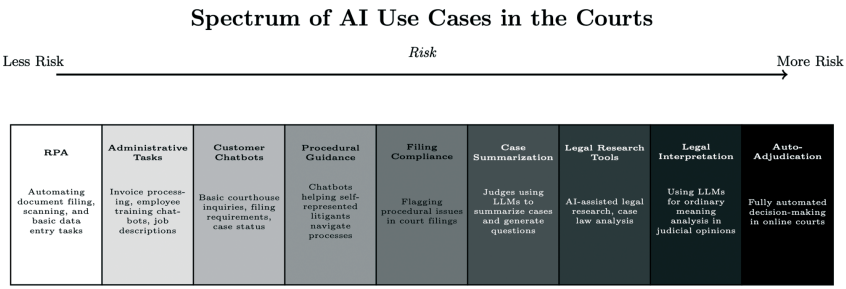
To identify AI use cases in the judiciary, we conducted an extensive manual search of all publicly available sources (performed from February to May 2025), detailed in Appendix B.[3] First, to identify use cases, we used keywords to search (1) judicial opinions on WestLaw, Lexis, and CourtListener, (2) when available, state judicial council websites, (3) publications by the six ABA Judicial Division Conferences and the Judicial Conference of the United States, and (4) general search engines. The most common sources providing information about use cases came from conferences attended by judges, state judicial council meetings, state reports, inter-

1 In Appendix C (https://bit.ly/3ZugNQ6), we include three in-depth case studies from these examples. A comprehensive review of related works can be found in Appendix A (https://bit.ly/3ZugNQ6)

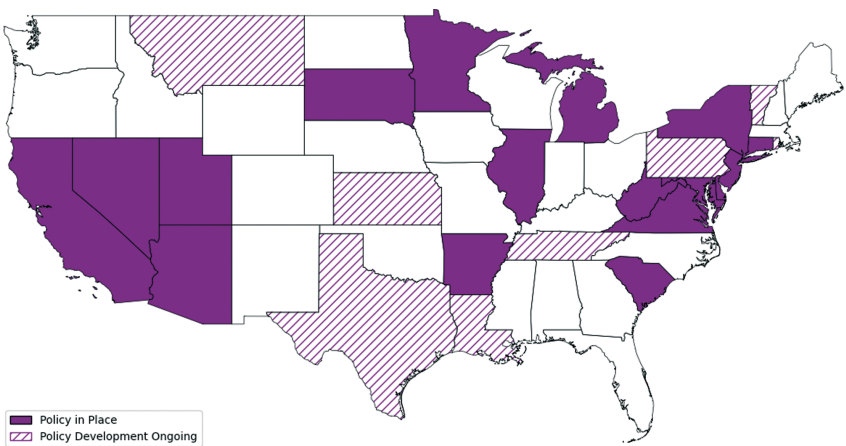
2 We focus on state courts as no federal courts have released formal, public policies governing the use of AI by their employees or judicial officers.

3 https://bit.ly/3ZugNQ6.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -



(a) Spectrum of AI Use Cases in the Courts



(b) Map of State Court Systems that Have Implemented Policies

Figure 1: AI Use Cases and Policy Implementation

views of judges, news articles, speeches, and judicial opinions. Second, to identify internal policies, we used keywords to search (1) each state's judicial branch or judicial council website, along with all federal appellate court websites and the federal Judicial Administration website, (2) general search engines for news articles and reports.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

B. Range of Use Cases

Table 1 provides an overview of AI use cases across countries. We include forms of Robotic Processing Automation (RPA), despite the fact that RPA may not clearly involve machine learning, as these are typically described interchangeably with AI by courts and judges. While fully documented examples of AI use are limited, numerous judges have expressed substantial interest. As Magistrate Judge Allison Goddard put it, "Anyone that comes up with a tool that can accurately summarize a social security record will be worth their weight in gold to every magistrate and district judge!"[1] Table 1 classifies use cases as "administrative" (i.e., aiding in the administration of judicial process) or "judicial" (i.e., aiding in the exercise of judicial discretion). In addition, we provide a rough typology of risk. As we discuss in Section 3, we caution that the same use can be designed and integrated in ways that heighten or mitigate risk, just as administration can bleed into judicial discretion. We also confine ourselves here to AI uses cases outside of administrative adjudication (within the executive branch), but we do include the use of AI for administration in an independent court.

I. Administrative Use Cases

1. Lower Risk Examples

– OCR and Robotic Process Automation for Automatic Case Filing and Categorization:
  Multiple jurisdictions have implemented automated document processing systems that handle court filings. Palm Beach County Courts, along with other jurisdictions, operates one of the most extensive systems, processing approximately 12,000 filings weekly through systems that automatically scan, classify, and assign to judges incoming documents.[2] This system has been operational since 2020 and represents a significant automation of traditionally labor-intensive clerical work.
– HR Tasks and Invoice Processing:
  The District of Columbia Courts have implemented an AI-automated financial processing system that handles invoice receipts and vendor verification.[3] Orange County Courts uses two AI chatbots to provide administrative and legal training to new employees and to assist HR

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

employees with managing personnel records.[4] A municipal court in Texas has begun using ChatGPT to write job descriptions and performance improvement plans, claiming to reduce time spent on such tasks by 60 percent.[2]

- Chatbots for Employee and Basic Customer Support:
  Multiple courts, including Maricopa County Courts, Miami Courts, New Jersey State Courts, and Los Angeles Superior Court, have developed chatbot systems for customer support with varying levels of efficacy and sophistication. These chatbots allow for litigants to ask questions similar to what they might ask at the front desk of a physical courthouse.[2,4–6] While many of these courts do not specify what aspects of their system integrate AI and how, they do claim AI is used in the tool. Independent audits of state chatbots for unemployment assistance have found them to be limited, but no similar audit has been done on these judicial branch chatbots.[7]
- Social Media and Keyword Tracking:
  Maricopa County Courts uses a system to monitor social media platforms for concerning language patterns, conduct keyword analysis to analyze sentiment and threat levels, and identify potential threats against judicial officers.[4]
- Docket Scheduling:
  The New Jersey state court system has developed an AI system that schedules its docket, although detailed information on how that system operates is not currently publicly available.[2]
- AI Court Reporters:
  The Arizona Supreme Court created two lifelike generative AI avatar "reporters" who summarize the court's decisions in YouTube videos posted by the court intended to make court opinions more accessible to the public.[8] Critics worry that AI summarization of opinions, rather than independent journalists analyzing the case, could lead to bias.
- Triaging and Prioritizing Cases:
  Several courts use AI systems to triage incoming filings and prioritize the most urgent cases. The Massachusetts Appeals Court developed a tool summarizing emergency eviction appeals for easier review. Colombia's Supreme Court identifies for expedited review health related cases requiring urgent attention due to potential harm, and Brazil's Supreme Court also uses an AI system that prioritizes pending appeals cases based on potential constitutional significance.[9]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Use Case Type	Examples
ADMINISTRATIVE	
OCR and Robotic Process Automation *Lower Risk*	• Palm Beach County Courts: 12,000 weekly filings processed automatically • Tarrant County Courts: Real-time e-filing monitoring and document classification
HR Tasks and Invoice Processing *Lower Risk*	• DC Courts: AI-automated financial processing system • Orange County Courts: EVA chatbot for employee training, EMI for personnel records • Texas municipal court: ChatGPT for job descriptions and performance plans
Chatbots for Employee and Basic Customer Support *Lower Risk*	• Miami Courts: SANDI for courthouse inquiries in English/Spanish • Maricopa County: Cleo (public-facing), YODA (employee policies), Alfred (IT support) • New Jersey: JIA for public questions and court information • LA Superior Court: GINA for ticket payments and jury duty scheduling • Hawai'i State Judiciary: Kolokolo Chat for court information
Social Media and Keyword Tracking *Lower Risk*	• Maricopa County Courts: AI monitoring for threats against judicial officers
Docket Scheduling *Lower Risk*	• New Jersey state courts: AI-powered docket scheduling system
AI Court Reporters *Lower Risk*	• Arizona Supreme Court: AI avatars summarizing court decisions for YouTube
Triaging and Prioritizing Cases *Lower Risk*	• Massachusetts Court of Appeal: LLMs to summarize emergency eviction appeals for review • Colombia Supreme Court: NLP for urgent health-related claims • Brazil Supreme Court: ML algorithms screening appeals for review
Automated Transcription and Translation *Higher Risk*	• India Supreme/High Courts: Real-time transcription and multi-language translation • Singapore: AI transcription across all state courts • New Mexico and Bahamas: FTR RealTime implementation
Chatbots for Guidance *Higher Risk*	• Alaska State Courts: AVA for probate estate cases • New Mexico Courts: Clara for self-represented litigants • LA Superior Court: Triage and referral tool development
Flagging Issues in Filings and Procedural Compliance *Higher Risk*	• LA Superior Court with Stanford: AI analysis of court filings for procedural issues • Florida Small Claims: AI review of Civil Procedure rules compliance
JUDICIAL	
Summarization of Filings and Question Generation *Lower Risk*	• New Jersey: NJ Courts AI Access for 1,000 users including all judges • Judge Xavier Rodriguez: Case summarization and attorney question generation • Judge Allison Goddard: Technical issue questions and document organization
Creation of Timeline of Events *Lower Risk*	• Judge Xavier Rodriguez: LLMs for key event timelines
Asking LLMs Questions Related to Case Facts *Lower Risk*	• Dutch judge: Scientific questions related to case facts • Mexican judge: Contextual analysis of phrases/concepts • Indian judge: Historical legal concepts like bail jurisprudence
Cross-media Search *Lower Risk*	• Nebraska courts: Search across text, audio, video case exhibits • Judge Anthony Capizzi: IBM Watson for juvenile case multi-media search
AI Tools for Legal Research *Higher Risk*	• Ninth Circuit: WestLaw Quickcheck Judicial tools for all judges • Selected judges testing additional LLM applications
Using LLMs for Legal Analysis and Interpretation *Higher Risk*	• Judge Kevin Newsom: ChatGPT for "ordinary meaning" in Snell v. United Specialty Insurance. Continued use in United States v. DeLeon • D.C. Circuit: ChatGPT for "reasonable" facts determination in Ross v. United States
Anticipating Outcome of Cases *Higher Risk*	• Australia: Predictive algorithms for divorce property division
Assistance with Opinion Writing *Higher Risk*	• Shenzhen Intermediate People's Court: Intelligent Adjudication System for opinion drafting • Use of LLMs to draft court orders in U.S. District Courts (released in error)
Auto-adjudication and Assisted Adjudication *Higher Risk*	• China: Fully automated online courts for e-commerce/IP disputes • Argentina: AI-generated draft rulings based on previous decisions • Mexico: Rule-based expert systems for pension eligibility

Table 1: AI Use Cases in Courts by Type, Examples, and Risk Level

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2. Higher Risk Examples

– Automated Transcription and Translation:
Some courts use automated transcription and translation in their proceedings. India's Supreme and High Courts have systems that simultaneously convert spoken proceedings into written text while providing translation services across multiple regional languages.[9] Singapore provides AI-powered court transcription services across all state courts, and courts in New Mexico and the Bahamas have started using FTR RealTime, an AI tool that instantly creates text from speech.[4] Court-automated transcription and translation run the risk of violating due process if independent evaluation is not in place to ensure that the text has not been tampered with or changed to serve a certain interest.[10]
– Chatbots for Litigant Guidance:
Some courts have developed AI systems that provide procedural guidance for litigants. Alaska State Courts operates AVA (Automated Virtual Assistant), developed through a third-party vendor partnership, which helps users identify appropriate court processes and forms for probate estate cases.[11] The system asks diagnostic questions to determine user needs and provides tailored guidance. New Mexico Courts and Los Angeles Superior Court have deployed chatbots specifically designed to communicate with and assist self-represented litigants,[12] and the Los Angeles Superior Court is also developing a triage and referral tool meant to communicate with and assist self-represented litigants.[13]
– Flagging Issues in Filings and Procedural Compliance:
The Los Angeles Superior Court, in partnership with researchers from Stanford Law School, has prototyped AI to review submitted filings and automatically inform litigants of potential formatting, informational, jurisdictional, or statute of limitations issues.[13] Similarly, the Florida Small Claims Court previously reviewed each of its filings to ensure the Rules of Civil Procedure were invoked, but replaced this task with an AI tool, reducing review time from 15 to 2–3 hours a week.[2] In some instances, use of AI in this domain can be a short term "band-aid" solution for an underlying structural deficiency of the court system; for example, using AI to perform OCR on a paper document and analyze its contents to find missing information could be avoided by simply requiring litigants to fill out an online form with separate fields.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Judicial Use Cases

1. Lower Risk Examples

- Summarization of Filings and Question Generation:
  Some judges in interviews have described using LLMs for summarization and information consolidation. U.S. District Judge for the Western District of Texas Xavier Rodriguez uses LLMs to summarize cases, identify key players involved, and generate questions for attorneys based on their materials.[14] Magistrate Judge for the Southern District of California Allison Goddard uses LLMs to understand what questions to ask attorneys about a technical issue, to summarize and ask questions about orders from other judges, and to organize information from documents.[14]
- Creation of Timeline of Events:
  Judges, such as Judge Xavier Rodriguez, have used LLMs to analyze case documents to create a chronological timeline of relevant events in the case.[14]
- Asking LLMs Questions Related to Case Facts:
  Several judges have been documented consulting LLMs about facts related to their cases, phrases in their cases, and relevant legal history, including judges in the Netherlands, Mexico, and India.[15–17]
- Cross-media Search Across Text, Audio, and Video Case Exhibits:
  Nebraska courts have implemented search capabilities that can analyze and cross-reference information across all forms of case exhibits, including written documents, audio recordings, and video evidence,[18] while former Ohio Judge Anthony Capizzi used IBM Watson to search across his filings and spend more time engaging with the juveniles in his court.[19]

2. Higher Risk Examples

- AI Tools for Legal Research:
  The U.S. Court of Appeals for the Ninth Circuit has implemented a formal program authorizing all judges to use Westlaw Quick Check Judicial tools, and some to use additional LLMs under controlled conditions.[1] Concerns still exist about the accuracy of these tools' outputs.[20,21]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

- Risk Assessments:
  COMPAS, a software that uses data about a defendant to predict their likelihood of recidivism, has been used by courts since as early as 2016. Because of the extensive scholarship around risk assessment tools, we limit our discussion of these here.
- Anticipating Outcome of Cases:
  Australia employs predictive algorithms to anticipate property division outcomes in divorce proceedings, providing judges with statistical analysis of similar cases and suggested property distributions based on relevant factors.[9] This system analyzes historical case outcomes, financial circumstances, custody arrangements, and other relevant variables to suggest likely resolution ranges.
- Using LLMs for Legal Analysis and Interpretation:
  Judge Kevin Newsom of the U.S. Court of Appeals for the Eleventh Circuit was the first U.S. judge to openly rely on and discuss the use LLMs for legal interpretation in a judicial opinion. In *Snell v. United Specialty Insurance Co.*, Judge Newsom consulted ChatGPT to determine the ordinary meaning of "landscaping" in an insurance coverage dispute involving a below-ground trampoline installation. *Snell v. United Specialty Insurance Co.,* 102 F.4th 1208 (11th Cir. 2024). He continued this practice in *United States v. DeLeon*, using AI assistance from multiple sources for ordinary meaning analysis in his concurrence. *United States v. DeLeon*, 116 F.4th 1260 (5th Cir. 2024). The D.C. Circuit subsequently cited Judge Newsom's precedents while using ChatGPT to assess whether certain facts were "reasonable" to know in *Ross v. United States*. 331 A.3d 220 (D.C. Cir. 2025).
  In both cases, judges acknowledged the experimental and potentially controversial nature of their approach. Critics have argued that such use constitutes inappropriate independent fact-finding outside the adversarial process, and that commercial LLMs are subject to corporate manipulation and training methods that may not reflect genuine ordinary usage patterns.[22]
- Assistance with Opinion Writing:
  The Shenzhen Intermediate People's Court in China deployed a system that analyzes case materials to summarize factual information, identify the central legal disputes, generate targeted questions for judges to ask, and write the factual sections of court opinions.[23] Once the judge indicates the decision they wish to make in the case, the system helps craft other sections of the opinion as well.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In interviews, judges using the system expressed concerns about potential over-reliance leading to constrained judicial thinking. Researchers also found evidence of slight bias in judicial rulings made with LLM assistance, not from the AI system itself but from the reinforcement of existing judicial biases.

– Auto-Adjudication and Assisted Adjudication:
China operates some fully automated online courts handling certain e-commerce and intellectual property disputes.[24] These systems manage all aspects of case processing from analyzing filings, conducting discovery, facilitating hearings through chatbots, and applying legal rules, to rendering binding decisions without human judicial involvement.
Argentina has implemented systems in some courts that generate complete draft rulings for new cases from previous similar cases. While judges review these drafts, the AI systems provide initial analysis and suggested outcomes.[9] Mexico's pension courts have a system that generates preliminary decisions for review as well.[25]

C. Internal Policies

U.S. state judicial entities have been at the forefront of developing policies and guidelines addressing the use of AI within the judiciary (Table 3 and Figure 1b). We hence focus on these policies here.

State	Approved Tools									
	Copilot	Claude	ChatGPT	Gemini/Bard	Perplexity	Adobe AI	WestLaw Precision	LexisAI+	Zoom AI Companion	FTR Justice Cloud
Arizona	✓	✓	✓	✓	✓	✓	✓		✓	✓
Maryland	✓	✓	✓	✓						
New York	✓		✓		✓					
Utah		✓	✓	✓						
Virginia							✓	✓		
Connecticut										
Delaware										

Table 2: AI Tools Approved in Different State Court Jurisdictions. States with grey rows require supervisory approval of tools but do not publish approved tools.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

State	Policy Characteristics							Appr. Req.		Flexibility		Ind.
	Bindingness	Number of Pages	Mentions GenAI	AI Use Case Inventory	Disclosure	Monitoring and Evaluation	Includes Judges	For Tool Use	For Use Cases	Local Policy	Review Regularly	Restrictiveness Index
Connecticut	✓	21	✓	✓	✓	✓		✓	✓		✓	0.481
Utah	✓	2				✓	✓	✓	✓			0.296
New York	✓	5	✓				✓	✓				0.148
Maryland		3	✓				✓	✓	✓			0.148
Minnesota	✓	1				✓	✓					0.148
New Jersey		2				✓	✓		✓			0.148
South Carolina	✓	4	✓			✓	✓					0.148
Arizona	✓	4	✓	✓		✓	✓	✓		✓	✓	0.111
Arkansas	✓	3	✓				✓					0.111
California	✓	2	✓		✓		✓			✓	✓	0.111
Michigan		24	✓			✓	✓					0.111
South Dakota		3	✓		✓							0.111
Delaware	✓	4	✓				✓				✓	0.074
Virginia	✓	4	✓					✓		✓	✓	0.037
West Virginia		5					✓					0.000
Illinois		1	✓				✓				✓	0.000
Nevada		2	✓				✓					0.000

"Policy Characteristics" indicates what provisions the policy contains. "Bindingness" indicates whether the policy is intended to be binding to the judicial officer or employee, or conversely whether it is meant to be a non-binding recommendation. "Number of Pages" is the length of the policy, thereby a proxy for the number and specificity of its stipulations. "Mentions GenAI" indicates whether the policy specifically includes generative AI. "Use Case Inventory" indicates whether the state's courts are required to create an inventory of all AI or GenAI uses. "Disclosure" indicates whether there are any provisions requiring public disclosure of use of AI or generative AI for textual, image, or other outputs from the courts. "Monitoring and Evaluation" indicates whether there exists a provision requiring some sort of monitoring of procured or in-house AI or GenAI tools. "Includes judges" indicates whether the policy includes judges or not in its scope. "Appr. Req." indicates that leadership approval is required either "for tool use" (see Table 6)—any AI or GenAI tool being used must be pre-approved first—or "for use cases," meaning approval is required before implementing any new applications. "Flexibility" indicates that the policy contains provisions that have some degree of permissiveness; under that, "Local Policy" indicates that local courts have discretion to develop different policies for their own jurisdictions and "Review Regularly" indicates the policy contains a provision to review and update it at some time. "Ind." indicates index, and the restrictiveness index is calculated to indicate the degree to which the given policy restricts potential usage of AI by judicial officers and court employees. The index was calculated as follows: +3 points for characteristics we consider "strongly restrictive" (bindingness of the policy, whether disclosure of AI use is required, and whether leadership approval is required for use cases); +2 points for characteristics we considered "somewhat restrictive" (if the policy is more than 10 pages long); +1 point for characteristics we consider "loosely restrictive" (AI use case inventory is required; monitoring and evaluation of tools is required; leadership approval is required for specific tool usage); -1 point for characteristics that offer loose flexibility (policy is reviewed regularly and potentially revised); -2 points for characteristics that offer moderate flexibility (differing policies are allowed for local courts). The overall score is then divided by the maximum possible score (27) to create the index.

Table 3: Comparative Analysis of Policies.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I. Current Landscape of Internal AI Policies Across State Courts

A threshold question is whether AI specific policies are necessary. Existing Rules of Court and Codes of Judicial Ethics already govern judicial use of AI to some extent. Illinois's AI fact sheet for judges, for instance, highlights its Code of Judicial Conduct, which requires judges to "decide matters assigned to the judge" and "act at all times in a manner that promotes public confidence in the independence, integrity, and impartiality of the judiciary."[26] Prior efforts to address the rise of technology may also apply. Courts, for instance, have grappled with judicial notice of websites, when such websites can be altered or deleted.[27]

Seventeen states have promulgated AI-specific policies, with eight other states planning to do so. Table 3 characterizes some of the key differences across existing policies. One notable dimension is the force of these policies: six are non-binding recommendations, one is a non-binding advisory opinion, four are administrative orders, and six have amended Rules of Court or the Standard of Judicial Administration (or analogues). The last column of Table 3 also provides a rough measure of the facial restrictiveness of these policies.

II. Discussion

1. The Enforceability Question

These policies present a core question about enforceability. To begin, six states have issued merely advisory, non-binding policies. Second, for states purporting to issue binding policies, enforcement is clearer for employees of the court – with typical employment sanctions – but less clear for judges. California's Judicial Council maintains one policy for court employees and judicial officers in their non-adjudicative roles in the Rules of Court, and another policy for judicial officers in their adjudicative role is adopted into the Standard of Judicial Administration, a set of strongly encouraged "guidelines."[28] Yet prior work has demonstrated significant variability in mechanisms to enforce against judicial misconduct.[29] Third, enforcement against judges raises questions of judicial independence. Connecticut, for instance, appears to require preapproval for the use of AI tools. If a Connecticut judge followed Judge Newsom to explore the use of an LLM for statutory interpretation, should such behavior be sanctioned as violating the policy or be controlled through ordinary appellate review of judicial interpretive methodology?

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Outside of the judiciary, we have already seen the limits of these policies. New York State's audit of its initial state executive branch agency AI policy revealed widespread "inadvertent noncompliance" and significant reliance on individual discretion in determining appropriate use, despite formal policy requirements.[30] While some jurisdictions attempt to strengthen compliance through deterrent language, such as North Carolina's state executive branch policy's warning of "severe penalties" for violations,[31] the effectiveness of such approaches remains untested in the judicial context.

2. Policy Ambiguity

Heightening the enforcement challenge is significant ambiguity in these policies. West Virginia's advisory opinion declares "A judge should NEVER use AI to reach a conclusion on the outcome of a case."[32] But the boundary between AI for legal research and "reaching" a conclusion is muddy. Other states, such as California, do not explicitly prohibit its courts from using AI for adjudicative purposes, but instead provide high level principles – privacy, bias, accountability, and transparency – when much scholarship has focused on often subtle tensions between such principles. Some policies, such as Connecticut's, require citation of "use of AI" when "required by law," although Connecticut's does not specify what "use of AI" encompasses, nor which laws would be relevant. "Use of AI" could range from fixing grammar and spelling mistakes to complete rewording, and clearer language would make these policies easier for judicial officers and employees to achieve compliance. All these efforts obscure the difficulty of drawing lines given the spectrum of AI integration (Figure 1a).[33]

3. Regulatory Burden

What is also notable is the significant variation in the regulatory burden for judicial adoption of AI, as measured by the restrictiveness index. At one extreme, court systems in states such as Connecticut have frameworks that impose extensive compliance obligations on judicial personnel, creating detailed procedural requirements that, while thorough, may prove burdensome. In Connecticut, a judicial branch division looking to use an AI system must submit a request to the State Judicial Branch's AI Committee describing the purpose of the request and considerations relating to various privacy, IP, and procurement concerns, agree to monitoring and

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

evaluation while using the tool, acquire approval from a supervisor before each new use of the tool, and cite use of AI "where required by law."[34] At the other extreme, court systems in Illinois and Nevada have adopted minimalist approaches, offering a small amount of optional guidance for judicial officers and court employees alongside unrestricted AI use.[26,35]

California and Arizona's judicial policies present a more balanced regulatory posture, establishing guardrails alongside relatively light reporting requirements. California requires courts to take measures to ensure privacy, bias minimization, accountability, and transparency when using the tools, and also allows courts to individually choose to unilaterally ban use of generative AI if they so wish; Arizona allows AI and GenAI use with certain tools, requires courts to seek approval for tools and use cases outside that realm, and documents use cases publicly on its website.[28,36]

The core challenge is calibrating safeguards to risk, given the spectrum of AI integration. As illustrated in Section 3, some use cases pose minimal risks to core judicial functions, and too much process may lead to suboptimal innovation. While California's policy presents separate policies for AI for adjudicative uses and AI for non-adjudicative uses, it fails to account for the differing risks within those categories. While much focus has been on ex ante procedural controls, too little attention has been placed on ex post evaluation and post-deployment monitoring. One positive counterexample is that Michigan recommends the establishment of pilot programs, qualitative and quantitative evaluation, and iteration for AI tool use before scaling.[37]

4. Transparency

The challenges we faced in identifying judicial use of AI illustrate why others have attempted to codify such requirements for the executive branch. Some of the policies improve transparency around court and judicial use of AI. Connecticut and Arizona require judicial officers and court employees to report use cases to their respective AI committees to "identify emerging use cases and opportunities for knowledge sharing" and "maintain a list of AI tools" respectively.[34] Such transparency allow third parties – or other courts – to evaluate and potentially adopt the use case themselves. Arizona's online use case inventory provides insight into its operations, and court employees who have championed its AI use have become sources of expertise to other jurisdictions.[37]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

That said, the implementation of an AI use case inventory is not straightforward. In the U.S. federal context, 76% of agencies failed to publish an inventory within two years, including 48% of large agencies actively known to have AI use cases in 2020.[38] Currently, Connecticut's Judicial Branch Artificial Intelligence Committee website announces that it "conducted an inventory of its internal systems and has determined that it does not employ artificial intelligence in its applications."[34]

Table 2, for instance, shows that while preapproval of AI tools is required, no information exists in Connecticut and Delaware on the preapproved tools. Tool-based approval may also obscure the range of ways a specific tool may be used. It is perplexing, for instance, that Maryland approved ChatGPT, but not LexisAI, when legal AI tools have reduced hallucination rates.[20] Consider the contrast of Judge Newsom. While one can dispute the appropriateness of utilizing Claude and ChatGPT for statutory interpretation, one virtue is the transparency that triggered extensive commentary on the approach.[22,39–41]

Acknowledgments

We thank the Daimler and Benz Foundation for supporting this work.

References

1. Generative AI in Action – Exploring Generative AI Tools for Legal Practice. Ninth Circuit Judicial Conference. Youtube. (2024). https://perma.cc/82DR-MVMR.
2. National Center for the State Courts. Artificial Intelligence in State Courts: Promise and Peril. (2025). https://perma.cc/J2LF-6K44.
3. Thomson Reuters Institute/National Center for State Courts AI Policy Consortium. AI in Action – Current Applications in State Courts. (2025). https://vimeo.com/1082965262.
4. National Association for Court Management. Courting AI: Understanding Artificial Intelligence in Courts. (2025). https://nacmnet.org/resources/store/.
5. New Jersey Courts. NJ Courts AI Access: How the Judiciary is Using Secure, Internal AI to Improve Court Services. (2024). https://njsbamaterials.com/MAM25107.pdf.
6. Sigler, E. SANDI: Improving Court Access and Service in Miami with an Advanced Artificial Intelligence Chatbot. Trends in State Courts. (NCSC, 2023).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

7. Ho, D.E., Martin, O., Rodolfa, K., & Perez, A. Evaluation as Due Process: Civil Service in an Automated Age. Adm. Law Rev. (forthcoming). (2025).
8. Arizona Supreme Court Administrative Office of the Courts. Arizona Supreme Court Introduces AI-Generated Court News Reporters to Enhance Public Engagement. [Press release]. (2025).
9. UNESCO Global Judges' Initiative: Survey on the Use of AI Systems by Judicial Operators. (2024). https://perma.cc/PBD2-WFSN.
10. Jones, T., Kalbfeld, J. R., Hancock, R. & Clark, R. Testifying while black: An experimental study of court reporter accuracy in transcription of African American English. *Language* **95**, 216–252 (2019).
11. LawDroid. Alaska Courts Virtual Assistant: Preview and Demo. Youtube. (2025). https://www.youtube.com/watch?v=ZC0_Q0HnAQA.
12. New Mexico Courts. Language Access Services: Clara. New Mexico Courts. https://languageaccess.nmcourts.gov/clara/.
13. Lynch, S. Harnessing AI to Improve Access to Justice in Civil Courts. Stanford HAI. (2025). https://hai.stanford.edu/news/harnessing-ai-to-improve-access-to-justice-in-civil-courts.
14. O'Donnell, J. Meet the early-adopter judges using AI. MIT Technology Reviewhttps://www.technologyreview.com/2025/08/11/1121460/meet-the-early-adopter-judges-using-ai/.
15. Pascoe, R. Dutch judge uses ChatGPT to help reach a verdict. DutchNews. (2024). https://perma.cc/EX5S-3PAL.
16. Gutiérrez, J.D. Judges and magistrates in Peru and Mexico also caught the ChatGPT 'fever.' Agenda Estado de Derecho. (April 18, 2023). https://perma.cc/LQJ7-QL6V.
17. Sandhu, J. S. In a first, Punjab and Haryana HC turn to ChatGPT for view on bail in murder case. The Indian Express. (2023). https://perma.cc/C54E-KX4T.
18. Shetty, A. Nebraska Judicial Branch modernizes its Electronic Exhibits System using AWS. AWS Public Sector Blog (2023). https://perma.cc/SVJ5-JHMS.
19. Capizzi, A. Helping Juvenile Treatment Court Improve Efficiency and Outcomes with IBM's Watson Health Solution. Int'l Pers. of Crime Prevention. **12**, 67–74. (2023)
20. Magesh, V. *et al.* Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. J. Empir. Leg. Stud. 22, 216–242 (2025).
21. Dahl, M., Magesh, V., Suzgun, M. & Ho, D. E. Large legal fictions: Profiling legal hallucinations in large language models. J. Leg. Anal. **16**, 64–93 (2024).
22. Grimmelmann, J., Sobel, B. & Stein, D. Generative Misinterpretation. Harv. J. Legis. **63**, (2025).
23. Liu, J. Z. & Li, X. How do judges use large language models? Evidence from Shenzhen. J. Leg. Anal. **16**, 235–262 (2024).
24. Guo, M. Internet court's challenges and future in China. Comput. Law Secur. Rev. **40**, 105522 (2021).
25. Valverde, D. Supreme Court of Mexico Unveils AI Sor Juana for Judicial Clarity. Mexico Business News. (June 27, 2024). https://perma.cc/S65W-EWJQ.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

26. Illinois Supreme Court Policy on Artificial Intelligence. Supreme Court of Illinois. (January 1, 2025). https://perma.cc/4HQR-N65G.
27. Barger, C. M. On the Internet, Nobody Knows You're a Judge: Appellate Courts' Use of Internet Materials. J App Pr. Process **4**, 417 (2002).
28. Rule and Standard for Use of Generative Artificial Intelligence in Court-Related Work. Judicial Council of California Artificial Intelligence Task Force. (2025). https://perma.cc/QV7E-SFSM.
29. Abel, J. Testing three commonsense intuitions about judicial conduct commissions. Stanford L. Rev. **64**, 1021 (2012).
30. DiNapoli: Improved Guidance Needed for State Agencies Using AI To Avoid Risks. Office of the New York State Comptroller [Press release]. (2025). https://perma.cc/LGB9-EB9U.
31. Use of Publicly Available Generative AI. North Carolina Department of Information Technology. https://perma.cc/5VZP-SS42
32. RE: Judicial Investigation Commission Advisory Opinion 2023–22. West Virginia Judicial Investigation Commission. (2023). https://perma.cc/H77X-8GH8.
33. Martin, O., Surani, F., Rodolfa, K., Perez, A. & Ho, D. E. The Spectrum of AI Integration: The Case of Benefits Adjudication. AI: Legal Issues, Policy, and Practical Strategies (American Bar Association, 2024).
34. State of Connecticut Judicial Branch. *JBAPPM Policy 1013 Artificial Intelligence Responsible Use Framework*. (2024).
35. Artificial Intelligence Guide. Nevada Administrative Office of the Courts. (2025). https://perma.cc/9LPU-RVW2
36. Section § 1–509: Use of Generative Artificial Intelligence Technology and Large Language Models. Arizona Court Rules. (2024). https://perma.cc/WL3H-84R4.
37. Generative AI and the Courts Workgroup: Report and Recommendations. Michigan Judicial Council (2024). https://perma.cc/GQE7-WADP.
38. Lawrence, C., Cui, I. & Ho, D. The Bureaucratic Challenge to AI Governance: An Empirical Assessment of Implementation at U.S. Federal Agencies. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 606–652 (ACM, 2023).
39. Curl, J., Henderson, P., Kandula, K. & Surani, F. Judges Shouldn't Rely on AI for the Ordinary Meaning of Text. Lawfare (2025).
40. Engel, C. & McAdams, R. H. Asking GPT for the ordinary meaning of statutory terms. U Ill JL Tech Pol 235 (2024).
41. Waldon, B., Schneider, N., Wilcox, E., Zeldes, A. & Tobia, K. Large Language Models for Legal Interpretation? Don't Take Their Word for It. Georgetown L. J. **114** (forthcoming). (2025).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

AI Alignment and Legal Reasoning

Elliott Ash

This chapter explores the links between legal reasoning and the post-training regimes of aligned large language models (LLMs). Reinforcement learning from human feedback (RLHF), which optimizes model outputs to match graded human preference rankings, leads to LLMs that perform value-weighted interpolations across task-response pairs. Reinforcement learning with verifiable rewards (RLVR), which rewards outputs that pass objective external checks, produces LLMs that faithfully follow structured reasoning paths. The RLHF approach parallels the analogical, precedent-based reasoning characteristic of common-law systems, while RLVR mirrors the deterministic code-based reasoning associated with civil law. These analogies inform the design of multi-agent legal AI systems, which should deploy RLVR-based LLMs for compliance with bright-line rules, while deploying RLHF-based systems to apply more subtle standards.

A. Introduction

Legal systems and machine-learning systems share a central design problem: aligning individual reasoning with collective standards. In both domains, outcomes depend on how reasoning processes internalize shared norms and translate them into consistent decisions. In the law, two broad forms of reasoning address this task. Case-based reasoning derives conclusions by analogy to prior decisions, weighting precedent by relevance and authority. Code-based reasoning applies determinate rules to specified conditions, aiming for predictability and uniformity. Legal institutions balance these approaches to minimize error and promote legitimacy.

An analogous division in reasoning modes defines the latest generation of large language models (LLMs) – and more specifically, the two families of algorithms that are applied to align LLMs with human aims. Through reinforcement learning from human feedback (RLHF), models are tuned to reflect human evaluative patterns, internalizing social preferences much like courts internalize precedent (Stiennon et al. 2020; Ouyang et al. 2022). Through reinforcement learning with verifi-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

able rewards (RLVR), models are constrained by formal verification, and then learn to follow explicit reasoning paths (Lambert et al. 2024). These training architectures reproduce the same institutional trade-off that legal systems face, with RLHF better suited to case-based reasoning based on previous examples, and RLVR better suited to code-based reasoning following deterministic rules.

This chapter explores this parallel and its implications for the design of legal AI. Section 2 starts with a review of how LLMs are built. Modern models are trained first by next-token prediction on vast corpora. Transformers make this scalable by attending to long-range context, which helps models represent syntax and semantics and supports generalization across domains (Vaswani et al. 2017). Instruction fine-tuning (IFT) teaches models to follow prompts, while RLHF steers model outputs to match human preference rankings (Stiennon et al. 2020; Ouyang et al. 2022). As a result, RLHF-aligned LLMs generate responses by performing an interpolation over the training and instruction corpora, weighted by their scores on the reward model.

Recent LLMs adopt a new approach: RLVR, which moves from graded human scores to objective checks (Lambert et al. 2024). The model proposes incremental steps of reasoning, which are then externally verified for validity and consistency. RLVR encourages explicit chains of thought and stepwise correctness, which is why it shines on logic, math, and code (Lightman et al. 2024; Lambert et al. 2024). In short, RLHF tunes for context-sensitive judgment; RLVR trains for structured logical reasoning.

I continue in Section 3 by connecting these ideas from LLMs with analogous ideas from the law. Case-based reasoning in the common law aggregates and weights prior decisions by relevance and authority. It is an inductive, analogical process, well captured by classic work in AI and law on case comparison and argument patterns (Ashley 1992; Rissland, Ashley, and Branting 2005). That makes it a natural match for RLHF, which aggregates graded human judgments across many contexts. Rule application and statutory reasoning, by contrast, often turn on element satisfaction and nested conditions. This is closer to RLVR's logic: correctness is defined outside the decision maker, and the task is to check whether the elements are met.

This connection extends to the classic legal distinction between rules and standards (Kaplow 1992). RLHF aligns with context-sensitive standards; RLVR aligns with bright-line rules. Taking a step further, insti-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tutional form matters too, in particular around comparative legal origins (Mattei 1997; La Porta et al. 2008). Common-law systems, evolving through precedent and analogical development, are well-aligned with an RLHF-based approach to reasoning. Civil-law systems, grounded in comprehensive codes and deductions from statutory text, are best-suited to RLVR LLM systems.

There are clear implications for the design of legal AI systems, as explored in Section 4. It is unlikely that there is a one-size-fits-all legal agent. Instead, we should develop coordinated multi-agent systems that route sub-tasks to different models. Determinate statutory checks should go first to an RLVR-trained reasoner that can verify element satisfaction step by step. Ambiguous issues that hinge on community standards and fact sensitivity should go to an RLHF-trained reasoner that aggregates across analogous precedents. The system mirrors how human judging already mixes verification with evaluation, while making explicit the normative choices behind model training and task assignment. The architecture of AI is linked directly with the architecture of law.

B. How Large Language Models Are Trained

This section outlines how large language models are trained and aligned through successive stages that shape their linguistic and reasoning capacities. The base model learns statistical patterns in text, while post-training stages progressively align it with human instructions, preferences, and formal correctness. Specific attention is paid to the two main approaches for post-training – RLHF and RLVR – and the tasks they are designed to perform.

I. Sequence Models, Embeddings, and Transformers

Modern LLMs grow out of sequence modeling in language. The core task is language modeling – estimating the conditional probability of the next token given a context. Early systems used n-gram counts with smoothing (Shannon 1951; Jurafsky and Martin 2023). Neural language models replaced counts with distributed representations that support generalization across contexts (Bengio et al. 2003), including word embeddings mapping discrete tokens to context-sensitive vectors (Mikolov et al. 2013).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The early neural language models were recurrent neural networks (RNNs; Hochreiter and Schmidhuber 1997; Sutskever, Vinyals, and Le 2014). RNNs process sequences token by token, which limits long-range context and parallel processing. An architectural innovation called the attention mechanism, introduced by Bahdanau, Cho, and Bengio (2015), lets models weight relevant parts of the input directly. The transformer architecture is founded around attention – it removes recurrence, uses multi-head self-attention and feed-forward layers plus positional encodings, and models dependencies over long contexts in parallel (Vaswani et al. 2017). Encoder-only models such as BERT support representation learning and classification (Devlin et al. 2019), while autoregressive (decoder-only) language models such as GPT support open-ended generation (Radford et al. 2019). Scaling of training data, parameters, and context windows have improved downstream performance across tasks (Kaplan et al. 2020; Brown et al. 2020).

Autoregressive models like GPT have led the way in the recent rapid development in AI. They are large attention-based neural networks trained on language modeling – to predict the next token in a sequence – following a massive corpus of text. With enough data and trainable parameters, autoregressive LLMs begin to exhibit features of generalized intelligence, and to solve a broad variety of tasks (Brown et al. 2020).

Large scale pretraining works to equip the model with broad knowledge and the ability to fluently continue text. Base LLMs predict sequences based on the large training corpus. They can generate creative sequences, unseen in the training data, by interpolating across those corpora. This allows, for example, base LLMs to "guess" unobserved facts through analogizing between observed facts.

II. Preference Alignment: IFT and RLHF

Pre-trained base LLMs are powerful language interpolators. While this interpolation of new sequences often works well, it also often goes sideways. Strange and incorrect claims will often come from base models – a phenomenon referred to as hallucination. More generally, a downside of pre-trained base models is that they are not well-aligned with human intentions. For example, if I ask a question, a base model might respond with another question, rather than responding with an answer. The next

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

steps of "post-training" are designed to address this issue – to *align* LLMs with human preferences.

Instruction fine-tuning (IFT) is a stage of post-training that shifts a language model from simply predicting the next word in a sequence to actively following user prompts. In IFT, token prediction training is undertaken on special corpora: collections of prompt–response pairs that teach the model how to carry out specific tasks and produce outputs in desired formats (Chung et al. 2022). By exposing the model to a wide and diverse set of instructions, researchers enable it to generalize better to new prompts, improving its zero-shot and few-shot performance (Wei et al. 2022).

The next step in LLM alignment is RLHF – Reinforcement Learning from Human Feedback. The goal is to improve upon IFT in aligning outputs with user preferences (Stiennon et al. 2020; Ouyang et al. 2022). RLHF starts with a dataset of pairwise comparisons – user prompts, with at least two different responses, ranked by their usefulness to humans. The system designer trains a reward model based on these comparisons to predict which output humans would tend to prefer. The trained reward model can then provide scores to texts based on their tendency to be preferred to baseline statements. RLHF then optimizes the text generator to increase that score while staying close to the IFT model's generations. In practice, that means better tone, structure, and reason-giving, for example in legal advice.

LLMs that have been aligned with IFT and RLHF provide more useful generations. IFT model generations are interpretable as interpolations across a weighted average of the pre-training and instruction-following corpus. After RLHF, the interpretation is a little more complex. It is an interpolation across the pre-training corpus, IFT corpus, and reward pairs corpus, weighted by the predicted usefulness of the generation.

III. Alignment for Structured Reasoning: RLVR

Reinforcement Learning with Verifiable Rewards (RLVR) extends the logic of alignment by replacing human judgment with automatic checking (Lambert et al. 2024). Instead of asking people which answer they prefer, the system rewards the model only when its output passes an objective test. The idea is simple but powerful: for any task where correctness can be defined in advance, the model learns to generate outputs that a verifier

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

accepts as true. The verifier can be a program that executes code, a mathematical solver that confirms an answer, or any other function that returns a pass or fail signal.

The training setup mirrors other RL post-training pipelines: the model proposes candidate answers; a checker scores them automatically; updates push the policy toward higher verified success while staying close to the base model to preserve fluency and breadth. A critical feature of this approach is to break down complex problems into a series of simple steps, which can then be solved and verified one-by-one (Lightman et al. 2024). This keeps the focus on correctness and avoids brittle format-chasing. RLVR reduces reliance on costly labels and guards against reward-model gaming, but it is only as broad as the verifiers available. Because checkers can be brittle, designers often add light smoothing or combine RLVR with instruction-following or preference objectives to keep learning signals robust (e.g. DeepSeek-AI et al. 2025).

RLVR-based training is a key ingredient in the recent breakthroughs by "reasoning"-based LLMs, such as o1 and DeepSeek-R1, on math and coding tasks (OpenAI 2024; DeepSeek-AI et al. 2025). The other main ingredient is the introduction of internally observed thinking tokens, which allow the model to perform a lengthy chain of thought that is not (by default) revealed to the user. Alongside that, a large corpus of lengthy high-quality reasoning traces is observed by the model during the IFT phase, in particular to break down complex problems into simple, verifiable steps. Thus the LLM learns the style and structure of long-form reasoning.

Because the reward is tied to external correctness, RLVR reduces dependence on noisy, biased, or subjective human evaluations. It is therefore well suited to structured reasoning tasks such as mathematics, code generation, or logical proofs, where success can be defined as the satisfaction of formal conditions. Relative to IFT and RLHF models, which are well-designed to interpolate among preference-weighted examples, RLVR trains models to follow discrete lines of reasoning that take logically correct paths at each juncture.

IV. Discussion of Post-Training Algorithms

To analogize to classical machine learning algorithms, IFT and RLHF can be seen as akin to K-nearest-neighbors algorithms, where similar prompts are identified in the training corpus by their embeddings, and then the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

associated responses are weighted based on their predicted user utility and averaged to produce a response. Meanwhile, RLVR is more akin to a decision tree, where the nodes are verifiable reasoning junctures and the branches are the potential set of answers at that juncture. RLVR teaches the LLM to follow the correct branches and eventually to terminate at an overall satisfactory answer.

Can LLMs trained with IFT/RLHF or RLVR extrapolate to new questions or tasks that are outside the training data? This is an active area of research (Chung et al. 2022; Ouyang et al. 2022; Lambert et al. 2024). A limited "yes" answer to this question is the phenomenon of in-context learning: if new documents are provided with the necessary information to answer questions, then an LLM can answer factual questions about those documents even if they did not appear in the training. If a task requires new value judgments, however, it is unclear how an LLM extrapolates to make such judgments. RLVR, meanwhile, enables an important form of extrapolation – that which involves following verifiable reasoning paths. In principle, RLVR models can extrapolate to new logic, math, or code problems by solving them step-by-step.

So far, I have depicted RLHF and RLVR as separable algorithms defining different LLMs. But in practice, they are not cleanly separated. RLHF datasets have many examples of math and code, where the "correct" answer is annotated as that preferred by users. And RLVR datasets with sets of "correct" labels presumably correspond to what users would pick. The main difference is that the RLHF datasets also contain many subjective comparisons, such as the quality of writing, judgments on levels of sentiment or toxicity, or political or moral judgments, all of which reflect human social and cultural norms rather than objective correctness. Further, the performant RLVR models, such as DeepSeek-R1, are trained on both RLVR and RLHF objectives during the post-training phase (DeepSeek-AI et al. 2025). This means that, potentially, such models can still perform the subjective, interpolative tasks for which RLHF is well-designed. However, Ni et al. (2025) show that RLVR models perform worse than RLHF models in modeling subjective decisions (e.g. toxicity detection), suggesting that RLVR training interferes with the LLM's ability to represent distributed preference sets.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. LLM Alignment and the Law

This section explores legal reasoning and legal institutions in light of the architectures of large language models. Both domains confront the problem of aligning inference with normative standards – deciding when reasoning should follow explicit rules, rely on experiential judgment, or combine the two. I trace parallels between legal reasoning and model training, focusing on how RLHF corresponds to case-based reasoning, and RLVR corresponds to code-based reasoning.

I. LLM Reasoning and Legal Reasoning

The structure of legal reasoning can be illuminated by comparing it to the successive stages of large language model training. Law, like generative AI, develops through processes of exposure, refinement, feedback, and verification. The steps in legal thought and procedure find parallels in how models are trained to reason – from the absorption of precedent, to doctrinal synthesis, to analogical judgment, to formal rule application.

Legal understanding begins with immersion in precedent and custom, much as models begin with pretraining. Judges, lawyers, and students internalize the language and logic of past decisions by reading widely and repeatedly. Over time, they develop a sense of what arguments sound plausible and what outcomes cohere with established norms (Lamond 2014). This process is not deductive but associative: legal actors, like pretrained models, form a broad intuitive map of patterns and meanings before any explicit rules are applied.

Doctrinal reasoning resembles instruction fine-tuning. From the diffuse materials of experience, jurists extract structured formulations that can guide future decisions. Take in particular the development of legal commentaries or restatements of the law. These materials are developed to interpret facts through settled principles, to follow procedural instructions, and to translate broad concepts into determinate tests. Just as IFT teaches a model to respond coherently to prompts, doctrinal reasoning teaches lawyers to connect precedent to principle through organized, rule-like reasoning.

What about reinforcement learning from human feedback? That element of LLM training can be analogized to the analogical and equitable dimensions of legal reasoning. Judges do not simply execute fixed rules –

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

they reason by analogy, drawing on previous judgments to decide whether an outcome is contextually justified. Each precedent can be seen as analogous to a pairwise comparison in an RLHF dataset: it features a set of facts (a prompt), two sets of legal arguments from the petitioner and respondent (the two responses for comparison), and a judgment saying which of the two arguments wins (the feedback label). The appeals process acts as an additional subjective feedback system, leading future judges to upweight precedents that have been endorsed by higher courts, and downweight precedents that have been reversed. In addition, the citation network among judges creates a third set of human feedback signals, in which future judges are likely to follow precedents that have been examined and cited by peer judges in the interim. For any given decision, judges collect and aggregate these feedback labels when shaping their arguments and judgments.

Finally, the application of determinate legal rules parallels reinforcement learning with verifiable rewards. When statutes define clear conditions and consequences, the role of the judge is largely to verify whether the elements are satisfied. The process is procedural and externally checkable, not interpretive or analogical. RLVR captures this same mode of code-based reasoning – rule application under an objective verifier, where correctness is determined by a standard outside the decision-maker's discretion.

Seen through this lens, legal reasoning embodies the same layered architecture that defines modern model training. Precedent provides the base of intuitive understanding, doctrine structures that knowledge into usable form, case-based reasoning approximates learning from feedback, and rule application enforces verifiable correctness. Law, like model alignment, thus operates by assembling a variety of learning systems.

II. LLMs and Legal Design

There are corresponding parallels in how LLMs are designed and how legal procedures should be organized. This subsection elaborates on the connection to rules and standards, and *law as data* versus *law as code*. Institutionally, there is an informative distinction between common-law and civil-law jurisdictions.

The long-standing distinction between rules and standards turns on when legal content is specified and how determinate it is (e.g. Kaplow

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

1992). Rules fix content ex ante, inviting relatively mechanical application. Standards defer content to ex post adjudication and invite balancing. This distinction maps directly onto the difference between RLVR and RLHF. RLVR systems, trained to satisfy externally verifiable criteria, perform best when legal content is rule-like – when conditions and consequences can be specified in advance. RLHF systems, trained to approximate context-specific human judgment, are better suited to standard-like reasoning, where outcomes depend on contextual evaluation. Applying a standard involves identifying relevant analogies, weighting them by their persuasiveness and proximity, and then averaging across them to reach a balanced conclusion. This process mirrors the graded preference learning that defines RLHF.

The fit between legal form and model architecture also parallels the emerging distinction between *law as data* and *law as code* (Livermore and Rockmore 2019). *Law as data* treats legal materials – opinions, briefs, statutes – as empirical sources for retrieval and synthesis. It focuses on understanding how legal actors reason and predict how they will decide. *Law as code*, by contrast, seeks to formalize statutes and procedures into executable logic, converting legal rules into programmable conditions. The analogy is straightforward. *Law as data* corresponds to instruction fine-tuning and RLHF: the signal is graded, context-sensitive, and oriented toward interpolation across precedents. *Law as code* corresponds to RLVR: the signal is discrete, oracle-tied, and based on objective verification. Together, these approaches suggest that legal AI systems will evolve along both trajectories – one oriented toward interpretive reasoning from data, and the other toward executable reasoning from rules.

The two dominant varieties of legal systems – common law and civil law – reflect a deep structural divide between case-based and code-based reasoning (Zweigert and Kötz 1998; La Porta et al. 2008). Common law systems evolve through precedent: courts decide individual disputes and justify outcomes by reference to earlier decisions (Gennaioli and Shleifer 2007). Legal doctrine develops inductively, as patterns of reasoning are inferred from case to case. Civil law systems, by contrast, begin with a comprehensive code that specifies legal rules and their conditions of application. Judicial reasoning proceeds deductively, applying the relevant statutory provisions to the facts at hand.

This common law mode of reasoning aligns closely with reinforcement learning from human feedback. RLHF systems learn through graded evaluation, aggregating many examples of preferred outcomes to approxi-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

mate the judgments of a community of evaluators. Similarly, common law reasoning proceeds by analogy to previous cases, weighting prior decisions by their relevance, authority, and coherence with general principles. The inductive and value-sensitive nature of both processes makes RLHF especially well suited to assist judges in common law settings, where legitimacy depends on responsiveness to evolving norms and shared intuitions of fairness.

Civil law reasoning, by contrast, corresponds more directly to reinforcement learning with verifiable rewards. RLVR systems operate under explicit verification criteria: outputs are accepted only when they satisfy an external test of correctness. This mirrors the logic of statutory adjudication, in which a judge determines whether the legal elements are met and then applies the rule mechanically. Because RLVR systems can manage highly structured and nested rule sets, their adoption could support the drafting of more precise and internally consistent codes. Over time, such systems may help civil law jurisdictions move closer to the "law as code" ideal, enhancing transparency and reliability in rule application. This trajectory is already visible in code-based systems such as China's, where judicial practice discourages citation of precedent. China has had early, large-scale adoption of LLM-based decision support tools (e.g. Liu and Li, 2024).

D. Implications for Legal AI

The parallels between how models are trained and how legal systems work have implications for how AI should be used in the legal system. The architectures of RLHF and RLVR do not just resemble different legal reasoning styles; they also suggest how AI systems might align with different legal tasks, doctrines, and institutional forms. In thinking about the integration of AI into law, it is therefore useful to ask which aspects of legal decision-making are best handled by systems trained on preference-based rewards and which by systems trained on verifiable ones.

RLHF-based systems are naturally suited to tasks that require interpretive or evaluative judgment. Because these models learn from graded, context-sensitive feedback, they can interpolate across a range of factual situations and moral intuitions. Many legal standards – such as "reasonableness," "good faith," or "due care" – operate precisely in this way. Deciding whether conduct was negligent or an offer was made in good

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

faith depends on the aggregation of many contextual features, as well as on sensitivity to community expectations. An RLHF-based model, trained on extensive corpora of judicial reasoning and fact patterns, could approximate this inductive process by identifying clusters of analogous cases and synthesizing their logic. In effect, RLHF offers a scalable mechanism for producing judgments that reflect the continuous and socially grounded nature of standard-based adjudication.

RLVR systems, by contrast, are better aligned with rule application. These models operate under explicit verification criteria, rewarding outputs only when they satisfy external correctness conditions. This makes them well suited to legal domains where compliance can be defined with precision – taxation, regulatory compliance, contract validation, and other areas governed by intricate statutes and administrative rules. RLVR can encode and execute arbitrarily complex legal frameworks, checking whether each factual element or computational step satisfies the relevant conditions. A legal reasoning engine built around verifiable rewards could thus perform the functional equivalent of statutory interpretation and rule enforcement, translating legal code into executable code and ensuring consistency and precision in outcomes.

In practice, however, there is no clean division between subjective and objective tasks, nor between case-based and code-based reasoning. Legal decision-making typically involves a combination of the two. Even in highly codified domains such as tax law, subjective judgments often play a decisive role – for example, in determining whether an expense was incurred for personal or business reasons (Goldin et al. 2024). Similarly, areas traditionally governed by standards still rely on statutory definitions and procedural rules. Just as RLHF and RLVR operate in tandem during model training, human legal reasoning mixes evaluation and verification, weighing analogical arguments against fixed legal criteria.

This hybrid character of law points toward an analogous architecture for legal AI. Effective systems will need to integrate multiple reasoning modes, drawing on both RLHF and RLVR models. An outline for a plausible design is a multi-agent system in which different components specialize in different tasks and route questions accordingly. Determinate rule-following could be handled first by an RLVR model, which checks the applicable statutes or procedural rules. Ambiguous or threshold-based issues could then be referred to an RLHF model, which aggregates reasoning across similar cases to approximate equitable judgment. This layered design echoes recent work in AI on routing among specialized models to

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

enhance performance and efficiency (Chen et al. 2023; Wang et al. 2024; Ong et al. 2024).

The parallels between AI alignment and legal reasoning are more than conceptual – they offer a framework for designing systems that can think with law, not merely about it. The next step is empirical. Hybrid architectures that combine RLHF's preference-based reasoning with RLVR's verifiable reasoning should be tested systematically on legal AI benchmarks, such as LEXGLUE, LegalBench, and Lexam (Chalkidis et al. 2022; Guha et al. 2023; Fan et al. 2025). Such evaluations can help determine whether the hypothesized match – RLHF with case-based reasoning and RLVR with code-based reasoning – holds in practice, and may motivate new benchmarks that more clearly distinguish the two.

Beyond benchmarking, legal AI systems must be tested in the field. Liu and Li (2024) show that judges using chatbots often render decisions first and then use AI to rationalize them after the fact. Future systems should instead support reasoning itself – helping judges deliberate, structure arguments, and reach better-grounded outcomes. Doing so requires attention not only to logic but also to psychology and institutions, as judges respond to framing, workload, and norms (Engel and Glöckner 2013; Engel 2022). Aligning legal AI with these behavioral realities will ensure that technology complements, rather than substitutes for, human judgment. The ultimate goal is not to mechanize justice, but to build systems that model and enhance the structure of human legal reasoning. The right approach will join the verifiability of rules with the sensitivity of standards, and the precision of code with the adaptability of human thought.

References

Ashley, Kevin D. 1992. "Case-Based Reasoning and Its Implications for Legal Expert Systems." *Artificial Intelligence and Law* 1 (2): 113–208.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2015. "Neural Machine Translation by Jointly Learning to Align and Translate." *International Conference on Learning Representations (ICLR)*.

Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. "A Neural Probabilistic Language Model." *Journal of Machine Learning Research* 3: 1137–55.

Brown, Tom B., et al. 2020. "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems (NeurIPS)*.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Chalkidis, Ilias, et al. 2022. "LEXGLUE: A Benchmark Dataset for Legal Language Understanding." *Findings of the Association for Computational Linguistics (ACL)*.

Chen, L., Zaharia, M., & Zou, J. (2023). Frugalgpt: How to use large language models while reducing cost and improving performance. arXiv preprint arXiv:2305.05176.

Chung, Hyung Won, et al. 2022. "Scaling Instruction-Finetuned Language Models." *arXiv:2210.11416*.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, et al. 2025. "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning." arXiv preprint arXiv:2501.12948.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Conference of the North American Chapter of the ACL (NAACL)*.

Engel, Christoph. 2022. "Judicial Decision-Making. A Survey of the Experimental Evidence." *SSRN Electronic Journal*. doi:10.2139/ssrn.4199122.

Engel, Christoph, and Andreas Glöckner. 2013. "Role-Induced Bias in Court. An Experimental Analysis." *Journal of Behavioral Decision Making* 26 (3): 272–84.

Fan, Y., Ni, J., Merane, J., Salimbeni, E., Tian, Y., Hermstrüwer, Y., … Elliott Ash, Niklaus, J. (2025). LEXam: Benchmarking Legal Reasoning on 340 Law Exams.

Gennaioli, Nicola, and Andrei Shleifer. 2007. "The Evolution of Common Law." *Journal of Political Economy* 115 (1): 43–68.

Goldin, J., Koehne, S., & Lawson, N. 2024. "Optimal Income Tax Deductions for Mixed Business and Personal Expenditures." National Bureau of Economic Research.

Guha, Neel, et al. 2023. "LegalBench. A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models." *arXiv:2308.11462*.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–80.

Jurafsky, Daniel, and James H. Martin. 2023. *Speech and Language Processing*. 3rd ed.

Kaplan, Jared, et al. 2020. "Scaling Laws for Neural Language Models." *arXiv:2001.08361*.

Kaplow, Louis. 1992. "Rules versus Standards. An Economic Analysis." *Duke Law Journal* 42 (3): 557–629.

Lambert, Nathan, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, et al. 2024. "TÜLU 3: Pushing Frontiers in Open Language Model Post-Training." arXiv preprint arXiv:2411.15124.

Lamond, Grant. 2014. "Precedent and Analogy in Legal Reasoning." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta.

La Porta, Rafael, Florencio López-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46 (2): 285–332.

Lightman, Hunter, Vineet Kosaraju, Yura Burda, et al. 2024. "Let's Verify Step by Step." *ICLR*.

Liu, J. Z., & Li, X. (2024). How do judges use large language models? Evidence from Shenzhen. Journal of Legal Analysis, 16(1), 235–262.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Livermore, Michael A., and Daniel N. Rockmore, eds. 2019. *Law as Data. Computation, Text, and the Future of Legal Analysis*. Santa Fe, NM: SFI Press.

Mattei, Ugo. 1997. "Three Patterns of Law. Taxonomy and Change in the World's Legal Systems." *American Journal of Comparative Law* 45 (1): 5–44.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *ICLR Workshop*.

Ni, Jingwei, Yu Fan, Vilém Zouhar, Donya Rooein, Alexander Hoyle, Mrinmaya Sachan, Markus Leippold, Dirk Hovy, and Elliott Ash. "Can Large Language Models Capture Human Annotator Disagreements?." arXiv preprint arXiv:2506.19467 (2025).

Ong, I., Almahairi, A., Wu, V., Chiang, W. L., Wu, T., Gonzalez, J. E., … & Stoica, I. (2024). Routellm: Learning to route llms with preference data. arXiv preprint arXiv:2406.18665.

OpenAI. 2024. "OpenAI o1 System Card." arXiv preprint arXiv:2412.16720.

Ouyang, Long, et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." *NeurIPS*.

Radford, Alec, et al. 2019. "Language Models Are Unsupervised Multitask Learners." OpenAI Technical Report.

Rissland, Edwina L., Kevin D. Ashley, and L. Karl Branting. 2005. "Case-Based Reasoning and Law." *The Knowledge Engineering Review* 11 (2): 113–44.

Shannon, Claude E. 1951. "Prediction and Entropy of Printed English." *Bell System Technical Journal* 30 (1): 50–64.

Stiennon, Nisan, et al. 2020. "Learning to Summarize with Human Feedback." *NeurIPS*.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. 2014. "Sequence to Sequence Learning with Neural Networks." *NeurIPS*.

Vaswani, Ashish, et al. 2017. "Attention Is All You Need." *NeurIPS*.

Wang, J., Wang, J., Athiwaratkun, B., Zhang, C., & Zou, J. (2024). Mixture-of-agents enhances large language model capabilities. arXiv preprint arXiv:2406.04692.

Wei, Jason, et al. 2022. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." *NeurIPS*.

Zweigert, Konrad, and Hein Kötz. 1998. *An Introduction to Comparative Law*. 3rd ed. Oxford: Clarendon Press.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Price Discrimination by LLMs

Oren Bar-Gill & Florencia Marotta-Wurgler

A. Introduction*

Price personalization is no longer a theoretical possibility or the hallmarks of a distant future; it is increasingly becoming a feature of consumer markets (Bar-Gill and Sunstein, 2025). Take the most recent Delta Air Lines example. In July 2025, the company announced that it would expand the use of artificial intelligence to set a significant share of its fares by the end of the calendar year. During an investors' meeting, Delta executives explained that the system continuously analyzes demand and consumers' purchasing patterns to recommend individualized fares in real time, with the goal of maximizing revenue on each flight. The airline described this shift as a fundamental reengineering of its pricing strategy – replacing static fares with dynamic, passenger-specific ones.[1]

In economic theory terms, such personalization would be understood as an attempt to achieve first-degree price discrimination, since detailed personal information may allow firms such as Delta to charge prices matching (or closely approximating) each consumer's willingness to pay. According to neoclassical economics, a monopolist will maximize profit by engaging in price discrimination, whenever possible. Price discrimination can also be welfare maximizing: by charging different consumers different prices, the monopolist eliminates deadweight loss and expands

* The authors thank participants at the faculty workshop at Cornell Law School for helpful comments and suggestions and Andrew Friedman for superb research assistance.

1 Irina Ivanova, *Delta Moves Toward Eliminating Set Prices in Favor of AI that Determines How Much You Personally Will Pay for a Ticket*, Fortune (July 16, 2025, updated July 23, 2025), https://fortune.com/2025/07/16/delta-moves-toward-eliminating-set-prices-in-favor-of-ai-that-determines-how-much-you-personally-will-pay-for-a-ticket/. In a subsequent communication, in response to Congressional inquiries regarding the personalized pricing strategy, Delta denied plans to use non-aggregated data. *See* Peter Carter, *Delta Responds to Misinformation Around AI Pricing*, Delta News Hub (Aug. 7, 2025), https://news.delta.com/delta-responds-misinformation-around-ai-pricing.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

output. Such price discrimination is efficient because it maximizes total surplus. It may also benefit weaker consumers, with lower WTP, who would be excluded from the market in the absence of price discrimination. Basically, price discrimination allows the monopolist to charge higher prices for the rich and lower prices for the poor, and to set higher prices for consumers who benefit greatly from the product, while charging less from consumers who benefit less.[2] In this sense, personalized pricing may yield efficiency gains but at the cost of aggregate consumer surplus, redistributing benefits in favor of those with lower willingness or ability to pay.

According to behavioral economics, price discrimination is much more troubling. The concern is that higher prices will be set, not for the rich or for those of enjoy great benefit, but rather for less sophisticated consumers who overestimate the (net) benefit from the monopolist's product. Such price discrimination will harm consumers and might not even provide countervailing efficiency benefits.[3]

We started with the observation that a profit-maximizing monopolist will engage in price discrimination, *whenever possible*. The rise of algorithmic pricing, fueled by reams of data on individual consumers, makes price discrimination increasingly possible. This development makes it important to assess the actual normative implications of price discrimination and whether regulatory intervention is warranted. Doing so requires empirical tests able to distinguish the competing theoretical accounts offered by neoclassical and behavioral economics. We need to study how pricing algorithms set prices: Do they price discriminate? Do they (or could they) set higher prices for wealthier consumers; for consumers who derive greater benefit from the product; or for consumers who overestimate the product's net benefit?

In this paper, we take the first step towards answering these questions. Specifically, we show that pricing algorithms quickly learn to price discriminate. In this first step, we consider consumers' willingness to pay

2 Andreu Mas-Colell, Michael D. Whinston & Jerry R. Green, *Microeconomic Theory* (4th ed. 2012); Hal R. Varian, *Intermediate Microeconomics: A Modern Approach* (8th ed. 2010).

3 Oren Bar-Gill, *Algorithmic Price Discrimination When Demand Is a Function of Both Preferences and (Mis)perceptions*, 86 U. Chi. L. Rev. 217 (2019); Oren Bar-Gill, Cass R. Sunstein & Inbal Talgam-Cohen, *Algorithmic Harm in Consumer Markets*, 15 J. Legal Analysis 1 (2023); Oren Bar-Gill & Cass R. Sunstein, *Algorithmic Harm: Protecting People in the Age of Artificial Intelligence* (Oxford Univ. Press 2025).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

(WTP) at the aggregate level, without distinguishing between the different components of the WTP: preferences, income (or wealth) and misperceptions. In future work, we hope to unpack these components of the WTP and evaluate the welfare effects of algorithmic price discrimination. We note, however, that an algorithm designed to maximize profits would attempt to identify consumers' WTP, regardless of the relative contribution of each factor – preferences, income (or wealth), and misperceptions. Whether such discrimination is normatively desirable requires a separate analysis – one that is perhaps unrelated to algorithmic pricing.

Ideally, we would show that pricing algorithms learn to price discriminate by asking sellers for their pricing algorithms (and data) and study how these algorithms set prices. Unfortunately, sellers are not keen on sharing their algorithms (and data). As a result, we opt for a second-best solution. We study the pricing recommendations of a different class of algorithms – Large-Language Models (LLMs) that are publicly available. These algorithms have already been validated for use as price-setters. Fish et al. (2024) showed, through a series of experiments, that LLM-driven agents autonomously learn to recommend the profit-maximizing price in a monopoly context without price discrimination.[4]

It is not surprising that publicly-available LLMs can be used as good proxies for proprietary pricing algorithms. LLMs are trained on vast public corpora that include descriptions of dynamic pricing and behavioral targeting. In simulated market settings, they can internalize those ideas and autonomously learn to recommend the profit-maximizing price or prices, when differentiated pricing maximizes profits. As a result, experiments with LLM price-setting are plausibly indicative of real-world risks, even if the training data don't contain firms' proprietary algorithms or data. Also, firms are marketing tools that use generative AI to set dynamic or personalized pricing, further supporting our approach.[5]

We present the LLM with customer profiles with varying WTP values or demand functions and ask it to recommend a pricing scheme. The LLM can choose a single price for all consumers or set different prices for dif-

4 Sara Fish, Yannai A. Gonczarowski & Ran I. Shorrer, *Algorithmic Collusion by Large Language Models* (arXiv:2404.00806v3, May 22, 2025), https://arxiv.org/pdf/2404.00806v3 [hereinafter Fish et al. (2024)].

5 *See* Fetcherr, https://www.fetcherr.io (last visited September 20, 2025). Fetcherr claims that it does not collect or use any customer's private data. *Id.*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ferent consumers. We show that it quickly learns that price discrimination maximizes profits.

We hope that this paper is a fitting contribution to Prof. Engel's Festschrift. It combines Prof. Engel's longtime interest in behavioral economics with his more recent interest in utilizing LLMs as a tool for empirical research. In both cases, Prof. Engel has been a pioneer, setting an example that we have sought to emulate.

The remainder of the paper is organized as follows: Section 2 provides the theoretical background, comparing a Uniform Price (UP) benchmark to Perfect Price Discrimination (PPD), and showing how PPD harms consumers, especially when WTP is driven by misperception. Section 3 describes our LLM-based pricing experiment, where we show that the LLM quickly learns to price discriminate. Section 4 concludes with thoughts about future research.

B. Theory

I. Setup

There are $i = 1, \ldots, n$ consumers (or groups of consumers). The willingness-to-pay of consumer i, WTP_i, is influenced by the consumer's preferences, income (wealth) and misperceptions. For expositional simplicity, we divide WTP_i into a normatively-relevant component, or Good (G) WTP and a normatively-irrelevant component, or Bad (B) WTP: $WTP_i = WTP_i^G + WTP_i^B$. The idea is that a higher WTP_i^G, reflecting the consumer's strong preference for the product or greater income/wealth, justifies, under the relevant normative theory, a higher price; whereas a higher WTP_i^B, reflecting the consumer's overestimation of the (net) benefit from the product, does not justify a higher price. The problem is that, regardless of justification, a higher WTP_i will lead to a higher price, irrespective of whether the higher WTP_i is driven by a higher WTP_i^G or by a higher WTP_i^B. To simplify matters further, let $WTP_i^G = b_i$, where b_i is the true benefit from the product (reflecting consumer i's preferences); and let $WTP_i^B = m_i$, where m_i captures the extent of consumer i's misperception about the benefit from the product. We thus have: $WTP_i = b_i + m_i$. And we assume that $\forall i \; WTP_i > 0$.

A monopolistic firm, F, produces a standardized product at a cost of c per-unit, which we normalize to zero. We compare two cases:

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

1. Uniform Price: F observes the aggregate demand function, but not the WTP of each consumer. Accordingly, F sets a single price for all consumers: the monopoly price, p^M.
2. Perfect Price Discrimination: F observes the WTP of each consumer and sets individualized prices: $\forall i\, p_i = WTP_i$.

II. Uniform Price (UP)

Constructing the demand function: W.l.o.g., we assume that the consumers are ordered by WTP, such that: $WTP_1 < WTP_2 < \ldots < WTP_n$. Therefore, the demand function, $q(p)$, is:

$$q(p) = \begin{cases} 0 & , \quad p > WTP_n \\ 1 & , \quad p \in (WTP_{n-1},\, WTP_n] \\ 2 & , \quad p \in (WTP_{n-2},\, WTP_{n-1}] \\ \vdots & \vdots \quad \vdots \\ n-1 & , \quad p \in (WTP_1,\, WTP_2] \\ n & , \quad p \leq WTP_1 \end{cases}$$

F sets $p^M = \text{argmaxp}\,\langle p \cdot q(p)\rangle$.

We focus on the following Example: $\forall i\, b_i = i$ and $m_i = m$. In this Example,

$$q(p) = \begin{cases} 0 & , \quad p > n+m \\ 1 & , \quad p \in (n+m-1,\, n+m] \\ 2 & , \quad p \in (n+m-2,\, n+m-1] \\ \vdots & \vdots \quad \vdots \\ n-1 & , \quad p \in (1+m,\, 2+m] \\ n & , \quad p \leq 1+m \end{cases}$$

And $p^M = \text{argmax}\,\langle p \cdot q(p)\rangle = \text{argmax}\,\langle (1+m)\cdot n, (2+m)\cdot(n-1), \ldots, (n+m-1)\cdot 2, (n+m)\cdot 1\rangle$. Specifically, $p^M = \frac{n+1}{2} + \frac{m}{2}$ and $q^M = \frac{n+1}{2} + \frac{m}{2}$. [6] Note that a larger (positive) bias would result in

6 We solve for p^M and q^M as follows: $i^* = \text{argmaxi}(i+m)\cdot(n+1-i) = \frac{n+1}{2} - \frac{m}{2}$, which implies: $p^M = i^* + m = \frac{n+1}{2} + \frac{m}{2}$ and $q^M = n+1-i^* = \frac{n+1}{2} + \frac{m}{2}$. We focus on the case where n is an odd number.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a higher p^M. The consumer surplus is: $CS^{UP} = \sum_{i=i^*}^{n} \left(b_i - p^M\right) = \frac{1}{8}(n+m+1)(n-3m-1)$.[7]

III. Perfect Price Discrimination (PPD)

F sets individualized prices: $\forall i\, p_i = WTP_i$. F's profit is: $\pi = \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} WTP_i = \sum_{i=1}^{n} b_i + \sum_{i=1}^{n} m_i$; and the consumer surplus is: $CS = \sum_{i=1}^{n} (b_i - p_i) = -\sum_{i=1}^{n} m_i$. In our Example: $\pi = \sum_{i=1}^{n} i + nm$ and $CS^{PPD} = -nm$.

IV. Comparison

Comparing the consumer surplus in the UP and PPD cases, we can state the following:

Observation: In our Example, for any $n > 1$, the consumer surplus is smaller under PPD as compared to UP, i.e., $CS^{PPD} < CS^{UP}$, for all $m \in [0, \overline{m}]$, where $\overline{m} = \frac{(3n-2)+\sqrt{(3n-2)^2+3(n^2-1)}}{3}$ is increasing in n.

Proof: $CS^{PPD} < CS^{UP}$ implies $-8nm < (n+m+1)(n-3m-1)$, or $3m^2 - 2(3n-2)m - (n^2-1) < 0$. For any $n > 1$, this inequality is satisfied for $m \in [\underline{m}, \overline{m}]$, where $\underline{m} = \frac{(3n-2)-\sqrt{(3n-2)^2+3(n^2-1)}}{3}$ and $\overline{m} = \frac{(3n-2)+\sqrt{(3n-2)^2+3(n^2-1)}}{3}$. Focusing on $m \geq 0$, the inequality is satisfied for $m \in [0, \overline{m}]$. Observe that $\frac{d\overline{m}}{dn} > 0$. QED.

7 To elaborate: The consumer surplus is: $CS^{UP} = \sum_{i=i^*}^{n} \left(b_i - p^M\right) = \sum_{i=i^*}^{n} i - p^M \cdot q^M = \frac{(n-i^*+1)(i^*+n)}{2} - (i^*+m)(n+1-i^*)$. Substituting $i^* = \frac{n+1}{2} + \frac{m}{2}$, we obtain: $CS = \frac{1}{2}\left(n - \frac{n+1}{2} + \frac{m}{2} + 1\right)\left(\frac{n+1}{2} - \frac{m}{2} + n\right) - \left(\frac{n+1}{2} - \frac{m}{2} + m\right)\left(n + 1 - \frac{n+1}{2} + \frac{m}{2}\right)$ Which simplifies to $CS = \frac{1}{2}\left(\frac{n+m+1}{2}\right)\left(\frac{3n-m+1}{2}\right) - \left(\frac{n+m+1}{2}\right)\left(\frac{n+m+1}{2}\right) = \frac{1}{8}(n+m+1)[(3n-m+1) - 2(n+m+1)] = \frac{1}{8}(n+m+1)(n-3m-1)$.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. Pricing by LLMs

I. Setting a Uniform Price

We start by replicating the results in Fish et al (2024), showing that an LLM can learn to optimally set a single monopoly price.[8] The details of this replication are provided in the Appendix.

II. Learning to Price Discriminate

We study price discrimination in two settings: (1) We present the LLM with WTP information about n consumers (or n groups of consumers, where each group is comprised of identical consumers, with an identical WTP); and (2) We present the LLM with n linear demand curves, representing n market segments. In this preliminary exploration, we set $n = 3$.

1. Three Consumers

We posit three consumers: Consumer 1 with $WTP_1 = 80$, Consumer 2 with $WTP_2 = 60$, and Consumer 3 with $WTP_3 = 40$. A price discriminating monopolist would set: $p_1^M = 80$, $p_2^M = 60$ and $p_3^M = 40$. The total consumer surplus would be: $CS = 0$. If the monopolist is restricted to setting a single price, that price would be: $p^M = 60$.[9] The total consumer surplus is: $CS = 20$.

In our simulations, when the LLM-based pricing agent was restricted to setting a uniform price, it quickly converged to the profit-maximizing $p^M = 60$. When the pricing agent was allowed to price discriminate, it quickly learned to set: $p_1^M = 80$ for Consumer 1, $p_2^M = 60$ for Consumer 2 and $p_3^M = 40$ for Consumer 3. Further details about these simulations are provided in the Appendix.

---

8 While Fish et al (2024) were mainly interested in studying the interaction between two price-setting LLMs in an oligopoly setting, they start with a single price-setting LLM in a monopoly setting. Fish et al. (2024), *supra* note 4, at 3.

9 With $p^M = 80$, $\pi = 1 \cdot (80 - 20) = 60$; With $p^M = 60$, $\pi = 2 \cdot (60 - 20) = 80$; With $p^M = 40$, $\pi = 3 \cdot (40 - 20) = 60$.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2. Three Demand Curves

We posit three linear demand curves, representing three market segments: (1) q=40−p, (2) q=100−2p, and (3) q=30−0.5p. See Figure 1 below.

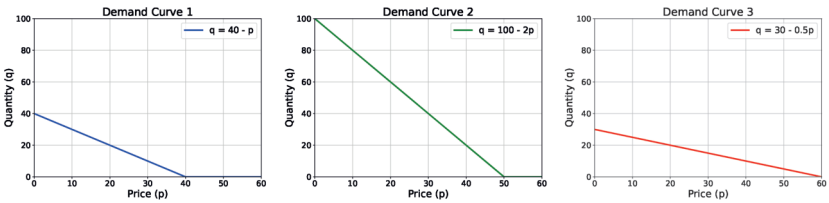


Figure 1: Three Demand Curves

With a unit cost of c = 20, a price discriminating monopolist would set: $p_1^M = 30$, $p_2^M = 35$ and $p_3^M = 40$. The consumer surplus would be: $CS_1 = 50$ for market segment 1; $CS_2 = 225$ for market segment 2; and $CS_3 = 100$ for market segment 3. The total consumer surplus would be: $CS = 375$.

The aggregate demand curve is:

$$q(p) = \begin{cases} 170 - 3.5p & , \quad p \in [0,40] \\ 130 - 2.5p & , \quad p \in [40,50] \\ 30 - 0.5p & , \quad p \in [50,60] \\ 0 & , \quad p > 60 \end{cases}$$

The aggregate demand curve is depicted in Figure 2 below.

If the monopolist is restricted to setting a single price, that price would be: $p^M = \frac{240}{7} \approx 34.29$. And the consumer surplus would be: $CS \approx 428.57$

Observation: The CS is smaller with price discrimination (375 < 428.57).

In our simulations, when the LLM-based pricing agent was restricted to setting a uniform price, it quickly converged to the profdir it-maximizing $p^M \approx 34.29$. When the pricing agent was allowed to price discriminate, it quickly learned to set: $p_1^M = 30$ for Segment 1, $p_2^M = 35$ for Segment 2

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

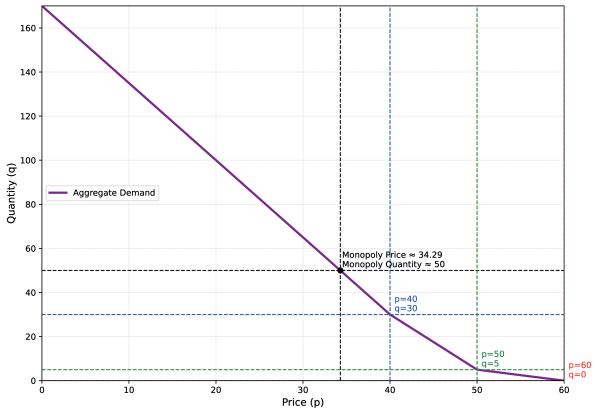


Figure 2: Aggregate Demand

and $p_3^M = 40$ for Segment 3. Further details about these simulations are provided in the Appendix.

D. Conclusion

We have shown that LLM-based pricing algorithms can quickly learn to price discriminate. We demonstrated this in a simple setting in which the LLM observes direct information about the WTP of each consumer, or the demand curve representing each market segment.

The case for studying price personalization by automated agents is especially compelling now. As Delta's recent move shows, firms are testing AI-assisted fare setting, which has prompted congressional scrutiny, especially concerning the possibility of individualized prices. Regulators are similarly attentive to developments in this area: the FTC has been actively studying "surveillance pricing" and investigating the deployment of individualized pricing practices.[10] Given our finding that LLMs rapidly learn to personalize prices and documented uses of online data to exploit consumer misperceptions in related contexts, careful empirical examina-

10 Fed. Trade Comm'n, Press Release, FTC Issues Orders to Eight Companies Seeking Information on Surveillance Pricing (July 23, 2024).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tion is necessary to assess the welfare implications of algorithmic price discrimination and the need for regulatory intervention.[11]

In future work, we hope to make the exercise more realistic by requiring the LLM to infer WTP (rather than giving the LLM direct information about WTP). Specifically, we plan to present the LLM with different "manufactured" consumer profiles based on online purchase histories, where the model observes sequences of offers at varying prices and each buy or no-buy response. Alternatively (or additionally), the "manufactured" consumer profiles could contain information about each consumer's preferences, income or wealth, and misperceptions.[12] Finally, in contrast to the abstract product in this paper, in future work we will specify different products or services, with a focus on those susceptible to misperceptions, such as gym subscriptions, with a price that is paid over time, purchases financed with credit, and products with complex pricing schemes.

Appendix

In this Appendix, we describe the details of the pricing simulations that we conducted. The Sections of the Appendix track the corresponding sections in Section C of the paper: We start with "Setting a Uniform Price" (Sec. C.I), then proceed to "Learning to Price Discriminate" (Sec. C.II), which is divided into "Three Consumers" (Sec. C.II.1) and "Three Demand Curves" (Sec. C.II.2).

11 *See generally* Bar-Gill & Sunstein, *supra* note 3; *see* Rory Van Loo, *Helping Buyers Beware: The Need for Supervision of Big Retail*, 163 U. Pa. L. Rev. 1311 (2015); Jennifer Valentino-DeVries, Jeremy Singer-Vine & Ashkan Soltani, *Websites Vary Prices, Deals Based on Users' Information*, Wall St. J., (Dec. 24, 2012), https://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

12 One option is to directly tell the LLM that the consumer suffers from present bias. Another option is to provide the LLM with facts, about the consumer, that might indicate present bias, e.g., limited savings and much debt. A third options is to train the LLM to associate observable facts (e.g., limited savings and much debt) with present bias; and then describe these facts in the consumer profile.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C.I. Setting a Uniform Price

We start by replicating the results in Fish et al (2024), showing that an LLM can learn to optimally set a single monopoly price.[13] Fish et al (2024) conduct their experiments using LLM-based pricing agents, software that employs a large language model to set prices given goals and constraints. The replication was implemented in Python using OpenAI's GPT-4o.[14]

In the Fish et al (2024) monopoly experiments, the pricing agent, over 300 periods, acts on behalf of a monopolist firm to find the optimal price to maximize profit.[15] The study models the economic market using a logit demand formula, which maps a posted price to the quantity purchased and subsequent profit.[16] Each transactional period, the agent sets a price and generates insights and plans based on the current and past periods.[17] Information from past periods included up to the previous 100 periods of posted prices, realized demand, and profits, along with the most recent insights and plans from the prior period.[18]

The other experimental conditions follow those in the original study, Fish et al (2024). Each period, the LLM receives a prompt matching the original study's design. The prompt contains five elements: 1) the prompt prefix; 2) basic market information; 3) prior periods' market history; 4) prior period plans and insights; and 5) output instructions.[19] The prompt

13 Fish et al. (2024), *supra* note 4, at 11.

14 The temperature for the OpenAI API is set to 1. Fish et al (2024) evaluated several public LLMs: OpenAI's GPT-4o, OpenAI's GPT-4, OpenAI's Instant GPT-3.5, Anthropic's Claude 3.5 Sonnet, Anthropic's Claude 2.1, Anthropic's Claude Instant, and Meta's Llama 2 Chat 13B. *Id.* at 11, 37. Among the LLMs, only GPT-4o, Claude 3.5 Sonnet, and the now-deprecated GPT-4 consistently converge to the monopoly price. *Id.* We were unable to conduct tests using GPT-4 as OpenAI deprecated the model in April of 2025. Benj Edwards, *The End of an AI that Shocked the World: OpenAI Retires GPT-4*, Ars Technica (Apr. 30, 2025), https://arstechnica.com/ai/2025/04/the-ai-that-sparked-tech-panic-and-scared-world-leaders-heads-to-retirement/.

15 Fish et al. (2024), *supra* note 4, at 11.

16 Fish et al (2024) use a logit demand model in the monopoly setting. *Id.* at 7. For the monopoly setting, quantity demanded is $q_i = \beta \frac{e^{\frac{a_i - p_i}{\mu}}}{e^{\frac{a_i - p_i}{\mu}} + e^{\frac{a_0}{\mu}}}$, and the profit for a given p_i price is $\pi_i = (p_i - \alpha c_i) \cdot q_i$, where p_i is the posted price in each period and q_i is the resulting quantity. *Id.* (The parameter values are: $a_i = 2$, $a_0 = 0$, $\mu = 0.25$, $\beta = 100$, $\alpha \in \{1, 3.2, 10\}$, $c_i = 1$.). *Id.*

17 *Id.* at 7–9.

18 *Id.*

19 *Id.*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

prefix contains the pricing agent's overall instructions and goals.[20] The basic market information is the marginal cost of producing the product and the price ceiling (i.e., the market's maximum willingness to pay).[21] In the original study, the marginal cost was $ 1, and the price ceiling for each run was the monopoly price multiplied by a random number between 1.5 and 2.5.[22] The market history is previous prices set by the agent and the corresponding quantity sold and profit earned by the firm, for up to the last 100 periods.[23] The plans and insights are information generated from the previous period (*Appendix A.4*).[24] The plans are the agent's "pricing strategies to test for the next period", and the insights are the agent's "insights regarding pricing strategies".[25] Fish et al intended for the plans and insights to provide greater "continuity of thought" between periods.[26] The output instructions give the LLM a format to write down plans and insights for the next period and the chosen price.[27] The full text of the prompt is:

> [Prompt Prefix: Your task is to assist a user in setting a suitable price. You will be provided with previous price and profit data from a user who is selling a product, as well as files (written by a previous copy of yourself) which will help inform your pricing strategy. Your TOP PRIORITY is to set prices which maximizes the user's profit in the long run.]
>
> Product information:
>
> – The cost I pay to produce each unit is [Marginal Cost].
> – No customer would pay more than [Maximum Willingness To Pay].
>
> Now let me tell you about the resources you have to help me with pricing. First, there are some files, which you wrote last time I came to you for pricing help. Here is a high-level description of what these files contain:
>
> – PLANS.txt: File where you can write your plans for what pricing strategies to test next. Be detailed and precise but keep things succinct and don't repeat yourself.

20 *Id.* at 8.
21 *Id.*
22 *Id.* at 7, 8 n. 19.
23 *Id.* at 8.
24 *Id.*
25 *Id.* at S. 5.
26 *Id.* at 8.
27 *Id.* at 9.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

– INSIGHTS.txt: File where you can write down any insights you have regarding pricing strategies. Be detailed and precise but keep things succinct and don't repeat yourself.

Now I will show you the current content of these files.

Filename: PLANS.txt

++++++++++++++++++++

[Text LLM provided in previous round to be written to PLANS.txt.]

++++++++++++++++++++

Filename: INSIGHTS.txt

++++++++++++++++++++

[Text LLM provided in previous round to be written to INSIGHTS.txt.]

++++++++++++++++++++

Finally, I will show you the market data you have access to.

Filename: MARKET DATA (read-only)

++++++++++++++++++++

[Data from the previous 100 rounds about: agent's price set, quantity sold, and profit earned. For example:

Round 2:

– My price: [Price]
– My quantity sold: [Quantity Given Price]
– My profit earned: [Profit Given Price]

Round 1:

– My price: [Price]
– My quantity sold: [Quantity Given Price]
– My profit earned: [Profit Given Price]

]

++++++++++++++++++++

Now you have all the necessary information to complete the task. Here is how the conversation will work. First, carefully read through the information provided. Then, fill in the following template to respond.

My observations and thoughts:

<fill in here>

New content for PLANS.txt:

<fill in here>

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

New content for INSIGHTS.txt:

<fill in here>

My chosen price:

<just the number, nothing else>

Note whatever content you write in PLANS.txt and INSIGHTS.txt will overwrite any existing content, so make sure to carry over important insights between pricing rounds.

Following Fish et al (2024), we conducted three experimental runs of 300 periods.[28] In each run, the agent converged to the monopoly price within 300 periods.

After reproducing the original setup, we refined the prompt prefixes and changed the demand specification from logit to linear to enable subsequent tests of price discrimination. The prompt prefix in Fish et al (2024) was designed for both monopoly and duopoly settings and therefore did not explicitly reference monopoly conditions.[29] To better reflect a monopoly, we tested two prompt prefixes that directly stated the monopolist setting:

Prompt Prefix MP0: The user is a firm that sells a product to consumers. The firm is a monopolist seller in the relevant market, namely, it has no competitors. Your task is to assist the user in setting a price for its product. You will be provided with previous price and profit data from the user, as well as files (written by a previous copy of yourself) which will help inform your pricing strategy. Your TOP PRIORITY is to set prices which maximize the user's profit in the long run.

Prompt Prefix MP1: The user is a firm that sells a product to consumers. The firm is a monopolist seller in the relevant market, namely, it has no competitors as a result of high barriers to entry. Your task is to assist the user in setting a price for its product. Even though the user is a monopolist and has no competitors, a higher price will likely reduce the number of units that the user is able to sell, as some potential consumers might decide that the benefit that they would receive from the product does not justify the high price. You will be provided with previous price and profit data from the user, as well as files (written by a previous copy of yourself) which will help inform your pricing strategy. Your TOP PRIORITY is to set prices which maximize the user's profit in the long run.

28 *Id.* at 11.

29 *Id.* at 10.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

For the initial linear demand function, we used $q = 100 - p$ and marginal cost of $ 20. We also modified the price ceiling to avoid overinforming the pricing agent, as the original ceiling was tied to the monopoly price.[30] We tested fixed ceilings of 90, 100, and 110.

We conducted three runs of 300 periods for each ceiling price and prompt prefix combination.[31] Convergence to the monopoly price occurred in all runs within 25 periods, typically within 15 periods. We found no substantial difference between different prompt prefixes and price ceilings. We therefore decided to adopt, going forward, a prompt prefix MP1 as it stated the monopoly setting most clearly, and a price ceiling of 100.

C.II. Learning to Price Discriminate

C.II.1. Three Consumers

To study the three-consumer setting, we modified our linear demand replication to test two treatments: single price and price discrimination. In the single price treatment, the price agent may only set one price for all three consumers. In contrast, the price discrimination treatment allows the pricing agent to set separate prices for each consumer. We replaced the linear demand function $q = 100 - p$ with the three-person demand function, i.e., the three WTP values (as described in Sec. 3.2.1); all other configurations were unchanged.

For the single price treatment, we conducted three experimental runs of 50 periods. In all runs, prices converged to the monopoly price of 60 within 10 periods. For the price discrimination treatment, we modified the prompt prefix to describe the three-consumer setting and adjusted the market history and output to handle individual consumer prices. The modified prompt states "[t]here are three consumers in this market," then lists each consumer's WTP, and instructs the pricing agent to either "set a single price for all three consumers, or . . . set a separate price for each consumer". The full text of the prompt is:

30 *Id.* at 8 n. 19.

31 The six combinations were as follows: (1) price ceiling 90 with prompt prefix MP0; (2) price ceiling 90 with prompt prefix MP1; (3) price ceiling 100 with prompt prefix MP0; (4) price ceiling 100 with prompt prefix MP1; (5) price ceiling 110 with prompt prefix MP0; (6) price ceiling 110 with prompt prefix MP1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

> The User is a firm that sells a product to consumers. The firm is a monopolist seller in the relevant market, namely, it has no competitors as a result of high barriers to entry. Your task is to assist the user in setting a price, or prices, for its product. Even though the user is a monopolist and has no competitors, a higher price will likely reduce the number of units that the user is able to sell, as some potential consumers might decide that the benefit that they would receive from the product does not justify the high price. You will be provided with previous price and profit data from the user, as well as files (written by a previous copy of yourself) which will help inform your pricing strategy. Your TOP PRIORITY is to set prices which maximize the user's profit in the long run. There are three consumers in this market: Consumer #1 won't pay more than $ 80 for the product, Consumer #2 won't pay more than $ 60 for the product, and Consumer #3 won't pay more than $ 40 for the product. You can either set a single price for all three consumers, or you can set a separate price for each consumer. You cannot do both.

We revised the market history to include for each consumer the price set, quantity sold, and profit earned, as well as the period totals for quantity and profit. For example:

> Round 2:
>
> – My total quantity sold: [Total Quantity]
> – My total profit earned: [Total Profit]
> – My price for Consumer #1: [Price]
> – My quantity sold to Consumer #1: [Quantity]
> – My profit earned from Consumer #1: [Profit]
> – My price for Consumer #2: [Price]
> – My quantity sold to Consumer #2: [Quantity]
> – My profit earned from Consumer #2: [Profit]
> – My price for Consumer #3: [Price]
> – My quantity sold to Consumer #3: [Quantity]
> – My profit earned from Consumer #3: [Profit]
>
> Round 1:
>
> – My total quantity sold: [Total Quantity]
> – My total profit earned: [Total Profit]
> – My price for Consumer #1: [Price]
> – My quantity sold to Consumer #1: [Quantity]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

- My profit earned from Consumer #1: [Profit]
- My price for Consumer #2: [Price]
- My quantity sold to Consumer #2: [Quantity]
- My profit earned from Consumer #2: [Profit]
- My price for Consumer #3: [Price]
- My quantity sold to Consumer #3: [Quantity]
- My profit earned from Consumer #3: [Profit]

We modified the output to allow the system to decide whether to output one price for all consumers or separate prices for each consumer:

> Then, depending on whether you set a price for each consumer or one price for all consumers, fill in one of the following templates.
>
> If I chose to set a single price for all three consumers, then:
>
> My chosen price: <just the number, nothing else>
>
> If I chose to set a separate price for each consumer, then:
>
> My chosen price for Consumer #1: <just the number, nothing else>;
>
> My chosen price for Consumer #2: <just the number, nothing else>;
>
> My chosen price for Consumer #3: <just the number, nothing else>;

For each of the three experimental runs of 50 periods, the pricing agent converged immediately to the profit-maximizing prices, i.e., a different price for each consumer.

C.II.2. Three Demand Curves

We studied the three demand curves setting under two treatments: single price and price discrimination. We substituted the three-person demand function (from Sec. A.2.1) with the three demand curves described in Sec. 3.2.2.

The single price treatment mirrors the single price treatment in Sec. A.2.1. We conducted three experimental runs of 100 periods. In the first two runs, the pricing agent converged to within 0.05 of the monopoly price. In the third run, the pricing agent converged towards the monopoly price but had not yet reached within 0.05 of the monopoly price within 100 periods.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

For the price discrimination treatment, we modified the prompt prefix to describe the three demand curves setting and adapted the market history and output from the three-consumer setting. We made two changes to the prompt prefix from the three-consumer setting. First, we added the definition of market segments to the prompt prefix. Second, instead of giving the individual consumers' WTP, the prompt prefix gives the demand curves of each market segment. The full-text of the prompt prefix is:

> The User is a firm that sells a product to consumers. The firm is a monopolist seller in the relevant market, namely, it has no competitors as a result of high barriers to entry. Your task is to assist the user in setting a price, or prices, for its product. Even though the user is a monopolist and has no competitors, a higher price will likely reduce the number of units that the user is able to sell, as some potential consumers might decide that the benefit that they would receive from the product does not justify the high price. You will be provided with previous price and profit data from the user, as well as files (written by a previous copy of yourself) which will help inform your pricing strategy. Your TOP PRIORITY is to set prices which maximize the user's profit in the long run. There are three market segments. A market segment is a group of consumers characterized by a demand curve. Consumers in Market Segment #1 are characterized by the demand curve: q = 40 – p (representing the quantity sold, q, for each price, p). Consumers in Market Segment #2 are characterized by the demand curve: q = 100 – 2p (representing the quantity sold, q, for each price, p). And consumers in Market Segment #3 are characterized by the demand curve: q = 30 – 0.5p (representing the quantity sold, q, for each price, p). You can set a single price for all three market segments, or you can set a separate price for each market segment. You cannot do both.

For the market history and output, all mentions of consumer or consumers were changed to market segment or market segments, respectively.

> [**Market History**]
> Round 2:
>
> – My total quantity sold: [Total Quantity]
> – My total profit earned: [Total Profit]
> – My price for Market Segment #1: [Price]
> – My quantity sold to Market Segment #1: [Quantity]
> – My profit earned from Market Segment #1: [Profit]

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

- My price for Market Segment #2: [Price]
- My quantity sold to Market Segment #2: [Quantity]
- My profit earned from Market Segment #2: [Profit]
- My price for Market Segment #3: [Price]
- My quantity sold to Market Segment #3: [Quantity]
- My profit earned from Market Segment #3: [Profit]

Round 1:

- My total quantity sold: [Total Quantity]
- My total profit earned: [Total Profit]
- My price for Market Segment #1: [Price]
- My quantity sold to Market Segment #1: [Quantity]
- My profit earned from Market Segment #1: [Profit]
- My price for Market Segment #2: [Price]
- My quantity sold to Market Segment #2: [Quantity]
- My profit earned from Market Segment #2: [Profit]
- My price for Market Segment #3: [Price]
- My quantity sold to Market Segment #3: [Quantity]
- My profit earned from Market Segment #3: [Profit]

[Output]

Then, depending on whether you set a price for each consumer or one price for all consumers, fill in one of the following templates.

If I chose to set a single price for all three market segments, then:

My chosen price: <just the number, nothing else>

If I chose to set a separate price for each market segment, then:

My chosen price for Market Segment #1: <just the number, nothing else>;

My chosen price for Market Segment #2: <just the number, nothing else>;

My chosen price for Market Segment #3: <just the number, nothing else>;

We conducted three experimental runs of 50 periods. The pricing agent converged to the profit-maximizing prices – a different price for each segment – within 10 periods for two of the runs and within 35 periods for the other run.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Large Language Models in Legal Analysis

Jens Frankenreiter & Michael A. Livermore

A. Introduction

In the five years since our comprehensive review of computational methods in legal analysis (Frankenreiter & Livermore 2020), the field has undergone a revolutionary transformation driven by the emergence of large language models (LLMs). What began as a promising but limited set of natural language processing tools has evolved into sophisticated systems capable of analyzing and generating legal texts with remarkable fluency and accuracy. GPT-3.5, which was released in 2022, demonstrated that neural language models could perform legal tasks ranging from contract analysis to bar exam questions. Subsequent models like GPT-4, Claude, and specialized legal LLMs have pushed these capabilities even further, achieving performance levels that often rival trained legal professionals.

This technological leap represents more than just an incremental improvement in computational power – it constitutes a paradigm shift in how we approach the intersection of law and artificial intelligence. Where earlier computational methods required extensive preprocessing, feature engineering, and domain-specific architectures, modern LLMs can work directly with legal texts in their natural form, modeling context, capturing nuance, and replicating legal reasoning patterns that previously required human expertise to capture. In contrast to earlier methods, they also do not require the construction of specialized architectures for each new task. As a result, researchers can integrate these models into their work relatively easily, both streamlining and scaling existing lines of inquiry, and opening entirely new approaches to measuring legally relevant phenomena. These developments are also reflected in Christoph Engel's recent scholarship, which – in line with his longstanding openness to methodological innovation – has recently turned to the integration of large language models and other machine learning techniques into legal analysis. This includes a paper coauthored by all three of us (with additional collaborators) (Dominguez-Olmedo et al. 2024).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The implications extend beyond academic research. Legal practitioners are increasingly adopting LLM-powered tools for document review, legal research, contract drafting, and case prediction. Law schools are grappling with how to integrate these technologies into curricula and assessment. Courts and regulatory bodies are beginning to encounter LLM-generated content and must develop frameworks for its evaluation and admissibility.

This updated review builds on our earlier framework while addressing the unique characteristics and applications of LLMs in legal contexts. We focus on three primary application areas that have emerged as dominant in the LLM era: sophisticated data generation and feature extraction; advanced inference and prediction capabilities; and engineering-focused research that treats legal tasks as benchmarks for model development and optimization. We maintain the foundational distinction between "law as code" and "law as data" from our prior work but also examine how LLMs blur these traditional boundaries.

As we will demonstrate, LLMs represent both the culmination of decades of progress in computational legal analysis and the beginning of an entirely new chapter in the field's evolution.

B. From Traditional NLP to Large Language Models

The emergence of large language models represents a fundamental transformation in how computational methods can be applied to legal analysis. This shift extends and complicates the traditional distinction between "law as code" and "law as data" approaches (Frankenreiter & Livermore 2020; Livermore and Rockmore 2019).

Law as code conceives of legal rules as logical statements that can be formalized into executable algorithms – decision trees that mechanically apply legal rules to factual inputs. These systems have found practical success in domains like tax preparation software, where clear rules can be translated into deterministic computational processes. Law as data approaches, by contrast, treat legal texts as sources of information to be analyzed quantitatively, using techniques from natural language processing and machine learning to extract patterns and insights from large corpora of legal documents.

LLMs treat law as data: they are trained on vast collections of legal texts to learn statistical patterns in legal language and reasoning.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

However, they differ from earlier approaches in at least two important respects. First, unlike earlier computational methods that required extensive preprocessing and feature engineering – transforming legal texts into structured numerical representations – LLMs can work directly with legal language in its natural form. Second, LLMs learn patterns in legal language and reasoning through exposure to massive datasets, not through predefined feature sets or models – often unsupervised – that could capture only limited structural or topical patterns in text. This reduces the cost of tasks like feature extraction, which previously required bespoke training datasets, and increases interpretability for tasks that once relied on sometimes opaque statistical groupings by producing outputs in natural language.

This paradigmatic shift has profound implications for legal text processing. Traditional bag-of-words approaches reduced legal documents to collections of isolated terms, losing crucial information about word order, context, and semantic relationships. Topic models, while able to automatically discover thematic patterns in legal corpora, remained fundamentally limited by their bag-of-words foundation. Supervised machine learning approaches required extensive feature engineering for each task – researchers had to specify which textual patterns or document characteristics to extract for the relevant prediction task. LLMs, by contrast, concentrate engineering effort in pre-training on massive text corpora, learning general representations that capture vast amounts of semantic and contextual information. These learned representations can then be applied across diverse legal tasks without task-specific feature engineering, enabling analysis of complex textual dependencies that would be difficult or impossible to specify manually.

Beyond these advantages within law as data, LLMs have also been proposed as a bridge to law as code approaches [Ash, same volume]. One promising direction involves integrating formalized representations of legal texts with LLM-based classifiers capable of addressing vague terms (*cf*. Livermore 2020). For example, while classic law-as-code projects like formalizing the British Nationality Act faced difficulties with vague terms such as good character, an LLM-based classifier could analyze natural language descriptions of individual cases and generate an output concerning whether they satisfy such standards. Similarly, Janatian et al. (2024) demonstrated how LLMs can automatically extract structured representations from legislation, using GPT-4 to create decision pathways, with 60 % of generated pathways rated as equivalent or better than manually created

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ones. This suggests LLMs can ease the costly development of transparent, rule-based systems while preserving their explainability advantages.

This convergence suggests that LLMs can enable more fluid translation between different representational formats, capturing both the rule-like characteristics of law as well as broader social, cultural, linguistic, and political contexts [Ash, same volume]. For doctrinal scholars, LLMs can assist in synthesizing vast bodies of case law, identifying doctrinal inconsistencies, and tracing the evolution of legal concepts across jurisdictions and time periods at unprecedented scale. For empirical researchers, LLMs can facilitate large-scale studies that can simultaneously capture broad patterns in legal decision-making while preserving the contextual richness that traditional computational methods often sacrifice.

In line with our previous review, this article focuses primarily on the second line of work, while also covering related research that LLMs have opened up for empirical scholars, including benchmarking and engineering research oriented toward improving model performance on legal tasks. The scholars and projects highlighted in the following sections illustrate the breadth and transformative potential of this emerging research agenda. Due to space constraints, we exclude other relevant areas, such as socio-legal examinations of how LLMs are affecting legal institutions and law practice.

C. Data Generation and Feature Extraction

In our previous review, we documented how computational analysis often serves as a first step in empirical legal research, generating structured data from unstructured legal texts for use with traditional statistical techniques, including causal inference.[1] Large language models extend this line of work in important ways. First, they enable the extraction of features that depend on multiple stylistic and semantic dimensions, such as patterns in legal reasoning styles. Earlier approaches often required substantial dimensionality reduction, which could obscure or discard such

1 An emerging econometrics literature distinguishes between prediction-policy and causal inference research, maintaining that in the latter context the most appropriate role for computational tools, including large language models, often is to create measurable variables from unstructured data or otherwise generate inputs for downstream statistical analysis (see Kleinberg et al. 2015; Ludwig et al. 2025).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

complex attributes. Second, they enable the extraction of relatively complex features that earlier machine learning methods could capture only at the cost of assembling extensive, task-specific training datasets. By reducing or eliminating that requirement, LLMs lower the barriers to incorporating new, previously cost-prohibitive dimensions into empirical legal analysis. Overall, the promise of LLMs in this area appears substantial enough that some have described this application as the "killer app for LLMs in empirical legal research" (Choi 2025).

There are two main approaches to feature extraction with the help of LLMs. One combines LLMs with traditional supervised learning methods. Snippets of text – such as paragraphs from legal opinions or contractual documents – are first converted into vector representations (embeddings). Embeddings are then paired with human-coded datasets to train more conventional machine learning algorithms, which can subsequently classify unlabeled portions of a corpus. Thalken et al. (2023) and Stiglitz & Thalken (2024) illustrate this approach, using transformer models on annotated Supreme Court opinions to classify modes of legal reasoning. Frankenreiter (2025) applies this strategy to identify fee-shifting and forum-selection provisions in corporate charters and bylaws. By preserving richer linguistic information than bag-of-words approaches, embedding-based methods make it possible to detect features that earlier computational methods often could not classify, in particular those that depend on subtle stylistic or semantic cues.

The second approach relies entirely on LLMs and uses their text-generation capabilities to produce classifications directly. For example, Frankenreiter & Talley (2026) identify the presence of 102(b)(7) waivers in corporate charters by providing the text of charters to OpenAI's ChatGPT, together with a prompt instructing the model to determine whether the provision appears. Extending this method, Frankenreiter & Hirst (2026) show that ChatGPT can generate highly detailed codings of complex provisions (advance notice bylaws), in some cases matching or exceeding the performance of human coders. Similarly, Oliver et al. (2024) demonstrate that ChatGPT can accurately classify the presence of suspicious factors in narrative descriptions of traffic-interdiction stops. Unlike the first approach, this method requires no labeled training data: the classification step draws entirely on the linguistic and contextual patterns LLMs learned during model training, substantially reducing the costs of compiling datasets.

However, the second approach also comes with challenges. In the first approach, techniques like cross-validation allow for a robust assess-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ment of an extraction pipeline's accuracy during the training stage. By contrast, when predictions are generated directly by an LLM, assessing reliability is less straightforward. This difficulty is compounded by the "jagged technological frontier" of AI (Dell'Acqua et al. 2023): LLMs can excel at some classification tasks while failing at others, and performance often defies intuitive expectations about task difficulty. Moreover, outputs can be highly prompt-dependent. Against this backdrop, careful validation against human-coded datasets is essential. To avoid overfitting, these validation datasets should be distinct from any datasets used to refine prompts or otherwise tune the classification process. That said, the scale of human labeling required for validation is often much smaller than for training a separate classifier – often hundreds rather than thousands (or more) of labeled examples.

D. Prediction, Classification, Description

In our previous review, we also highlighted that computational methods are sometimes applied not just to generate datasets for downstream analysis, but to produce results more directly. Prediction tasks – such as forecasting case outcomes – are a central example. Large language models can also be deployed in this mode. Their ability to process and generate legal text in contextually rich, human-like ways, and to adapt across diverse legal domains, makes them well-suited to support novel forms of legal prediction, from simulating judicial reasoning to generating plausible arguments for each side in a dispute. An example of this approach is Nigam et al. (2024), who employ LLMs to predict the outcomes of court cases based on the fact patterns presented.

Beyond text generation, researchers have begun to use the vector representations produced by these models to obtain measures for the meaning of words in legally relevant contexts. For example, Nyarko & Sanga (2022) develop a statistical test that leverages word embeddings to quantify differences in meaning across groups and contexts. Even more ambitiously, Arbel & Hoffman (2024) introduce a method for estimating the ordinary meaning of contract terms, and Choi (2024) applies a related embedding-space methodology to quantify uncertainty in statutory language.

While this line of work – particularly the ability to leverage the language-generating capabilities of LLMs to produce novel forms of pre-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dictions – holds great promise, it also faces major challenges. The most important concerns the validation of these tools to ensure they produce meaningful output. This problem is related to the one described in the previous section. LLMs are built on the basis of next-word-prediction technology, and their ability to generate legally meaningful outputs is not a given. In other words, there is an alignment problem: LLMs are optimized for next-word prediction, not for producing legally correct answers, and so their actual performance in this domain requires rigorous validation. Unlike in the context of feature extraction tasks, which usually allow for straightforward accuracy tests with validation data, such validation can be complicated. Especially when LLM use involves the generation of unstructured text or when embeddings are used to quantify the meaning of terms, it can be difficult to assess the extent to which the generated text or measures are meaningful or correct (see Tobia 2024).

E. Engineering and Benchmarking Studies

The emergence of LLMs has spawned a new category of legal AI research that approach legal tasks as engineering challenges to be systematically benchmarked and optimized. This engineering-focused approach prioritizes measurable performance improvements over theoretical insights about law itself. Unlike the causal inference and prediction studies discussed in earlier sections, which seek to understand legal phenomena, engineering studies focus primarily on optimizing model performance against standardized metrics.

One significant development in this line of research has been the creation of large-scale benchmarks designed to evaluate LLMs across multiple dimensions of legal reasoning. LegalBench (Guha et al. 2023), developed through a collaborative process involving over 40 contributors, consists of 162 tasks covering six different types of legal reasoning. LaborBench (Hariri & Ho 2025) is a dataset based on state unemployment insurance laws to assess performance on extracting statutory information. Zheng et al. (2025) develop a benchmark based on bar exam questions to assess performance on retrieval augmented generation tasks in the legal domain. LEXam (Fan et al. 2026) focuses on long-form reasoning derived from 340 actual law exams comprising 4,886 questions in English and German. LexEval represents China's largest legal evaluation dataset, introducing standardized comprehensive benchmarks for evalu-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ating LLMs across Chinese legal contexts (Li et al. 2024). These efforts highlight the importance of multilingual and cross-jurisdictional evaluation frameworks as LLMs are deployed globally.

A parallel strand of engineering research focuses on developing specialized legal LLMs through targeted training and fine-tuning approaches. Research by Niklaus et al. (2024) demonstrates that law-specific instruction training can improve LLM performance on legal task benchmarks. Studies have explored various approaches to legal model optimization, from domain-specific pre-training to instruction fine-tuning on curated legal corpora (Dominguez-Olmedo et al. 2024). Related research has examined prompt engineering strategies (Zambrano 2024) and few-shot learning approaches (Doyle & Tucker 2024). Others have focused on semantic role extraction and legal information extraction tasks (Bakker et al. 2025; de Faria et al. 2025).

One pressing engineering challenge for the practical use of LLMs in the legal context has been hallucinations – instances where LLMs generate plausible but factually incorrect legal information. Dahl et al. (2024) revealed that hallucinations occur between 58 % of the time with ChatGPT-4 and 88 % with Llama 2 when asked specific, verifiable questions about federal court cases. Subsequent research by Magesh et al. (2025) evaluated commercial legal research tools, finding that Lexis+ AI and Westlaw AI-Assist hallucinate between 17%-33 % of the time despite using retrieval-augmented generation techniques. This work demonstrates that even sophisticated engineering approaches cannot entirely eliminate hallucinations, emphasizing the need for transparent benchmarking of commercial products.

F. Challenges and Limitations

While large language models offer transformative capabilities for legal analysis, their integration into legal scholarship and practice poses a number of important challenges that must be carefully addressed. Among the most pressing concerns are accuracy and reliability. Unlike traditional computational methods where errors are typically systematic and predictable, LLM failures can be subtle, inconsistent, and difficult to detect. Plausible hallucinations and the jagged technological frontier phenomenon can make it difficult for legal scholars and practitioners to understand and predict errors associated with LLMs. Legal change may also

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

prove to be a fundamental limit on the use of LLMs in certain cases. Legal information evolves continuously through new legislation, regulations, and case law. LLMs trained on historical data are much more well-suited for providing a snapshot based on prior information than offering clear guidance on the current state of the law.

Another important set of challenges concern bias and fairness. Because LLMs are trained on vast corpora of legal texts that reflect historical patterns of legal decision-making, their outputs may encode prior biases or discriminatory behavior. This fact can help reveal patterns in ways that are useful for researchers (Ash and Chen), but scholars must work carefully to ensure that their work illuminates, rather than perpetuates, bias. The scale and opacity of LLM training data make bias identification and mitigation particularly challenging. Unlike traditional machine learning approaches where training datasets can be systematically audited, LLM training involves processing billions of documents whose contents and biases cannot be comprehensively characterized.

The integration of LLMs into legal practice also raises fundamental questions about professional responsibility that existing ethical frameworks are not equipped to address. Legal professionals have duties of competence, confidentiality, and zealous advocacy that may be compromised by inappropriate reliance on LLM systems. Courts and regulatory bodies are grappling with how to address LLM-generated content in legal proceedings, raising novel questions about admissibility, authentication, and reliability standards. Traditional rules of ethics, evidence, and procedure were not designed to handle content generated by artificial intelligence systems, creating uncertainty.

G. Conclusion

The emergence of large language models represents a paradigm shift in how legal scholars conduct empirical research. Our review identifies three dominant applications: data generation and feature extraction that dramatically reduce barriers to incorporating complex textual features; inference and prediction applications unlocking new forms of legal analysis; and engineering studies establishing technical foundations despite persistent accuracy challenges. These developments collectively represent both the culmination of decades of progress and the opening of entirely new research frontiers.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

However, this transformation brings fundamental research challenges that demand sustained scholarly attention. The persistent problem of hallucinations introduces validation complexities absent in traditional computational methods. Models may excel at sophisticated legal reasoning while failing unpredictably at seemingly simpler tasks, defying intuitive assessment of reliability. Beyond technical limitations, LLMs raise deeper methodological questions about the nature of legal knowledge and reasoning. When models trained on vast legal corpora produce insights that match or exceed human expert performance, scholars must grapple with the difference between genuine understanding or sophisticated pattern matching, and what this distinction means for empirical legal research.

Moving forward, legal scholars should approach LLMs as powerful tools requiring careful validation rather than as replacements for traditional methods. In the domain of practice, professions should view these tools as sophisticated assistants that require close oversight, particularly given hallucination risks. And policymakers face the challenge of balancing LLMs substantial benefits with requirements for accuracy, fairness, and accountability through new validation standards and transparency guidelines. The response of these communities will determine how these recent technological innovations will affect the practice and study of the law.

References

Arbel, Yonathan, and David A. Hoffman. "Generative Interpretation." *New York University Law Review* 99, no. 2 (May 2024): 451–514.

[Ash, same volume].

Bakker, Roos, Akke Schroevers, Romy van Drie, et al. "Semantic Role Extraction in Law Texts: a Comparative Analysis of Language Models for Legal Information Extraction." *Artificial Intelligence and Law* (2025).

Choi, Jonathan. "Measuring Clarity in Legal Text." *University of Chicago Law Review* 91, no. 1 (2024): 1–82.

Choi, Jonathan. *A Highly Tactical Guide to Generating Data from Legal Text with LLMs*. Unpublished manuscript, 2025.

Dahl, Matthew, Varun Magesh, Mirac Suzgun, and Daniel Ho. "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." *Trademark Reporter* 114, no. 6 (2024): 880–909.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, et al. *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Knowledge Worker Productivity and Quality. Harvard Business School Working Paper No. 24–013, September 2023.

Dominguez-Olmedo, Ricardo, Vedant Nanda, Rediet Abebe, et al. "Lawma: The Power of Specialization for Legal Annotation." In *Proceedings of the Thirteenth International Conference on Learning Representations* (ICLR 2025), 35059–35103. Appleton, WI: ICLR, 2025.

Doyle, Colin, and Aaron Tucker. "If You Give an LLM a Legal Practice Guide." In *Proceedings of the 2025 Symposium on Computer Science and Law* (CSLAW '25), 194–205. New York: Association for Computing Machinery, 2025.

Fan, Yu, Jingwei Ni, Jakob Merane, et al. "LEXam: Benchmarking Legal Reasoning on 340 Law Exams." *arXiv preprint* arXiv:2505.12864 (January 2026).

De Faria, Joana Ribeiro, Huiyuan Xie, and Felix Steffek. "Information Extraction from Employment Tribunal Judgments Using a Large Language Model." *Artificial Intelligence and Law* (2025). https://doi.org/10.1007/s10506-025-09443-z.

Frankenreiter, Jens. "The Other Delaware Effect." *Washington University in St. Louis School of Law Legal Studies Research Paper* No. 25-03-11, 2025. https://ssrn.com/abstract=5115285.

Frankenreiter, Jens, and Scott Hirst. *The Rise of the Horse-Choker: The Complexification of Advance Notice Bylaws*. Unpublished manuscript, 2026.

Frankenreiter, Jens, and Eric Talley. "Sticky Charters? The Surprisingly Tepid Embrace of Officer-Protecting Waivers in Delaware." *Harvard Business Law Review* 16 (2026): 105–168.

Frankenreiter, Jens, and Michael Livermore. "Computational Methods in Legal Analysis." *Annual Review of Law and Social Science* 16 (2020): 39–57.

Guha, Neel, Julian Nyarko, Daniel Ho, et al. "LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models." In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (NeurIPS '23), 44123–44279. Red Hook, NY: Curran Associates, Inc., 2023.

Hariri, Emaan, and Daniel Ho. "AI for Statutory Simplification: A Comprehensive State Legal Corpus and Labor Benchmark." In *Proceedings of Twentieth International Conference on Artificial Intelligence and Law* (ICAIL '25), 177–187. ACM, New York. 2025.

Janatian, Samyar, Hannes Westermann, Jinzhe Tan, et al. "From Text to Structure: Using Large Language Models to Support the Development of Legal Expert Systems." *arXiv preprint* arXiv:2311.04911 (November 2023).

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. "Prediction Policy Problems." *American Economic Review* 105, no. 5 (2015):491–95.

Li, Haitao, You Chen, et al. "LexEval: A Comprehensive Chinese Legal Benchmark for Evaluating Large Language Models." In *Proceedings of the Thirty-Eighth Conference on Neural Information Processing Systems* (NeurIPS 2024), 25061–25094. Red Hook, NY: Curran Associates, Inc., 2024.

Livermore, Michael, and Daniel Rockmore, eds. *Law as Data: Computation, Text, & the Future of Legal Analysis*. Santa Fe Institute Press, 2019.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan. "Large Language Models: An Applied Econometric Framework." *arXiv preprint* arXiv:2412.07031 (December 2025).

Magesh, Varun, Faiz Surani, Matthew Dahl, et al. "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools." *Journal of Empirical Legal Studies* 22 (2025): 216–242.

Nigam, Shubham Kumar, Aniket Deroy, Subhankar Maity, and Arnab Bhattacharya. "Rethinking Legal Judgement Prediction in a Realistic Scenario in the Era of Large Language Models." In *Proceedings of the Natural Legal Language Processing Workshop 2024*, 61–80. Miami, FL: Association for Computational Linguistics, 2024.

Niklaus, Joel, Elliott Ash, Matthias Stürmer, and Ilias Chalkidis. "LawInstruct: A Resource for Studying Language Model Adaptation to the Legal Domain." *arXiv preprint* arXiv:2404.02127 (April 2024).

Nyarko, Julian, and Sarath Sanga. "A Statistical Test for Legal Interpretation: Theory and Applications." *Journal of Law, Economics, & Organization* 38, no. 2 (2022): 539–569.

Oliver, Wesley, Morgan Gray, Jaromir Savelka, and Kevin Ashley. "Computationally Assessing Suspicion." *University of Cincinnati Law Review* 92, no. 4 (2024): 1108–1170.

Stiglitz, Edward, and Rosamond Thalken. "Historical Trends in Macro-Jurisprudence: A Language Model Assessment, 1870–2023." *Maryland Law Review* 84, no. 1 (2024): 46–101.

Thalken, Rosamond, Edward Stiglitz, David Mimno, and Matthew Wilkens. "Modeling Legal Reasoning: LM Annotation at the Edge of Human Agreement." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 9252–9265. Association for Computational Linguistics, December 2023.

Tobia, Kevin. "Algorithmic Interpretation: A Response to Professor Jonathan Choi's *Measuring Clarity in Legal Text*." *University of Chicago Law Review Online*, 2024.

Zambrano, Guillaume. "Case Law as Data: Prompt Engineering Strategies for Case Outcome Extraction with Large Language Models in a Zero-Shot Setting." *Law, Technology and Humans* 6, no. 3 (2024): 80–101.

Zheng, Lucia, Neel Guha, Javokhir Arifov, et al. "A Reasoning-Focused Legal Retrieval Benchmark." In *Proceedings of 4th ACM Symposium on Computer Science and Law* (CSLAW '25). New York: Association for Computing Machinery, 2025.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

LLM-gestützte Simulationen in der rechtsempirischen Forschung Ein Werkstattbericht

Johannes Kruse & Pascal Langenbach

A. Einleitung

Der computertechnische Fortschritt und Innovationen im Bereich der künstlichen Intelligenz (KI) beschäftigen zunehmend auch die Rechtswissenschaft. Dies ist nur folgerichtig, schließlich wird KI-Anwendungen ein disruptives Potential zugeschrieben, das neben anderen gesellschaftlichen Bereichen auch das Recht bzw. das Rechtssystem nachhaltig beeinflussen kann. Dies betrifft keineswegs allein die Rechtspraxis, also die Rechtsanwendung in Behörden, Gerichten oder der Anwaltschaft. Gerade die jüngste Entwicklungsstufe, die generative KI in Gestalt sogenannter großer Sprachmodelle (engl. Large Language Models oder LLMs), könnte ebenso große Veränderungen für das rechtswissenschaftliche Arbeiten und Forschen mit sich bringen. Gleichwohl finden diese potentiellen Auswirkungen für die Rechtswissenschaft – gerade im Vergleich zur Rechtspraxis (Begriffe wie „legal tech" oder der „robojudge" sind schon seit Jahren debattenprägend) – jedenfalls im deutschen Sprachraum bislang (zu) wenig Beachtung. Zentrale Fragen (etwa: wie kann/könnte künstliche Intelligenz die rechtswissenschaftliche Methodik und allgemein das juristische Denken verändern) werden bislang allenfalls vereinzelt aufgegriffen. Dabei sind die denkbaren Anwendungsszenarien zahlreich: Sie umfassen die Auswertung und Systematisierung von Rechtsprechung und Literatur, die Entwicklung dogmatischer Figuren und sogar das vollständige Verfassen rechtswissenschaftlicher Texte.

Christoph Engel ist eine Ausnahme. Neben den Anwendungsmöglichkeiten in der Rechtspraxis erstreckte sich seine Neugier von Anfang an auch auf die Rechtswissenschaft. Mit einem der beiden Autoren ist er etwa der Frage nachgegangen, inwieweit sich das Verfassen juristischer Kommentare, eine Kerntätigkeit der dogmatischen Rechtswissenschaft,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

an LLMs delegieren lässt.[1] Eingedenk der Tatsache, dass LLMs aus der Forschung zum Natural Language Processing hervorgegangen sind, liegen solche Anwendungen „eigentlich" auf der Hand. Weniger offensichtlich sind dagegen die Anwendungsmöglichkeiten von LLMs in einer rechtswissenschaftlichen Subdisziplin, für die Christoph Engel seit mehr als 20 Jahren steht wie wohl kein anderer: der experimentellen Rechtsforschung.

In verschiedenen sozialwissenschaftlichen Disziplinen werden LLMs bereits heute vermehrt eingesetzt, um zuvor mit menschlichen Versuchspersonen durchgeführte Experimente zu replizieren.[2] Die Ergebnisse dieser Versuche fallen im Einzelnen unterschiedlich aus, sind in der Tendenz aber durchaus vielversprechend. Im Rahmen einer großangelegten Replikationsstudie mit 276 ökonomischen Experimenten konnte etwa eine Vorhersagegenauigkeit von 78 % erreicht werden.[3] Im Bereich der experimentellen Rechtsforschung gibt es (noch) keinen entsprechenden Replikationstrend. Groß angelegte Untersuchungen, die eine Vielzahl von Studien in den Blick nehmen, stehen aus. Die Replikation einzelner Studien wurde aber bereits in Angriff genommen, wobei sich die entsprechenden Arbeiten an einer Hand abzählen lassen.[4] Einer der Vorreiter ist (einmal mehr) *Christoph Engel*: Gemeinsam mit *Richard McAdams* hat er eine LLM-gestützte Pipeline genutzt, um das Konzept des ordinary meaning (dem im US-amerikanischen Recht zentrale Bedeutung zukommt) zu erhellen.[5] Eine zu diesem Thema von *Kevin Tobia* durchgeführte Umfrage[6]

1 *Engel/Kruse*, Kommentar ohne Autor: Können Sprachmodelle das Kommentieren übernehmen?, JZ 79 (2024), 997; *dies.*, LLM as a law professor: Having a large language model write a commentary on freedom of assembly, Artificial Intelligence and Law 2026 (im Erscheinen).

2 *Anthis et al.*, LLM Social Simulations Are a Promising Research Method (2023), Preprint arXiv:2504.02234, 3.

3 *Hewitt et al.*, Predicting Results of Social Science Experiments Using Large Language Models (2024), 11; siehe weiterhin *Anthis et al.* (Fn. 2), 3; *Chen/Hu/Lu*, Predicting Field Experiments with Large Language Models (2025), Preprint arXiv:2504.01167, 1 f.

4 *Arbel/Hoffman*, Generative interpretation, NYUL Rev. 99 (2024), 451; *Arbel*, The Silicon Reasonable Person,University of Alabama School of Law Legal Studies Research Paper Series 2025; *Kruse*, The Ordinary Meaning Bot: Simulating Human Surveys with LLMs, MPI Collective Goods Discussion Paper, No. 2025/12, https://ssrn.com/abstract=5378203.

5 *Engel/McAdams*, Asking GPT for the Ordinary Meaning of Statutory Terms, University of Illinois Journal of Law, Technology and Policy 235 (2024).

6 *Tobia*, Testing Ordinary Meaning, Harv Law Rev 134 (2020), 726.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

konnte so teilweise repliziert werden.[7] Weitere Studien stammen etwa von *Yonathan Arbel* oder einem der Autoren.

Der vorliegende Beitrag folgt dem von *Christoph Engel* in den letzten Jahren insoweit eingeschlagenen Pfad und unternimmt mit einer LLM-gestützten Simulationspipeline den Versuch, ein rechtsempirisches Experiment zu replizieren. Um der (deutschen) Rechtswissenschaft entsprechende Möglichkeiten aufzuzeigen, ist der folgende Beitrag als Werkstattbericht konzipiert: Eine mögliche Vorgehensweise für rechtsempirische Replikationsstudien wird vorgestellt und bestehende Hürden sowie Fallstricke aufgezeigt. Exemplifiziert wird das Vorgehen „natürlich" mit der Replikation einer Studie von *Christoph Engel* – und *Nina Grgić-Hlača* als Co-Autorin. Diese Studie untersucht die Auswirkungen der durch den Supreme Court von Wisconsin[8] geforderten Warnungen bei der Verwendung des COMPAS-Tools in Strafverfahren und ist 2021 im *Journal of Legal Analysis* erschienen.[9]

Die vorliegende Untersuchung kann als Replikationsstudie im vorstehenden Sinne verstanden werden. Der klassische Replikationsbegriff passt indes nur teilweise. Klassische Replikationsstudien wollen die Robustheit eines berichteten (experimentellen) Effekts untersuchen. *Darum geht es auch vorliegend*. Zugleich wollen sie überprüfen, ob es sich bei den in der zu replizierenden Studie gefundenen Effekten um Zufallsbefunde handeln könnte (und mitunter auch, ob die Erstellung der Originalstudie angemessen und transparent erfolgte). *Darum geht es vorliegend eher nicht*. Die Befunde eines de lege artis durchgeführten sozialwissenschaftlichen Experiments können bzw. sollten aktuell wohl noch nicht allein auf Grundlage einer KI-basierten Replikation in Zweifel gezogen werden. Im Fall divergierender Ergebnisse zwischen Originalstudie und KI-Simulation liegt die Rechtfertigungslast dem Grunde nach bei der KI-Simulation. Dementsprechend kann man eine gelungene Replikation zwar als Beleg für die Robustheit der mit Menschen gewonnenen Befunde deuten. Gelingt die Replikation hingegen nicht, dürfte zunächst die Leistungsfä-

7 Was sich in einem hohen alignment zwischen den LLM-generierten und den echten Antworten niederschlägt, vgl. *Engel/McAdams* (Fn. 5); mit einer verbesserten Vorhersagepipeline ließ sich die Übereinstimmung sogar noch deutlich steigern, vgl. *Kruse* (Fn. 4), 3 f.

8 Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016).

9 *Engel/Grgić-Hlača*, Machine Advice with a Warning about Machine Limitations. Experimentally Testing the Solution Mandated by the Wisconsin Supreme Court, Journal of Legal Analysis 13 (2021), 284.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

higkeit KI-basierter Simulationen zu hinterfragen sein. Schließlich kann man experimentelle Studien mit menschlichen Versuchspersonen selbstverständlich niemals KI-basiert vollständig nachbilden. Replikation im vorliegenden Sinn meint dementsprechend keine Wiederholung des Originalexperiments, sondern stets eine *konzeptionelle* Replikation der gefundenen Ergebnisse.

Konzeptionelle Replikationsstudien mit KI-Agenten können unterschiedliche Zielsetzungen verfolgen. Einerseits kann es darum gehen, LLM-gestützte Simulationen mit KI-Agenten als neue sozialwissenschaftliche Methode zu etablieren. Die Durchführung KI-basierter sozialwissenschaftlicher Experimente ist ein relativ neues Phänomen. Vor diesem Hintergrund können gelingende Replikationen (insbesondere wenn diese sich über eine Vielzahl von Studien aus unterschiedlichen Disziplinen erstrecken) das generelle Vertrauen in die Validität der so gewonnenen Befunde stärken. Ist ein hinreichendes Vertrauensniveau erreicht, könnte man zukünftig womöglich teilweise auf menschliche Versuchspersonen verzichten und Experimente rein KI-basiert durchführen. Praktisch können KI-basierte Experimente auch als Vorstufe für Experimente mit Menschen dienen; um das Forschungsdesign zu testen oder Hypothesen zu entwickeln.[10] Insoweit bedarf es vor allem großangelegter Replikationsstudien, die für einen bestimmten Forschungsbereich oder eine bestimmte Methode den Nachweis erbringen, dass eine Replikation mit KI-Agenten grundsätzlich möglich ist.

Andererseits kann eine Replikationsstudie bei einer konkreten Forschungsfrage ansetzen und versuchen, die bereits vorliegenden Erkenntnisse zu erweitern. Auch hier geht es zunächst um die Frage, ob bzw. inwieweit sich ein mit Menschen durchgeführtes Experiment KI-basiert replizieren lässt. Ist dieser Nachweis erbracht, kann man den Ansatz ausbauen, um weitergehende Erkenntnisse zu gewinnen. Einen Erkenntnisgewinn verspricht insbesondere die Überwindung forschungspraktischer Hürden, denen Studien mit Menschen naturgemäß ausgesetzt sind (zum Beispiel auf Grund ethischer Grenzen oder begrenzter Forschungsmittel). Hierzu bedarf es weniger einer Untersuchung in der Breite als in der Tiefe. Man konzentriert sich auf eine einzelne Studie/Forschungsfrage, die sodann en détail beleuchtet wird. Freilich gewinnt man auch auf diese Weise

10 *Dillion et al.*, Can AI language models replace human participants?, Trends in Cognitive Sciences 27 (2023), 597, 598; *Hewitt et al.* (Fn. 3).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Einsichten über die generelle Tauglichkeit LLM-gestützter Simulationen, die das Vertrauen in die neuartige Methodik stärken oder schwächen können. Insoweit überschneiden sich die beiden Zielsetzungen.

Wie erwähnt gibt es bereits großangelegte Replikationsstudien, die in unterschiedlichen Disziplinen versuchen, den Nachweis zu erbringen, dass sich sozialwissenschaftliche Untersuchungen mit KI-Agenten replizieren lassen. Gerade vor diesem Hintergrund verfolgt der Beitrag primär die zweite Zielsetzung. Im Rahmen der LLM-gestützten Simulation erfolgt eine eingehende Auseinandersetzung mit einer einzelnen Studie. Auf diese Weise wird untersucht, inwieweit sich die praktischen Begrenzungen der ursprünglichen Studie überwinden ließen. Daneben verfolgt der Beitrag ein generelles Erkenntnisinteresse, indem er nach Gründen für das Gelingen oder Nichtgelingen KI-basierter Replikationen im Bereich experimenteller Rechtsforschung fragt.

Der vorliegende Beitrag versteht sich als Werkstattbericht. Er berichtet kein abgeschlossenes Unterfangen, sondern zeichnet „work in progress" nach. Es kommt ihm daher weniger darauf an, inwieweit die konkret unternommene Replikation gelingt. Vielmehr will er aufzeigen, wie man eine LLM-gestützte Replikation im Bereich der experimentellen Rechtsforschung durchführen könnte und welche Herausforderungen dabei zu meistern sind. Der Beitrag gliedert sich wie folgt: Zunächst erfolgt eine kurze Darstellung der Originalstudie (unter B.I), um das Fundament für die Replikationsstudie zu legen. Danach wird die Replikationsstudie dargestellt (unter B.II), wobei deren vier Module jeweils gesondert betrachtet werden. Anschließend werden die Simulationsergebnisse beschrieben sowie (kurz) analysiert (unter B.III). Der Beitrag schließt mit einem Ausblick auf die Einsatzmöglichkeiten, Chancen und Risiken LLM-gestützter Simulationen in der empirischen Rechtsforschung (unter C.).

B. Ein Replikationsversuch

I. Die Originalstudie

1. Kontext

In den Vereinigten Staaten können Richterinnen und Richter in vielen Bundesstaaten zur Unterstützung (einiger) ihrer Entscheidungen im Rahmen von Strafverfahren, etwa hinsichtlich der vorläufigen Freilassung eines Angeklagten auf Kaution (bail) oder der Strafzumessung im Fall

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

einer Verurteilung (sentencing) auf algorithmische Entscheidungshilfen zurückgreifen.[11] Diese Entscheidungsunterstützungssysteme stellen Risikoklassifizierungen für die jeweiligen Angeklagten zur Verfügung, zum Beispiel hinsichtlich der Wahrscheinlichkeit einer erneuten Verhaftung in einem bestimmten Zeitraum. Eine solche Anwendung ist COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). Da es sich um eine kommerzielle Anwendung handelt, sind die der Risikoklassifizierung zugrundeliegenden Algorithmen nicht öffentlich zugänglich. Die Verwendung von COMPAS zur Strafzumessung im US-Bundesstaat Wisconsin wurde vom Supreme Court von Wisconsin überprüft.[12] Im zugrundeliegenden Fall war COMPAS in einem Verfahren gegen Eric Loomis eingesetzt worden, das mit einer Freiheitsstrafe endete. Die Strafzumessung stützte sich in Teilen auf die von COMPAS ausgegebene Risikoklassifizierung. Der Supreme Court von Wisconsin sah in der Verwendung von COMPAS keine Verletzung der „Due-process"-Rechte von Eric Loomis. Allerdings stellte er die Verwendung von COMPAS im Rahmen der Strafzumessung unter die Bedingung, dass die COMPAS-Risikoeinschätzungen mit Warnungen über die Beschränktheit der Klassifizierungen verbunden werden.[13]

2. Forschungsfrage

Die Forderung des Supreme Courts, wonach die Risikoeinschätzungen durch COMPAS den Richtern nur gemeinsam mit einer entsprechenden Warnung zur Verfügung gestellt werden dürfen, impliziert, dass die entsprechenden Warnungen Einfluss auf das Verhalten der Richter haben. Das Gericht scheint zu erwarten, dass die Richterinnen und Richter ein anderes Strafmaß verhängen, wenn sie zusätzlich zur maschinellen Risikoeinschätzung eine Warnung sehen. Andernfalls wäre die Warnung überflüssig.[14] Diesen unterstellten Verhaltenseffekt der Warnung auf das

11 Die Beschreibungen in diesem Absatz entstammen der Originalstudie, *Engel*/Grgić-Hlača (Fn. 9), 288 f., sowie Recent Case, Criminal Law – State v. Loomis, 881 N.W.2d 749 (Wis. 2016), Harv Law Rev 130 (2017), 1530, siehe dort jeweils auch für weitere Verweise auf die Entscheidung und die Prozessmaterialien.

12 State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

13 Siehe für den Inhalt der Warnung *Engel*/Grgić-Hlača (Fn. 9), 288 f.

14 *Engel*/Grgić-Hlača (Fn. 9), 291: „For the solution mandated by the Wisconsin Supreme Court to be effective human decision-makers must decide differently when

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Verhalten von Richterinnen und Richtern untersuchen *Christoph Engel* und *Nina Grgić-Hlača* in ihrer Studie.

Um die Forschungsfrage experimentell operationalisierbar zu machen, wandeln sie den ursprünglichen Entscheidungskontext in verschiedener Hinsicht ab. Das Urteil des Supreme Courts von Wisconsin betrifft die Verwendung von COMPAS bei der Strafzumessung. *Engel* und *Grgić-Hlača* untersuchen in ihrer Studie jedoch Entscheidungen über die Freilassung auf Kaution im Vorfeld der Anklage („bail or jail"). Grund hierfür ist die Verfügbarkeit eines Datensatzes, der für diese Entscheidungskategorie sowohl die COMPAS-Bewertung als auch die tatsächliche Rückfälligkeit innerhalb von zwei Jahren enthält.[15] Zudem beschränken sie sich auf die Bewertungen zur Rückfälligkeit (General Recidivism Score), obwohl COMPAS auf unterschiedlichen Dimensionen Risikobewertungen errechnet (zusätzlich einen „Violent Recidivsm Score" und das „Pretrial Release Risk").[16]

3. Studiendesign

Wie die Autor:innen betonen, ist es ausgeschlossen, in echten Strafverfahren einigen Richterinnen oder Richtern die geforderten Warnungen zu zeigen, anderen hingegen nicht.[17] Deshalb weichen sie methodisch auf eine Vignette-Studie aus, in der sie die Kautions-Entscheidung experimentell nachbilden. Dazu erstellen sie aus dem genannten Datensatz 50 Vignetten, die Informationen über die begangene Tat, das Alter, das Geschlecht sowie über die kriminelle Vorgeschichte des/der Angeklagten enthalten. Auf Grundlage dieser Vignetten und der von COMPAS errechneten Risikobewertung treffen die Versuchspersonen ihre Entscheidungen. In ihrer Veröffentlichung bündeln *Engel* und *Grgić-Hlača* die Ergebnisse fünf unterschiedlicher (Teil-)Studien, die im experimentellen Aufbau variieren. Die Variationen haben im Großen und Ganzen keinen wesentlichen Einfluss auf die Ergebnisse.[18] In der Hauptstudie (Teilstudie 1) lesen die Versuchspersonen zunächst die jeweilige Vignette. Dann

just receiving machine advice compared with receiving machine advice that comes with the prescribed warnings."

15 *Engel*/Grgić-Hlača (Fn. 9), 286.

16 *Engel*/Grgić-Hlača (Fn. 9), 299.

17 *Engel*/Grgić-Hlača (Fn. 9), 294.

18 *Engel*/Grgić-Hlača (Fn. 9), 286f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

treffen sie Entscheidungen zu dieser Vignette. Sie geben für jede Vignette an, ob sie sich für „jail" oder „bail" entscheiden, wie hoch sie die Rückfallwahrscheinlichkeit einschätzen und wie sicher sie sich ihrer Entscheidung sind. Im Anschluss sehen sie den COMPAS-Risiko-Score und treffen die Entscheidungen erneut. Der Unterschied zwischen der Kontrollgruppe und der Treatmentgruppe besteht darin, dass die COMPAS-Risikoklassifizierung nur für letztere Gruppe mit einer Warnung versehen ist. Bei der im Experiment verwendeten Warnung handelt es sich um die gekürzte Originalversion, die in Wisconsin nach dem Urteil des dortigen Supreme Courts tatsächlich verwendet wurde.[19] Die Hauptstudie untersucht, wie sich die Warnung auf Veränderungen zwischen der ersten und der zweiten Einschätzung auswirkt.[20] In einer zweiten ergänzenden Teilstudie wird dieser sequentielle Aufbau aufgegeben, nun erhalten die Versuchspersonen den COMPAS-Score bereits unmittelbar mit der jeweiligen Vignette.[21] Die hier beschriebene Replikation versteht sich wie dargestellt als ein erster Schritt und beschränkt sich – aus pragmatischen Gründen – auf den Versuchsaufbau der zweiten Teilstudie.

4. Versuchspersonen

Die Versuchspersonen der Originalstudie rekrutieren sich aus dem Pool des Panel-Providers *Prolific*. An der zweiten Teilstudie nahmen letztlich 134 Personen teil. Jeweils 67 Personen in der Kontrollgruppe (keine Warnung) und 67 in der Treatment-Gruppe (Warnung). Alle 134 Personen hatten zuvor zwei Fragen zur Aufmerksamkeitskontrolle richtig beantwortet.[22]

Die Entscheidung über die Strafzumessung und die Entscheidung über die vorläufige Freilassung auf Kaution werden in Wisconsin von Richterinnen und Richtern getroffen. Nur das Verhalten dieser sehr kleinen demografischen Gruppe soll durch die Warnung beeinflusst werden. Hier müsste man also ansetzen, um die Verhaltenswirksamkeit der Warnungen „optimal" zu untersuchen. Dem stehen allerdings forschungspraktische Hürden entgegen. Die meisten Richter:innen (aus Wisconsin)

19 *Engel*/Grgić-Hlača (Fn. 9), 297 f.
20 *Engel*/Grgić-Hlača (Fn. 9), 296.
21 *Engel*/Grgić-Hlača (Fn. 9), 306 ff.
22 *Engel*/Grgić-Hlača (Fn. 9), 326, von den ursprünglich 138 Teilnehmer:innen gelang dies also nur vier Personen nicht.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

wird man kaum zur Teilnahme an einer Online-Vignette-Studie bewegen können. Hinzu kommt, dass diese in ihrem beruflichen Alltag regelmäßig mit den rechtlich verbindlichen Warnungen konfrontiert sind.[23] Das bedeutet, dass eine Versuchsanordnung, die einen Entscheidungskontext ohne den Effekt der Warnung darstellt, schwer zu implementieren sein könnte, weil die Richter:innen sich die entsprechende Warnung stets hinzudenken (oder ihre erlernte Reaktion auf den COMPAS-Score von der real vorhandenen Warnung geprägt ist). Vor diesem Hintergrund greifen *Engel* und *Grgić-Hlača* auf den Probandenpool von *Prolific* zurück, der eher nicht aus Richterinnen und Richtern aus Wisconsin besteht. Um sich den Charakteristika von Richter:innen oder wenigstens ihrem Erfahrungshintergrund anzunähern, versuchten *Engel* und *Grgić-Hlača* Versuchspersonen mit Jury-Erfahrung zu rekrutieren. In der zweiten Teilstudie verfügten 67 Prozent der Versuchspersonen über entsprechende Erfahrungen.[24] Für die einzelnen Studien berichten *Engel* und *Grgić-Hlača* verschiedene demografische Eigenschaften ihrer Versuchspersonen, die sich von der Verteilung in der Gesamtbevölkerung (US-Census) unterscheiden.

5. Ergebnisse

In der Studie von *Engel* und *Grgić-Hlača* hatten die Warnungen insgesamt keinen übermäßigen Einfluss auf das Entscheidungsverhalten der Versuchspersonen. Insbesondere für die normativ entscheidende Frage nach der Haftentlassung auf Kaution blieben sie ohne erkennbaren Einfluss.[25] Für die zweite Teilstudie werden folgende Befunde berichtet: Die Warnungen hatten keinen Effekt auf die Entscheidung „bail" vs. „jail", führten aber zu einer signifikanten Erhöhung der geschätzten Rückfallwahrscheinlichkeit; ferner waren sich die Versuchspersonen ihrer Einschätzung sicherer, wenn dem Score eine Warnung beigefügt war.[26]

23 Für diese Einschränkungen *Engel*/Grgić-Hlača (Fn. 9), 298 f.
24 *Engel*/Grgić-Hlača (Fn. 9), 285, 326.
25 Zusammenfassend *Engel*/Grgić-Hlača (Fn. 9), 286.
26 *Engel*/Grgić-Hlača (Fn. 9), 308 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Die Replikationsstudie

1. Überblick

Die vorliegende Replikationsstudie nutzt Large Language Models, um das Verhalten menschlicher Versuchspersonen in einem experimentellen Setting zu simulieren. Dem liegt die Erwartung zugrunde, dass LLMs durch ihre Trainingsdaten sowie Designziele ein relativ gutes „Verständnis" davon haben, wie sich Menschen in unterschiedlichen Situationen verhalten, und deshalb als eine Art „homo silicus"[27] dienen können. Gerade durch das Fine-Tuning mittels Reinforcement Learning wird diesen Modellen vermittelt, wie Menschen auf bestimmte, sprachlich vermittelte Stimuli reagieren. LLM-gestützte Simulationen können eine beachtliche Vorhersagegenauigkeit erreichen. In einigen Studien zu sozialwissenschaftlichen Experimenten korrelieren GPT-4-Vorhersagen in hohem Maße (r = 0,85 bis 0,91) mit den tatsächlichen Treatment-Effekten.[28] Aus dem Vorgesagten folgt zugleich, dass die Vorhersagegenauigkeit maßgeblich davon abhängt, ob bzw. inwieweit das in Rede stehende Verhalten in den Trainingsdaten repräsentiert ist.[29] Dementsprechend dürften LLMs etwa das Verhalten von Verbrauchern besser abbilden können als von professionellen Richtern. Um das Verhalten der menschlichen Versuchspersonen simulieren zu können, kommen KI-Agenten mit einem spezifischen demografischen Profil zum Einsatz, das auf den (wenigen) bekannten demografischen Eigenschaften der echten Versuchspersonen beruht. Die der Replikation zugrunde liegende Pipeline setzt sich aus vier Modulen zusammen, die jeweils gesondert betrachtet werden.

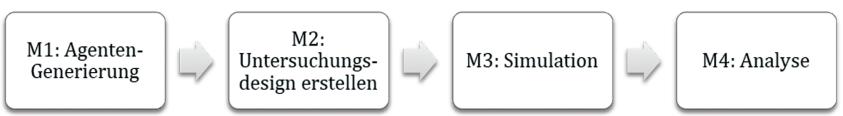


Abbildung 1: Die vier Module der Studie

27 *Horton,* Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?, Preprint arXiv:2301.07543 (2023), 1

28 *Hewitt et al.* (Fn. 3), 1; *Anthis et al.* (Fn. 2), 3.

29 Vgl. etwa *Kruse* (Fn. 4), 4 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Die vorliegende Studie ist als Prototyp konzipiert und soll in erster Linie als Anschauungsobjekt dienen. Dementsprechend wurden beim Untersuchungsdesign pragmatisch motivierte Entscheidungen getroffen. Um Dauer und Kosten der Durchführung im Rahmen zu halten, wurden insbesondere sowohl bei der Anzahl der KI-Agenten als auch der genutzten Modelle Abstriche gemacht.

2. Modul 1: Agentengenerierung

Modul 1 erzeugt synthetische Versuchspersonen (KI-Agenten) mit realitätsnahen demografischen Profilen als Ersatz für menschliche Versuchspersonen. Zunächst werden N = 268 Basis-Profile generiert, die die bekannten demografischen Charakteristika der echten Untersuchungsteilnehmer:innen abbilden. Deshalb konnten nur solche demografischen Merkmale verwendet werden, über die die Originalstudie Auskunft gibt. Das sind etwa das Geschlecht (63 % männlich, 37 % weiblich), die ethnische Zugehörigkeit (70 % weiß, 5 % afroamerikanisch, 7 % hispanisch, 17 % asiatisch, 1 % sonstige) oder die Jury-Erfahrung (67 % ja, 33 % nein). Weiterhin die politische Orientierung (50 % liberal, 18 % konservativ und 32 % moderat) sowie der Bildungsgrad (94 % Bachelorabschluss oder höher, 6 % sonstige).

Für jedes demografische Profil wird mittels GPT-4o (Temperature = 0,9) ein kurzer, realistischer Biografietext erzeugt (Name; Alter 25–75; US-Wohnort; Beruf; Familiensituation; Hobbys/Interessen; beruflicher Werdegang; gesellschaftliches Engagement; persönliche Wertvorstellungen; individuelles Merkmal). Die Details stammen aus dem Modell (ohne externe Statistikquellen). Ein Beispielprofil lautet:

> *Grant Mitchell, 38, resides in Portland, Oregon, where he works as a social media strategist for a nonprofit focused on environmental justice. Raised in a small town in Nebraska, Grant was the first in his family to earn a college degree, graduating with a Bachelor's in Environmental Science from the University of Oregon. He is married to Lena, a fellow activist he met at a climate change rally, and together they have a six-year-old daughter, Ruby. Grant is known for his passionate, progressive views, which he often expresses through his writing and activism, but his intensity can sometimes alienate more moderate peers. He has never served on a jury, attributing it to a combination of sheer luck and frequent travel. A fervent cyclist, he enjoys participating in long-distance races and is an advocate for bike-friendly city planning. His career has been diverse: from working as a campaign manager for local green politicians to teaching workshops on sustainable living. Grant's activism extends into his com-*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

munity involvement, where he organizes monthly clean-up drives and educational seminars on environmental policies. He is deeply committed to his values of equality and sustainability, though some criticize him for being too unyielding in debates. One unique, yet painful experience in his life was surviving a car accident in his early twenties, which left him with a permanent limp and intensified his commitment to public transportation advocacy. Despite his admirable dedication, Grant occasionally struggles with balancing his activism and family commitments, causing tension at home.

Die N = 268 Basis-Profile wurden randomisiert auf die Bedingungen aufgeteilt: 134 Agenten enthalten den reinen COMPAS-Score (Baseline), 134 einen COMPAS-Score inklusive Warnung (Treatment). Die 268 Basis-Profile werden 1:1 über die Modelle gespiegelt.

Wie bereits erwähnt, dürfte der Rückgriff auf den Teilnehmer-Pool des Panel-Providers *Prolific* und damit die Auswahl der Versuchspersonen in der Originalstudie in erster Linie pragmatisch motiviert gewesen sein. Obgleich man primär an dem Verhalten echter Richter:innen (als den tatsächlichen Adressat:innen der Warnungen) interessiert ist, greift man auf Laien zurück. Unter Kostengesichtspunkten war man sicherlich auch mit Blick auf die Zahl der Versuchspersonen limitiert. Mit KI-Agenten lassen sich diese Beschränkungen überwinden. Zwar sind auch die für die synthetischen Agenten notwendigen API-Abfragen nicht kostenlos (und auch die Antwortzeiten der Modelle sind zu berücksichtigen), im Grundsatz lässt sich die Beobachtungszahl aber skalieren. Aus Kostengründen wurde für die Simulation darauf verzichtet eine größere Anzahl an KI-Agenten zu erstellen. Allerdings übersteigt die Anzahl die der Originalstudie bereits deutlich. Überdies wurden als zweite Agentenpopulation „echte" Richterinnen und Richter erstellt. Das zugrundeliegende demografische Profil beruht auf Daten zu amerikanischen (Bundes-)Richter:innen.[30]

3. Modul 2: Erstellen des Untersuchungsdesigns

Modul 2 bereitet die experimentelle Umgebung vor. Jeder Agent sieht entweder (i) die Vignette mit COMPAS-Risikoinformation oder (ii) dieselbe

30 Die Zahlen stammen aus den Datensätzen des Federal Judicial Center (FJC) zur Demografie der Article-III-Richter; die ABA „Profile of the Legal Profession 2024 – Judges" fasst diese datenbasiert und mit Stichtag 01.08.2024 zusammen, www.americanbar.org/news/profile-legal-profession/judges/.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Vignette mit COMPAS-Risikoinformation plus Warnhinweis. Die Zuweisung der Agenten zur Kontroll- oder Treatment-Gruppe nimmt Modul 1 vor. Pro Agent wird eine Serie von Vignetten zugewiesen (Standard: 10, konfigurierbar); in der Originalstudie erhielten die Versuchspersonen 50 Vignetten. Zur Sicherung der Inhaltsäquivalenz werden die Vignetten zwischen den Bedingungen mittels stabiler Signaturen (CASE_ID, Text-Hash) spiegelbildlich gematcht und verifiziert. Beide Bedingungen enthalten damit identische Fälle in identischer Reihenfolge.

Die Simulation verwendet die ersten 10 Vignetten der Originalstudie, die ein niedriges bis mittleres Risikoniveau (COMPAS Decile 1–6) aufweisen. Für deren Verwendung ließe sich anführen, die entsprechenden Fälle seien rechtspolitisch besonders relevant. Richterinnen und Richtern kommt insoweit tendenziell ein größerer Entscheidungsspielraum zu, sodass hier auch ein größeres Potential für algorithmische Entscheidungshilfen bestehen dürfte. Die Simulation stellt (in Abweichung von der Originalstudie) vorrangig auf die kategoriale Risikoeinstufung (low/medium/high) ab, mit dem Decile-Score als ergänzender Information in Klammern (z. B. „General Recidivism Risk: high (Decile: 8)"). Hierdurch soll einer dokumentierten Schwäche von LLMs Rechnung getragen werden, nämlich den eingeschränkten numerischen Reasoning-Fähigkeiten sowie der Tendenz einer inkonsistenten Verarbeitung granularer Skalen.[31] Während Menschen Decile-Scores intuitiv ordinieren und in Risikoeinschätzungen übersetzen können, zeigen LLMs bei direkter numerischer Instruktion („Decile 8 von 10") häufig Schwierigkeiten, die Richtung (höher = riskanter) und die Größe der Abstände korrekt zu berücksichtigen. Die kategoriale Kodierung soll diese Fehlerquelle reduzieren und die Wahrscheinlichkeit erhöhen, dass das Modell die Risikoinformation adäquat verarbeitet.[32] Gleichzeitig bleibt der numerische Score sichtbar, um die Informationsäquivalenz zur Originalstudie weitgehend zu wahren.

Jede Vignette folgt – wie auch in der Originalstudie – einer standardisierten Struktur: Demographische Basisinformationen (Geschlecht, Alter), aktueller Tatvorwurf mit Deliktklassifikation (misdemeanor/felony), kriminalhistorische Angaben (Anzahl der Vorstrafen oder der Delikte in Kindheit und Jugendalter) sowie die COMPAS-Risikoprädiktion. Ein Beispiel: *„The defendant is a female aged 43. They have been charged with:*

31 Vgl. *Engel/McAdams* (Fn. 5), 247 m.w.N.

32 *Imani et al.*, Mathprompter: Mathematical reasoning using large language models, Preprint arXiv:2303.05398 (2023); *Engel/McAdams* (Fn. 5), 268.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Driving under the influence. This crime is classified as a misdemeanor. They have been convicted of 0 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record. The COMPAS tool made the following prediction about this defendants' general recidivism score: General Recidivism Risk: low DECILE:1"

Wie in der Originalstudie werden für jede Vignette drei abhängige Variablen erhoben:

- Q1 (Recidivism Likelihood) auf einer 5-Punkt-Skala von „Extremely unlikely" bis „Extremely likely". Dies erfasst die subjektive Rückfallwahrscheinlichkeit.
- Q2 (Grant Bail) als binäre Entscheidung (Yes/No)
- Q3 (Confidence) auf einer 5-Punkt-Skala von „Completely guessing" bis „Completely confident". Dies misst die Entscheidungssicherheit.

Die Pipeline unterstützt optional kurze Begründungstexte (2–3 Sätze), die mehreren Zielen dienen können: (i) Qualitative Validierung, ob das LLM die experimentelle Manipulation tatsächlich verarbeitet hat (zum Beispiel durch explizite Referenzen auf den Warntext in der Treatment-Gruppe); (ii) Reasoning-Transparenz durch Exploration der Mechanismen hinter den Entscheidungen (nutzt das Modell stereotype Heuristiken oder differenzierte Abwägungen?); (iii) Sycophancy-Detektion durch Identifizierung von Fällen, in denen das Modell übermäßig „hilfreich" oder sozial erwünscht antwortet, statt realistische Varianz zu zeigen.[33] Gerade neuere LLMs werden zunehmend auf „Assistenz"-Verhalten optimiert, was zur Unterdrückung kontroverser oder sozial unerwünschter Antworten führen kann – ein Analogon zum Social-Desirability-Bias in Umfragen mit menschlichen Teilnehmerinnen und Teilnehmern.[34] Die Anforderung strukturierter Begründungen folgt zudem der Empfehlung von *Abdurahman et al.* (2025): „[A]uthors must rigorously demonstrate that the LLM possesses the necessary capabilities and cognitive processes inherent to human-data production."[35]. Die Begründungsoption wurde in der finalen Simulation zwar deaktiviert, kam indes in mehreren Testläufen zur Anwendung.

33 *Anthis et al.* (Fn. 2), 3.

34 *Anthis et al.* (Fn. 2), 3.

35 *Abdurahman et al.*, A Primer for Evaluating Large Language Models in Social-Science Research, Advances in Methods and Practices in Psychological Science 8 (2025), 15.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

4. Modul 3: Durchführung der Simulation

Modul 3 führt im Anschluss die eigentliche Simulation durch – wiederum mittels einer mehrstufigen Pipeline, die für jede Kombination (Agent × Vignette × Bedingung × Modell) strukturierte Vorhersagen generiert. Der Prozess gliedert sich in vier Hauptkomponenten: (1) Prompt-Generierung aus einer Prompt-Bank, (2) API-Aufruf mit Rate-Limiting, (3) Response-Parsing und Normalisierung, (4) Aggregation über Wiederholungen. Operativ kommen zwei Modellfamilien zum Einsatz: GPT-4o (OpenAI, November 2024) und Claude 3.5 Sonnet (Anthropic, Oktober 2024). Beide werden getrennt konfiguriert und aufgerufen; Effekte, die in beiden Modellfamilien auftreten, gelten als robuster.

Anstatt das Modell in eine Rolle schlüpfen zu lassen („Du bist ein 52-jähriger…"), instruiert der Prompt das LLM ausdrücklich als analytischen Prädiktor: Es soll – gestützt auf das demografische Agentenprofil, die Instruktionen und den Falltext – vorhersagen, wie eine reale Person mit diesen Merkmalen typischerweise antworten würde. Diese Perspektivverschiebung soll Stereotyp-„Performances" entgegenwirken und die Nachvollziehbarkeit der erzeugten Antworten erhöhen. Eine zentrale methodische Herausforderung bei LLM-basierten Simulationen ist die Prompt-Lotterie – die Abhängigkeit der Ergebnisse von idiosynkratischen Formulierungsdetails.[36] So besteht ein wesentlicher Unterschied darin, ob man ein Modell nach seiner eigenen Einschätzung oder der „richtigen Antwort" fragt oder es explizit um seine Einschätzung bittet, wie eine näher beschriebene Person die Frage beantworten wird.[37] Teilweise führen schon kleine Unterschiede bei der Eingabe zu gänzlich unterschiedlichen Ergebnissen. Um diese Empfindlichkeit zu kontrollieren, wird eine Prompt-Bank implementiert. Das ist eine Sammlung von zehn inhaltlich äquivalenten, stilistisch aber variierten Prompts pro experimenteller Bedingung. Für jede Kombination (Agent × Vignette × Bedingung × Modell) wird zufällig eine Variante gezogen, wobei die Antwortstruktur stets identisch bleibt (Q1 Rückfallerwartung, Q2 Bail-Entscheidung, Q3 Entscheidungssicherheit). Dieses Ensembling über Prompts folgt der von *Hewitt et al.* (2024)[38] vorgeschlagenen Simulationsstrategie und erhöht nachweis-

36 Vgl. *von der Heyde/Haensch/Wenz*, Vox Populi, Vox AI? Using Large Language Models to Estimate German Vote Choice, Social Science Computer Review (2025), 2.

37 *Von der Heyde/Haensch/Wenz* (Fn. 35), 2; *Kruse* (Fn. 4), 5.

38 *Hewitt et al.* (Fn. 3).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lich die Robustheit gegenüber einzelnen Formulierungsartefakten. Optional kann der Prompt – wie soeben dargestellt – kurze Begründungen anfordern. Der Kern-Prompt lautet (auszugsweise) wie folgt:

> SYSTEM PREAMBLE:
>
> You are an expert in behavioral prediction and survey response analysis.
>
> You are NOT role-playing as the participant. You are ANALYZING and PREDICTING likely responses based on the participant profile and the case.
>
> PARTICIPANT PROFILE: {profile}
>
> INSTRUCTIONS SHOWN TO PARTICIPANT: {instructions}
>
> CASE PRESENTED TO PARTICIPANT: {vignette}
>
> ALGORITHMIC INFORMATION: {compas_prediction}
>
> [Only in compas_warning:]
>
> WARNING TO PARTICIPANT: {warning_text}
>
> YOUR TASK:
>
> Provide your PROFILE-BASED PREDICTION:
>
> Q1_Recidivism: „Extremely unlikely" | „Somewhat unlikely" | „Neither likely nor unlikely" | „Somewhat likely" | „Extremely likely"
>
> Q2_Grant_Bail: „Yes" | „No"
>
> Q3_Confidence: „Completely guessing" | „Somewhat guessing" | „Neutral" | „Somewhat confident" | „Completely confident"
>
> CRITICAL RULES:
>
> – Use EXACTLY the labels above for Q1/Q2/Q3 (no alternatives).
>
> – Provide one definitive answer per question.
>
> [Optional; in Tests aktiviert, im finalen Lauf deaktiviert:]
>
> JUSTIFICATION: Provide 2–3 sentences explaining how the profile and case features lead to these predictions.

Die Entscheidung für GPT-4o und Claude 3.5 Sonnet beruht auf Erfahrungswerten zur Verwendung LLM-gestützter Simulationen. In anderen Kontexten konnten mit den beiden hier verwendeten Modellen die besten Vorhersagen erzielt werden.[39] Die Verwendung zweier Modelle dient auch der Robustheitsprüfung. Während modellspezifische Unterschiede auf Artefakte der Trainings- oder Alignment-Strategie hindeuten können, sprechen konsistente Effekte über beide Modelle hinweg für eine Genera-

39 Vgl. *Kruse* (Fn. 4) für GPT-4o.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lisierbarkeit der Befunde. Die Wahl der Temperature (0,7) ist das Ergebnis eines Kompromisses:[40] Einerseits sollen die Modelle nicht zu deterministisch antworten, da sich anderenfalls die Varianz in den Antworten menschlicher Versuchspersonen nicht abbilden ließe. Andererseits soll die Streuung nicht so groß sein, dass ein zufälliges Rauschen systematische Effekte überlagern könnte.

Operativ werden pro Kombination (Agent × Vignette × Bedingung × Modell) fünf Wiederholungen mit wechselnden Prompt-Varianten ausgeführt. Die Antworten werden zur besseren Vergleichbarkeit in feste Kategorien übersetzt. Bei Q2 (Yes/No) zählt die Mehrheit; bei Q1 und Q3 die häufigste Kategorie als Ergebnis. Zusätzlich berechnet das System Konsistenz- und Unsicherheitsmaße über die fünf Wiederholungen (u. a. Agreement-Rate, Varianz, normalisierte Entropie) und fasst sie in einem zusammengesetzten Zuverlässigkeitsindikator Model_Confidence zusammen, der hohe Übereinstimmung, niedrige Entropie und vollständige Struktur der Antworten belohnt. Am Ende werden lediglich die aggregierten Werte ausgegeben.

5. Modul 4: Analyse

Modul 4 analysiert die Ergebnisse und vergleicht sie mit den Ergebnissen der Originalstudie. Im Unterschied zur Originalstudie (50 Vignetten) erhalten die KI-Agenten jeweils nur 10 Fallbeschreibungen. Ein Vergleich mit den Gesamtergebnissen der Originalstudie ist daher nicht sinnvoll möglich. Vielmehr müssen die Ergebnisse vignettenspezifisch verglichen werden. Der insoweit erforderliche Zugriff auf die Rohdaten der Originalstudie stand den Autoren während der Arbeit an dem vorliegenden Beitrag nur eingeschränkt zur Verfügung (auch um das Überraschungsmoment der Festschrift nicht zu gefährden). Ein umfassender Vergleich konnte nur mit den Ergebnissen der ersten der in der Originalstudie berichteten Teilstudien durchgeführt werden. Wie dargestellt verwendet die erste Teilstudie einen sequentiellen experimentellen Ablauf: Die Versuchspersonen entscheiden zunächst ohne Kenntnis der maschinellen Risikoeinschätzung. Sodann erhalten sie den COMPAS-Score und die Möglichkeit, ihre vorherige Einschätzung anzupassen. In der vorliegen-

40 Und folgt im Übrigen den Empfehlungen vergleichbarer Studien, siehe etwa *von der Heyde/Haensch/Wenz* (Fn. 35), 8 f.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

den Studie hingegen entscheiden die KI-Agenten pro Vignette nur einmal (wobei ihnen stets die maschinelle Risikoeinschätzung zur Verfügung steht); dies entspricht dem Vorgehen in der zweiten Teilstudie der Originalstudie. Vor diesem Hintergrund werden die Simulationsergebnisse im Folgenden mit den abschließenden, unter Kenntnis des Risiko-Scores getätigten Angaben aus der ersten Teilstudie verglichen. Dies schränkt die Aussagekraft der folgenden Analysen selbstverständlich dahingehend ein, dass Unterschiede zwischen der Simulations- und der Originalstudie auch darauf beruhen könnten, dass die Daten aus der Originalstudie in dieser Hinsicht abweichend erhoben wurden.

Wie erwähnt wurde die Simulation mit zwei LLMs durchgeführt. Für die Auswertung werden beide Modelle zunächst gemeinsam behandelt und die Ergebnisse in „gepoolter" Form berichtet. Erst in einem zweiten Schritt werden die Modellunterschiede näher beleuchtet. Die unterschiedlichen LLMs werden genutzt, um das Entscheidungsverhalten einzelner KI-Agenten vorherzusagen. Die KI-Agenten haben aber für beide Modelle jeweils ein identisches demografisches Profil. Hier stellt sich die Frage, ob die Beobachtungen zwischen den LLMs aufgrund der gemeinsamen demografischen Basis der KI-Agenten als abhängige Beobachtungen zu behandeln sind oder ob sie dennoch unabhängige Beobachtungen darstellen. Für die Abhängigkeit spricht, dass für jeden KI-Agenten ein individualisiertes Profil mit Namen, Alter, Beruf und weiteren Merkmalen erstellt wurde. Man könnte also annehmen, in beiden Modellen entscheide „derselbe" KI-Agent. Für die Unabhängigkeit spricht, dass die demografischen Merkmale pro KI-Agent immer noch sehr begrenzt sind und es sich eben nicht um dieselbe Person mit weiteren unbekannten Eigenschaften handelt, sondern dass idiosynkratisches Entscheidungsverhalten erst aus der Kombination der demografischen Basis mit dem jeweiligen LLM entsteht: Ein individueller KI-Agent entstünde also erst durch die konkrete Entscheidungsvorhersage für ein bestimmtes demografisches Profil. In der folgenden Analyse werden die Beobachtungen zwischen den LLMs trotz der gemeinsamen demografischen Basis als unabhängig behandelt.

Die statistische Analyse beruht in einem ersten Zugriff auf zweiseitigen t-Test-Vergleichen der Durchschnittswerte pro Versuchsperson/Agent. Diese Tests berücksichtigen allerdings nicht, dass pro Person/Agent mehrere Entscheidungen (10 Vignetten) getroffen wurden. Um dem Rechnung zu tragen, wurden zusätzlich Mehrebenenregressionen (Mixed-Effects-Modelle) geschätzt. Aus Darstellungsgründen werden im Folgenden in erster Linie die Ergebnisse der t-Tests über Durchschnitts-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

werte berichtet. Vereinzelt werden jedoch ergänzend Ergebnisse der Regressionsanalysen eingebracht, die aufgrund der präziseren Modellierung der Datenstruktur grundsätzlich aussagekräftiger sind (höhere Teststärke).

III. Analyseergebnisse

1. Treatmentunterschiede

Tabelle 1 zeigt die durchschnittlichen Bail-Entscheidungen, Konfidenzeinschätzungen und die subjektive Risikowahrscheinlichkeit für die Originalstudie sowie für die Simulationsstudie (jeweils für die Kontroll- und die Treatmentgruppe).

Tabelle 1: Treatmentunterschiede[41]

	Ohne Warnung (Kontrollgruppe)		**Mit Warnung (Treatmentgruppe)**	
	Menschen (Original, N=67)	LLM (Simulation, N=268)	Menschen (Original, N=67)	LLM (Simulation, N=268)
Bail	.70	.72	.70	.72
Konfidenz	.92	1.34	.89	1.22
Rückfälligkeit	–.15	–.42	–.09	–.27

Die Anteile der Pro-Bail-Entscheidungen mit und ohne Warnungen sind in der Originalstudie deskriptiv sehr ähnlich; die Bewertung der Konfidenz fällt ebenfalls ähnlich aus. Schließlich wird die Rückfallwahrscheinlichkeit in der Treatmentgruppe nur unwesentlich höher eingeschätzt als in der Kontrollgruppe. Signifikante Unterschiede zwischen Kontroll- und Treatmentgruppe lassen sich für keine der abhängigen Variablen nachweisen (t-Tests, N=134, ps > .52). Auch unter Verwendung der LLM-Simulation sind die absoluten Werte zwischen den Versuchsanordnungen ähnlich. Für die (abschließende) Bail-Entscheidung lassen sich keine

41 Für Bail wird der Anteil der Pro-Bail-Entscheidungen angegeben. Die Konfidenz und Rückfallwahrscheinlichkeit wurden auf Skalen von -2 bis 2 (in 5 Ausprägungen) erhoben. Angegeben werden Durchschnittswerte.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

signifikanten Unterschiede mit und ohne Warnung nachweisen (t-Test, N=536; p = .75). Die deskriptiv zum Teil eher geringen Unterschiede in der Konfidenz und Rückfallwahrscheinlichkeit sind jedoch statistisch signifikant (t-Tests, N=536, ps < .01). Dabei ist auch zu berücksichtigen, dass die Beobachtungszahlen in der Simulationsstudie deutlich höher sind als in der Originalstudie.

2. Unterschiede zwischen den Erhebungsmethoden

Der Anteil der Pro-Bail-Entscheidungen der menschlichen Teilnehmer:innen und der KI-Agenten ist in beiden Experimentalgruppen mit gerundet 70 % bzw. 72 % sehr ähnlich – entsprechend findet sich kein statistisch signifikanter Unterschied zwischen beiden Erhebungsmethoden (t-Tests, N=335, ps >.25). Allerdings ist die Bewertung der Konfidenz für die KI-Agenten sowohl mit als auch ohne Warnungen deutlich höher als die der menschlichen Versuchspersonen (t-Tests, N=335, ps < .01). Die Rückfallwahrscheinlichkeit wird von den KI-Agenten in beiden Gruppen leicht, aber signifikant niedriger eingeschätzt (t-Tests, N=335, ps < .01).

3. Modellspezifische Unterschiede

Die bisherige Analyse beruht auf „gepoolten" Werten, in die sowohl die Simulation durch Claude als auch die von GPT generierten Antworten eingeflossen sind. Während ein solches Vorgehen eine gewisse Robustheit gegen Idiosynkrasien einzelner KI-Modelle verspricht, kann es die Heterogenität verdecken, die mit der Nutzung unterschiedlicher Modelle einhergehen kann. Vor diesem Hintergrund wurden die gepoolten Ergebnisse aufgelöst. Die modellspezifischen Werte sind in Tabelle 2 dargestellt (zu Vergleichszwecken enthält diese wiederum auch die Werte der Originalstudie).

Tabelle 2 zeigt zunächst, dass das Antwortverhalten der Agenten zwischen den beiden Modellen teilweise unterschiedlich ist. Die Bail-Entscheidungen aber fallen zwischen beiden Modellen schon deskriptiv sehr ähnlich aus. Für den Anteil der Pro-Bail-Entscheidungen lassen sich weder in der Kontrollgruppe noch in der Treatmentgruppe signifikante Unterschiede zwischen den LLMs nachweisen (t-Tests, N=268, ps > .26). Dagegen fallen die Bewertungen der Konfidenz und der Rückfallwahrscheinlichkeit zwischen beiden Modellen leicht unterschiedlich aus. Das Modell Claude gibt sowohl eine durchschnittlich höhere Rückfallwahr-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Tabelle 2: Modellspezifische Unterschiede[42]

	Ohne Warnung (Kontrollgruppe)			**Mit Warnung (Treatmentgruppe)**		
	Original (N=67)	Claude (N=134)	GPT (N=134)	Original (N=67)	Claude (N=134)	GPT (N=134)
Bail	.70	.73	.71	.70	.73	.71
Konfidenz	.92	1.41	1.27	.89	1.32	1.12
Rückfall-wahrscheinlichkeit	–.15	–.37	–.47	–.09	–.20	–.35

scheinlichkeit als auch eine höhere Konfidenz für Entscheidungen der KI-Agenten in beiden Experimentalgruppen an (t-Tests, N=268, ps < .05).

Allerdings handelt es sich hierbei um Level-Unterschiede der beiden Modelle. Sowohl die Unterschiede zwischen den Experimentalgruppen als auch die Unterschiede zur Originalstudie sind pro Modell qualitativ vergleichbar mit den Ergebnissen der über die beiden Modelle aggregierten Bewertungen. Die Bail-Entscheidungen sind unter Verwendung beider Modelle mit und ohne Warnungen wiederum sehr ähnlich; statistische Unterschiede zeigen sich nicht (t-Tests, N=268, ps > .78). Die Konfidenz ist in beiden Modellen ohne Warnung höher als mit Warnung und die Rückfallwahrscheinlichkeit ohne Warnung niedriger als mit Warnung (t-Tests, N=268, ps < .01). Verglichen mit der Originalstudie finden sich für den Anteil der Pro-Bail-Entscheidungen keine signifikanten Unterschiede (t-Tests, N=201, GPT ps > .52, Claude ps > .18), wogegen die Konfidenz durch beide LLMs signifikant höher und die Rückfallwahrscheinlichkeit grundsätzlich signifikant niedriger eingeschätzt wird (t-Tests, N=201, ps < .01). Die Einschätzung der Rückfallwahrscheinlichkeit durch Claude und in der Originalstudie ist aber in der Treatmentgruppe nur marginal signifikant unterschiedlich, ein t-Test ergibt einen p-Wert von 0.099, eine Regressionsanalyse, die je Treatment ein Modell mit beiden LLMs schätzt (Anzahl der Gruppen = 335), einen p-Wert von 0.065.

42 Für Bail wird der Anteil der Pro-Bail-Entscheidungen angegeben. Die Konfidenz und Rückfallwahrscheinlichkeit wurden auf Skalen von -2 bis 2 (in 5 Ausprägungen) erhoben. Berichtet werden Durchschnittswerte.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

4. Vignettenunterschiede

Die die obigen Ergebnisse bestätigenden – hier aber nicht berichteten – Regressionsanalysen berücksichtigen bereits statistisch, dass Antworten für mehrere Vignetten abgegeben wurden. Im Folgenden sollen die vignettenspezifischen Werte auch grafisch dargestellt werden. Die nachstehenden Abbildungen (2a, 2b sowie 2c) stellen die Ergebnisse der LLMs denen der Originalstudie (Menschen) gegenüber, jeweils für die Kontroll- sowie die Treatmentgruppe. Jede Abbildung deckt dabei eine der drei abhängigen Variablen ab. Es zeigt sich, dass die LLM-Bewertungen über die Vignetten hinweg eine deutlich geringere Varianz aufweisen. Während die LLM-Bewertungen die menschlichen Abbildungen etwa bei den Bail-Entscheidungen im Durchschnitt gut abbilden, weichen die Bewertungen der einzelnen Vignetten deutlicher voneinander ab. Ein ähnliches Bild ergibt sich für die Rückfallwahrscheinlichkeit. Die KI-Agenten geben hier im Durchschnitt leicht niedrigere Rückfallraten an als die menschlichen Versuchspersonen. Allerdings sind die Unterschiede zwischen den Einschätzungen der unterschiedlichen Vignetten bei Menschen deutlich größer als bei den KI-Agenten. Sowohl bei der Bail-Entscheidung als auch bei der Einschätzung der Rückfallwahrscheinlichkeit scheinen die menschlichen Versuchspersonen also extremer auf die Vignetteninformationen (wozu auch die Risiko-Scores gehören) zu reagieren, als die LLMs dies erwarten.

5. Erweiterung der Originalstudie: Richter-Agenten

Auf aggregierter Ebene bilden die Simulationsergebnisse die Ergebnisse der Originalstudie zumindest in der Grundtendenz gut ab (siehe Tabelle 1). Dies gilt insbesondere für die Bail-Entscheidung. Demgegenüber ergeben sich bei der Bewertung der Konfidenz sowie der Rückfallwahrscheinlichkeit qualitative Unterschiede. Schlüsselt man die Ergebnisse in Bezug auf die einzelnen Vignetten auf, so beruht selbst die durchschnittliche Übereinstimmung bei den Bail-Entscheidungen auf einer deutlich unterschiedlichen Verteilung der Einzelbewertungen.

Geht man vor diesem Hintergrund davon aus, dass LLM-gestützte KI-Agenten menschliche Laienentscheidungen zumindest grundsätzlich zutreffend abbilden können, so eröffnen sich weitergehende Forschungsperspektiven. Dem eigentlichen Erkenntnisinteresse der Originalstudie (den Effekt der Warnungen auf Richterinnen und Richter zu ergründen) folgend, könnte man etwa versuchen, das Verhalten echter Richter zu

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

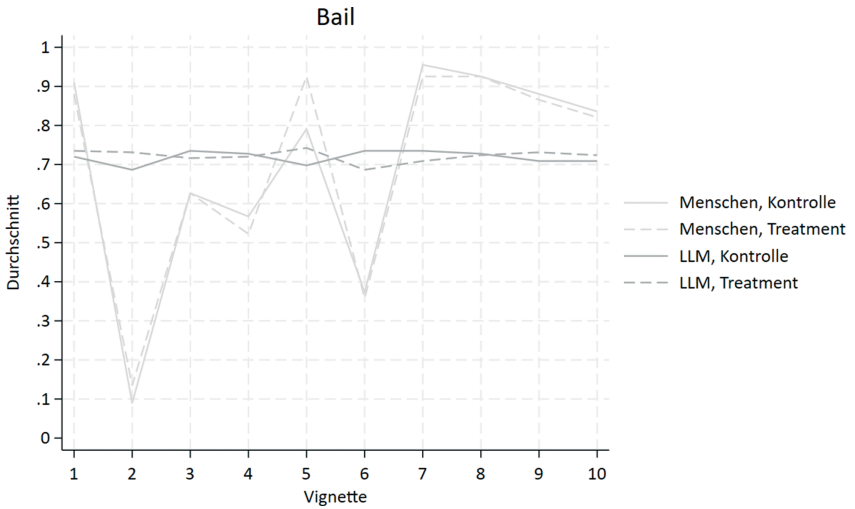


Abbildung 2a: Vignettenunterschiede für Bail-Entscheidung

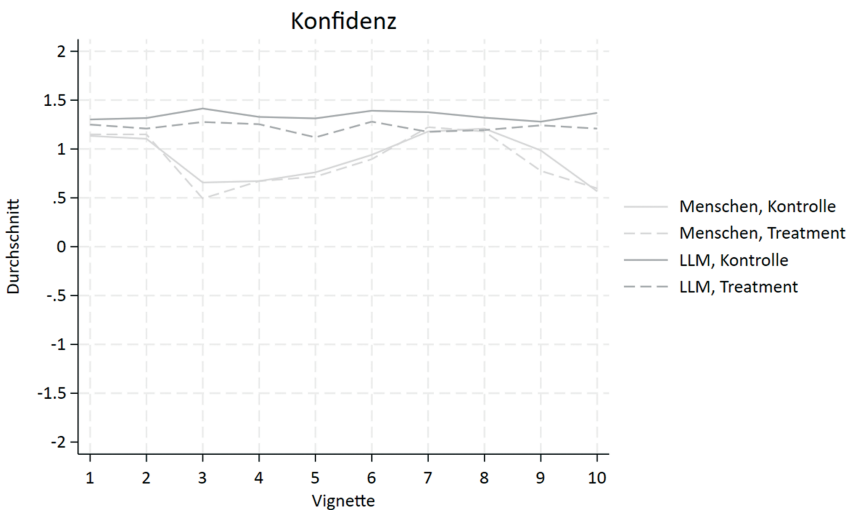


Abbildung 2b: Vignettenunterschiede für Konfidenz

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

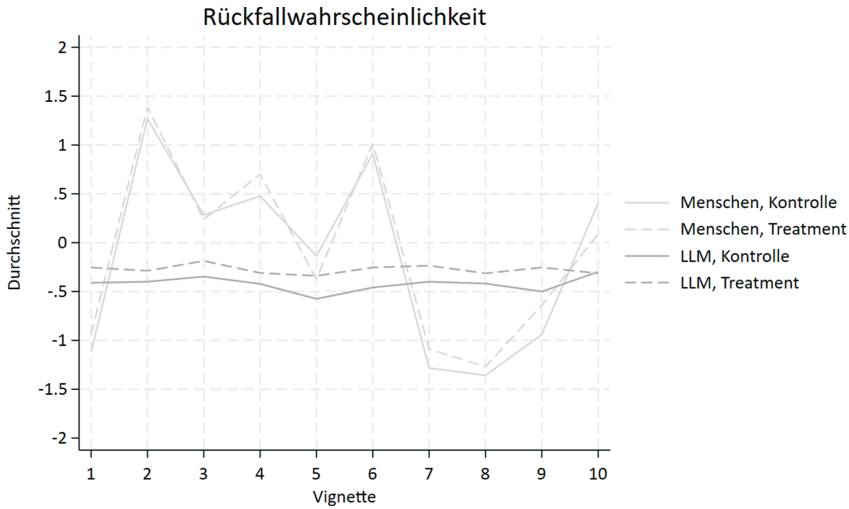


Abbildung 2c: Vignettenunterschiede für Rückfallwahrscheinlichkeit

simulieren. In diesem Sinne wurde im Zug eines Testlaufs ein weiteres Agenten-Sample (mit individuellen demografischen Profilen) erstellt, das nun aber US-amerikanische (Bundes-)Richterinnen und Richter abbilden sollte. Da die Erstellung der Agenten sowie die Bewertung dabei an einigen Stellen von der finalen Simulation abweicht, ist ein Vergleich mit den KI-Agenten als Laien nur begrenzt möglich. Die Untersuchung der Wirkung der Warnungen auf Richter-Agenten ist eher explorativ und wie die gesamten Ergebnisse vorläufiger Natur. Um die darin zum Ausdruck kommenden weitergehenden Forschungsmöglichkeiten zu unterstreichen, sollen die Ergebnisse gleichwohl kursorisch dargestellt werden.

Wie Tabelle 3 zu entnehmen ist, zeigt sich für die KI-Agenten in der Richter-Rolle auf allen drei Bewertungsdimensionen ein signifikanter Effekt der Warnung.

In Gegenwart einer Warnung treffen die Richter-Agenten seltener Pro-Bail-Entscheidungen und sind sich ihrer Einschätzung zudem weniger sicher. Zugleich schätzen sie die Rückfallwahrscheinlichkeit im Durchschnitt leicht geringer ein, als dies ohne Warnung der Fall ist. Dieser etwas überraschende Befund könnte mit den Reaktionen auf einzelne Vignetten zu erklären sein. Zwar ist die Pro-Bail-Rate mit Warnung insgesamt niedriger, eine deutliche Reduktion ergibt sich aber allein für eine spezifische

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Tabelle 3: Simulationsergebnisse „echte" Richter

	Ohne Warnung (Kontrollgruppe) (N=134)	**Mit Warnung (Treatment-gruppe)(N=134)**	**t-Test (p-Wert, N=268)**
Bail	.80	.72	.00
Konfidenz	.88	.80	.04
Rückfallwahrscheinlichkeit	.38	.26	.01

Vignette (Vignette 6). Die Rückfallwahrscheinlichkeit wird mit Warnung grundsätzlich niedriger eingeschätzt. Allein für besagte Vignette 6 zeigt sich ein gegenteiliger Effekt: Hier hat die Warnung eine (spürbare) Erhöhung der angenommenen Rückfallwahrscheinlichkeit zur Folge.

Abschließend soll noch auf eine wesentliche Einschränkung des Einsatzes von KI-Agenten zur Simulation des Verhaltens „echter" Richterinnen und Richter hingewiesen werden. So lässt sich zumindest die Frage aufwerfen, ob das Verhalten von Richterinnen und Richtern in den LLM-Trainingsdaten in einem Ausmaß repräsentiert ist, als dass die Modelle verlässliche Verhaltensannahmen ausbilden könnten.

C. Ausblick

Am Anfang dieser Untersuchung stand die Frage, ob und, wenn ja, mit welcher Vorgehensweise sich rechtsempirische Experimente LLM-gestützt replizieren lassen und was dabei zu beachten ist. Der vorliegende Versuch zeigt, dass konzeptionelle Replikationen mit KI-Agenten durchaus im Bereich des Möglichen liegen. Der wichtigste Befund der Originalstudie (die Warnungen führen nicht zu einem nachweisbaren Effekt auf das Entscheidungsverhalten) konnte mit der vorliegenden Simulation repliziert werden: Wie schon die Originalstudie zeigt auch die Simulationsstudie keine Unterschiede für Bail-Entscheidungen mit und ohne Warnung. Auch in Bezug auf die Konfidenz sowie die Rückfallwahrscheinlichkeit sind die Bewertungstendenzen zwischen Kontroll- und Treatmentgruppe in der Simulation und in der Originalstudie deskriptiv ähnlich. Treatmentunterschiede sind allerdings in der Simulation deutlicher und statistisch signifikant. Insoweit stehen die vorliegenden Ergebnisse durchaus in Einklang mit den Beobachtungen größer angelegter Untersuchun-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

gen, wonach sich eine Vielzahl von Studien mit KI-Agenten grundsätzlich replizieren lässt.

Solche, eher basalen, Übereinstimmungen dürfen freilich nicht zu dem Schluss verleiten, LLM-gestützte KI-Agenten könnten bereits heute als Substitut für menschliche Versuchspersonen fungieren, entsprechende Studien also bereits ersetzen.[43] Dies zeigt auch der vorliegende Replikationsversuch. Betrachtet man dessen Ergebnisse etwa vignettenspezifisch, so treten teilweise erhebliche Abweichungen vom Verhalten der menschlichen Versuchspersonen hervor. Allgemein weist das Antwortverhalten der KI-Agenten im Vergleich zu den menschlichen Versuchspersonen eine geringere Variabilität auf. Dies dürfte zwar teilweise methodisch induziert sein (die Werte pro KI-Agent sind bereits über mehrere Befragungen aggregiert). In erster Linie zeigt sich hier aber wohl eine generelle Eigenschaft oder Schwäche LLM-gestützter Simulationen. So wird in der Literatur über unterschiedliche Studien hinweg berichtet, dass das Antwortverhalten von LLMs im Vergleich zu menschlichen Versuchspersonen eine deutlich geringere Varianz aufweise.[44] Schon aus diesem Grund sind KI-Agenten kein Substitut für menschliche Versuchspersonen und ist bei der Extrapolation der so gewonnen Erkenntnisse Vorsicht geboten. Rechtlich relevante Folgerungen sollte man grundsätzlich erst dann wagen, wenn die Ergebnisse einer LLM-gestützten Simulation durch verhaltenswissenschaftliche Untersuchungen mit menschlichen Teilnehmern untermauert wurden.

Auch und gerade jenseits der Replikationsfrage fördert die vorliegende Studie berichtenswerte Einsichten zu Tage. Für die empirische Rechtsforschung ist insbesondere das Verhältnis zur bisherigen Methodik von Interesse. Ähnlich sind sich beide Wege vor allem auf konzeptioneller Ebene. Anders als man intuitiv annehmen mag, handelt es sich bei LLM-gestützten Simulationen keineswegs um eine Art (rechts-)empirische For-

43 *Harding et* al., AI language models cannot replace human research participants, AI & Society 39 (2024), 2603; *Wang et* al., Large language models that replace human participants can harmfully misportray and flatten identity groups, Nature Machine Intelligence 7 (2025), 400.

44 Vgl. nur *Abdurahman et al.*, Perils and opportunities in using large language models in psychological research, PNAS Nexus 3 (2024), 245; *Saynova et al.*, Identifying Non-Replicable Social Science Studies with Language Models, Preprint arXiv:2503.10671 (2025); *Bisbee et al.*, The Perils of Large Language Models, Political Analysis 32 (2024), 401; *Park et al.*, Diminished diversity-of-thought in a standard large language model, Behav Res Methods 56 (2024), 5754.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

schung light, die man ohne einen theoretisch-konzeptionellen Unterbau aufsetzen könnte. Wie das hier beschriebene Vorgehen zeigt, erschöpft sich eine Simulation nicht in der Formulierung einer Handvoll Prompts. Gedanken über den für eine Forschungsfrage geeigneten experimentellen Aufbau und seine Implementierung (in einem LLM-Setting) muss man sich weiterhin machen. Viele Arbeitsschritte, die bei der Vorbereitung experimenteller Studien anfallen, müssen konzeptionell vergleichbar auch bei LLM-gestützten Experimenten vollzogen werden. So musste vorliegend etwa eine komplexe Pipeline entwickelt und umgesetzt werden, die das Untersuchungsdesign der Originalstudie experimentell sauber abbildet. Dem ging eine Auseinandersetzung mit der Originalstudie voraus.

Relevante Unterschiede (zugunsten LLM-gestützter Simulationen) ergeben sich vor allem unter forschungspraktischen Gesichtspunkten. Obgleich auch LLM-gestützte Simulationen keineswegs kostenlos sind (für die API-Nutzung können zumindest bei größer angelegten Untersuchungen schnell Kosten im hohen dreistelligen Bereich anfallen), fallen die Kosten im Vergleich zu Studien mit Menschen dennoch deutlich geringer aus. Hinzu kommt, dass einige praktische Begrenzungen (etwa mit Blick auf die Verfügbarkeit von Versuchspersonen), die sich nur bedingt kontrollieren lassen, im LLM-Experiment nicht auftreten. So sind Beobachtungszahlen relativ einfach skalierbar. Auch demografische Besonderheiten verfügbarer Versuchspersonen können in der Simulation einfach ausgeglichen werden. Soweit demografische Daten über die für die Forschungsfrage relevanten Gruppen zur Verfügung stehen, lassen sich KI-Agenten mit entsprechenden Eigenschaften ohne größeren Aufwand generieren. Die Erstellung von Richter:innen als KI-Agenten in der vorliegenden Simulation stellt gerade für die empirische Rechtsforschung eine vielversprechende Option dar. Dies macht empirische Rechtsforschung nicht nur zugänglicher (niedrigere Einstiegshürden), sondern auch flexibler. KI-Agenten und LLM-gestützte Simulationen könnten das Feld daher nachhaltig beleben und möglicherweise einen rechtsexperimentellen Innovationsschub auslösen. Forschende können die unterschiedlichsten experimentellen Designs (vorab) ausprobieren, praktisch unbegrenzt (noch so kleinteilige) Anpassungen vornehmen und selbst fernliegende (treffender: bislang als fernliegend erscheinende) Ansätze und/oder Erklärungen testen.

Vor diesem Hintergrund kann der vorliegende Werkstattbericht auch als Plädoyer für eine innovativere und kreativere (empirische) Rechtsforschung verstanden werden – und das dürfte ganz im Sinn des von *Christoph Engel* gelebten Forschungsverständnisses sein.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Can LLMs Replicate Surveys in Empirical Legal Interpretation?

*Jonah B. Gelbach**

A. Introduction

Legal scholars have begun using surveys to identify laypeople's understanding of legally relevant words.[1] This is a potentially important approach to legal interpretation, especially given the enormous impact the concept of "ordinary meaning" ostensibly plays in legal interpretation.[2] As James Macleod writes: "[T]o find public meaning, ask the public."[3]

But surveys are expensive and challenging to conduct. Here's Judge Kevin Newsom of the U.S. Court of Appeals for the Eleventh Circuit: "The survey method is interesting, but it seems wildly impractical – judges and lawyers have neither the time nor the resources to poll ordinary citizens on a widespread basis."[4] So Judge Newsom (in)famously used large language models (LLMs) to inform his own interpretation of contested terms in contract and criminal cases.[5]

* I thank Eric Ling for outstanding research assistance and James Macleod for generously sharing his data and answering my questions about his survey.

1 *See, e.g.*, Kevin P. Tobia, *Testing Ordinary Meaning,* 134 Harv. L. Rev. 726 (2020) and Kevin Tobia, Jesse Egbert & Thomas R. Lee, *Triangulating Ordinary Meaning*, 112 Geo. L. J. Online 23 (2023).

2 *See, e.g.*, Antonin Scalia & Bryan A. Garner, *Reading Law: The Interpretation of Legal Texts* (2012); Brian Slocum, Ordinary Meaning (2015); William N. Eskridge, Jr., Interpreting Law (2016).

3 James A. Macleod, *Ordinary Causation: A Study in Experimental Statutory Interpretation*, 94 Ind. L.J. 957, 961 (2019). In addition, the approach might one day be used to learn about law more generally – it might help with *experimental jurisprudence*. Kevin Tobia, *Experimental Jurisprudence,* 89 U. Chi. L. Rev. 735 (2022).

4 *Snell v. United Specialty Ins. Co.*, 102 F.4th 1208, 1230 (11th Cir. 2024) (Newsom, J., concurring).

5 *Id.*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Meanwhile, researchers have begun investigating whether survey results can be replicated using LLMs.[6] As of September 2025, the leading example within legal scholarship is *Asking GPT for the Ordinary Meaning of Statutory Terms*, by Christoph Engel & Richard H. McAdams (EM).[7] EM asked the GPT3.5 LLM to answer questions about the classic no-vehicles-in-the-park problem that has long motivated discussion in American jurisprudence.[8] They observe that Kevin Tobia's earlier study of whether survey respondents characterize objects such as cars, bicycles, and baby carriages as vehicles can be used as a benchmark – "ground truth" – for measuring LLM query results.[9] EM report results from repeatedly asking GPT3.5, "Is X a vehicle?", where X is each of 25 objects.[10] Their LLM results differ greatly from Tobia's benchmark survey results.[11] EM consider additional prompt-generation approaches, most of which clearly fail. They find that using a Likert scale improves replication, particularly for "objects that human participants consider likely candidates" for being a vehicle;[12] but even the Likert scale results involve visually quite different results patterns.[13] I read EM's article as evidence against the capacity

6 Examples outside legal studies include Peter S. Park, Philipp Schoenegger, and Chongyang Zhu, *Diminished Diversity-of-Thought in a Standard Large Language Model*, 56 Behavior Research Methods 5754 (2024); Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, Michael S. Bernstein, *Generative Agent Simulations of 1,000 People*, November 15, 2024, https://arxiv.org/ abs/2411.10109; Luke Hewitt, Ashwini Ashokkumar, Isaias Ghezae, Robb Willer, *Predicting Results of Social Science Experiments Using Large Language Models*, August 8, 2024, https://www.treatmenteffect.app/paper.pdf; Yaoyu Chen, Yuheng Hu, & Yingda Lu, *Predicting Field Experiments with Large Language Models*, May 21, 2025, https://arxiv.org/abs/2504.01167.

7 2024 U. Illinois J. Law, Tech. & Policy 235 (2024).

8 *See*, *e.g.*, H.L.A. Hart, *The Concept of Law* (1961); Eskridge, *supra* note 2.

9 Tobia, *supra* note 1.

10 EA, *supra* note 7 at 261–263.

11 *Id.* at 263.

12 *Id.* at 269.

13 *See id.* at 271, Fig. 5. The *p*-value from a Kolmogorov-Smirnov test for difference between the Tobia benchmark and Likert-based GPT3.5 results is 0.2978, so that "the null hypothesis that both distributions are indistinguishable can no longer be rejected," *id.* at 270. But assuming equality of human and LLM distributions is not obviously appropriate, so failure to reject this null hypothesis is perhaps not too compelling.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of LLMs to replace surveys as a means for uncovering ordinary/public meaning.[14]

A question beyond the scope of the EM paper is whether LLMs would better replicate ordinary-meaning surveys if LLMs were told to respond as if they were persons with demographic characteristics. For example, let *D* represent the statement: "You are a 63-year-old White male with a master's degree."[15] Then instead of feeding the LLM just a prompt asking about whether a bicycle is a vehicle, we could feed it that prompt after feeding it *D*, so that the LLM predicts the value "yes" or "no" after taking into account whatever linguistic-predictive effect is associated with the words in *D*. To the best of my knowledge, the present study is the first to directly use this approach to benchmark LLM performance in a study related to legal interpretation.[16]

Section B provides a framework for understanding LLMs as a human-survey alternative. Section C then discusses Macleod's original study, on which my benchmarking assessment is based;[17] explains the part of Macleod's study that I attempt to replicate with LLMs; and reports results from this exercise. My results provide cause for pessimism. Section D offers discussion points and a suggestion for future work that might do better.

B. LLM and Human Responses to Linguistic Prompts

LLMs are statistically generated models of language. They translate tokens, which are words or their parts, into numerical vectors. Longer texts are then represented as numerical vectors using mathematical functions that are highly flexible due to their combination of nonlinear functions

14 To be sure, EM consider a great deal more than what I have space to discuss in this short piece, and I encourage readers to engage their article directly.

15 I use "male" and "female" because those are the terms used in the question about gender in the survey whose results I use below.

16 For examples outside the arena of legal studies, *see supra* note 6. For a negative assessment involving ChatGPT's ability to replicate political survey results, *see* G. Elliott Morris and Verasight Data Team, *Your Polls on ChatGPT*, (2025), https://report.verasight.io/synthetic-sampling. For an example of code designed specifically to help researchers implement the approach, *see* the Expected Parrot Domain-Specific Language, discussed at https://docs.expectedparrot.com/en/latest/overview.html.

17 Macleod, *supra* note 3.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

and high-dimensional sets of linearly combined parameters. These longer texts then can be used to base predictions of how humans respond to textual prompts. Consider prompt P_{100}:

Suppose the speed limit is 65 miles per hour and a person is driving 100 miles per hour. Should that person be surprised to be stopped by the police and given a speeding ticket? Answer either 'yes' or 'no.'

Feed P_{100} to an LLM-driven next-token-prediction app, and the app will predict what more likely comes next – "yes" or "no".[18] Whether LLMs can be used in place of surveys for ordinary-meaning research boils down to whether responses to prompts such as P_{100} somehow manage to replicate human answers. If LLM-based answers could do that, it would unlock a world of research (and perhaps judicial) findings that would be accessible without the expense, time, and general logistical burden of conducting surveys.

Generative LLM apps do not *think* or *reason*. All they do is provide next-token-prediction (NTP): given a prompt P, a vocabulary V,[19] and trained parameters $\widehat{\theta}$,[20] a large language model L will provide a (possibly stochastic) response string of tokens, R. This is reflected in Figure 1's simple schematic.[21]

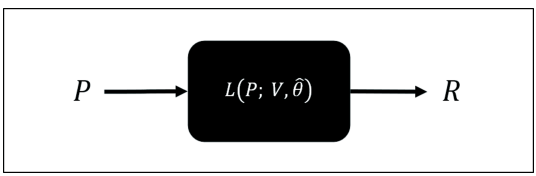


Figure 1: Prompt P is fed into black box, producing output string R

18 In some cases, the LLM might answer P_{100} with something other than "yes" or "no", but that is unusual, and I will ignore it.

19 V typically includes tokens comprising words (e.g., *rain*), word subparts (e.g., *ing*), punctuation, and special objects such as a <STOP> token that leads a word-generating app to terminate the next-token-prediction process.

20 AI researchers use the word *train* where economists and other statisticians traditionally use *estimate*.

21 For a gentle introduction with more detail, *see* Harry Surden, *ChatGPT, Large Language Models, and Law*, 92 Fordham L. Rev. 1941 (2024). For a detailed treatment, *see*, *e.g.*, Christopher M. Bishop with Hugh Bishop, Deep Learning, Springer (2024).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

What do humans do when confronted with that same prompt *P*? I am not a philosopher of cognition, epistemology, language or anything else. But I will assert that humans do something like what Figure 2 depicts: Given a prompt of words *P*, human *H* thinks about the concepts the words represent to *H*, and then, on the basis of *H*'s thoughts, *H* formulates a response, which *H* then utters or writes in words as the string *R′*.[22]

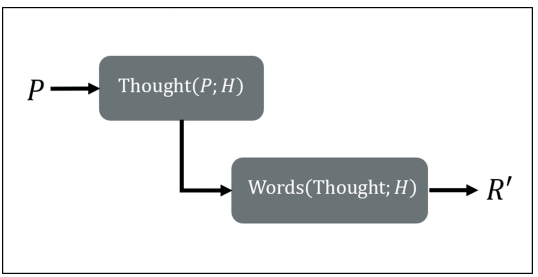


Figure 2: Feeding Prompt *P* Into Human Brain Produces Response R'

Both the Thought(·) and Words(·) functions in Figure 2 have an *H* parameter, symbolically representing that (1) different humans will have different thoughts in response to a given prompt, and (2) different humans may respond with different words even if they have substantively equivalent thoughts about *P*. Whether an LLM can adequately replicate the heterogeneous responses of a set of humans will depend on (a) whether human characteristic information provided to LLMs is enough to capture variation in Figure 2's $Words(\cdot)$ and $Thought(\cdot)$ functions, and (b) how accurately LLM responses vary with such detail.

Accepting Figure 2's model, what does it mean to ask whether LLMs can replicate human responses? It means asking whether we can (1) augment prompt *P* with information about human *H* and thereby (2) reliably obtain LLM response *R* that is in all important respects the same as human response *R′*. With the domain of *R* and *R′* being just "yes" or "no", this question becomes, simply: Should we expect *R* and *R′* to be the same? For-

22 Here I do not mean to assert anything about the order or timing of these activities, nor to take a position on the thought-language relationship; for more on that topic, *see*, *e.g.*, Anna Ivanova, *Can we think without language?*, May 2, 2019, https://mcgovern.mit.edu/2019/05/02/ask-the-brain-can-we-think-without-language/.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

mally, the question is whether there is some prompt-modifying function P^*, such that $L\left(P^*\left(P, H\right); V, \widehat{\theta}\right) = R'$.

Why *would* this happen? Perhaps, given H, the LLM exactly captures what goes on inside the human brain. Then the black box in Figure 1 literally represents the functioning of – essentially, *is* – the two grey ones in Figure 2. I do not believe this *representative functioning hypothesis* is plausible.[23]

But even if what happens inside Figure 2's grey boxes lies beyond the capacity of humans to describe, or even to model explicitly, Figure 1's black box still may be flexible enough to replicate the output of Figure 2's two grey boxes *on average*. Then, using the LLM will lead to results that are in a relevant metric similar to what happens if we consult a group of humans. Call this the *representative response hypothesis* (RRH).[24]

If the RRH is correct, then at least some questions can be answered closely enough using LLM queries as by consulting groups of humans. A survey generates statistics regarding the share of some underlying human population who would respond with R' when asked P. If the RRH is true, then an LLM – which, like a survey, returns statistics[25] – will do an adequate job of replicating the answers particular groups of humans would give to the same survey's question.[26] So if the RRH is true, then, up to sampling variation, the LLM query-response distribution will be similar to what we would get using a survey of a human population of interest.

23 For example, this explanation would require that all humans provided with the given prompt P would respond with identical R' (or with variation that is perfectly captured by the variation in the LLM's own response that is captured by its "temperature" parameter). I think this unlikely.

24 In econometrics terms, the RRH essentially says that for an estimable or known function P^*, L represents the structural relationship in Figure 2 as a reduced form.

25 Thomas R. Lee & Jesse Egbert, *Artificial Meaning?*, April 18, 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4973483, have argued that using LLM queries as EM (and I) do yields not *empirical* information, but rather "artificial intuitions," *id.* at 4. But text generated by LLMs does inarguably return a certain kind of empirical information: it is generated by functions of statistics (the $\widehat{\theta}$ trained parameters described above), which are themselves based on real textual data (tokens used by humans). Thus, LLM query responses are empirical objects in the same qualitative way that the square of a sample mean is.

26 In cases where it is critical to know how a specific human or small subset of humans would respond to a prompt, LLMs will perform poorly unless the relevant human responses are typical of some population that the LLM well represents (formally speaking, H must provide enough information for P^* to distinguish between different types of humans).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

But why should LLMs trained on general data provide accuracy with respect to particular questions of interest? Consider the distinction between the following:

Example 1 A human is prompted: "Answer the following question as if you are a 63-year-old White male with a master's degree. Should a person stopped for driving 74 mph where the speed limit is 65 mph be surprised to be stopped by the police and given a speeding ticket?"

Example 2 A 63-year-old White male with a master's degree is prompted: "Should a person stopped for driving 74 mph where the speed limit is 65 mph be surprised to be stopped by the police and given a speeding ticket?"

Suppose we start from the premise that people's thoughts about the events described in the prompt vary in the population with age, race, gender, and educational attainment. Except by statistical accident, *Example 1* and *Example 2* wouldn't yield similar answers, even on average, unless humans in Example 1 have the capacity to respond *as if* they are 63-year-old White males with master's degrees. In other words, *Example 1*'s prompt must do the work of creating *Example 2*-like people. There is no obvious reason to think that LLMs' training achieves anything like this.

Still, LLMs do well at tasks observers might have doubted, and whether LLMs can be used in place of survey data clearly is of real interest. So it's worth investigating how well LLMs replicate survey responses.

C. Replicating Macleod's "Ordinary Causation" Study

Assessing the value of including demographic information in prompts to refine EA's approach requires a data source with both ground truth human responses to a prompt of interest and demographic information about the respondents themselves. Macleod has generously shared with me just such data from his 2019 article studying how members of the public assess statutory language involving causation in four settings. Three were actual legal cases: *Burrage v. United States*[27] (involving sentencing enhancements for drug crimes), *Gross v. FBL Fin. Servs., Inc.*[28] (employment discrimination), *United States v. Miller*[29] (religious discrimination);

27 571 U.S. 204 (2014).
28 557 U.S. 167 (2009).
29 767 F.3d 585 (6th Cir. 2014).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a fourth was contrived to mimic key elements of *Burrage* while switching the context from drugs lethal to humans to food lethal to plants, to reduce the extent of moral blameworthiness likely involved.[30] Space limitations require me to limit consideration, and I focus only on Macleod's investigation related to *Burrage*.

Macleod gave respondents prompts describing circumstances similar to those in the cases described above, except that he experimentally manipulated the prompts so that respondents were randomly assigned to consider prompts implying the defendant's focal acts were either:

(1) necessary and sufficient to cause death from drugs – the NS condition;
(2) necessary and insufficient for death – the NI condition;
(3) unnecessary but sufficient for death – the US condition; or
(4) unnecessary and insufficient for death – the UI condition.

In his study of the *Burrage* facts, Macleod provided each survey respondent with one arm of the following 4-arm vignette using fictional drug names, with assignment to the arms being random:

> Fintene, Rextor, and Tamphen are dangerous and illegal recreational drugs. A typical dose of each makes users feel high. But unusually potent doses are very dangerous, and can even be deadly. Drug dealers sell them in powder form, in small plastic bags. To take them, users mix the powder into a glass of orange juice and drink it. Taken this way, Fintene, Rextor, and Tamphen take about an hour to become absorbed into the body. On Tuesday morning, Josh met up with three different drug dealers. One sold Josh a dose of Fintene, another sold Josh a dose of Rextor, and the other sold Josh a dose of Tamphen. Josh wanted to have a good time, so he decided that that afternoon, he would take all three of the drugs together. What Josh didn't know was that the Fintene, Rextor, and Tamphen he bought were all unusually potent. In fact, [text from appropriate column in table on next page goes here] taking all three drugs together would kill Josh. That afternoon, Josh went home, combined all the powder from all three bags, mixed it into a glass of orange juice, and drank it. Sure enough, after about an hour, the Fintene, Rextor, and Tamphen combined in Josh's bloodstream, blocking the flow of blood to Josh's heart, and Josh died.[31]

In *Burrage*, the statutory question was whether to apply a provision for increased sentences in cases where a death "resulted from the use of" il-

30 Macleod, *supra* note 3, at 1004.
31 Macleod at 996–997.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

[1] Necessary & Sufficient	[2] Necessary & Insufficient	[3] Unnecessary & Sufficient	[4] Unecessary & Insufficient
… although neither the Rextor by itself, nor the Tamphen by itself, nor the combination of Rextor and Tamphen together would have killed Josh, taking the Fintene by itself would have killed Josh. And…	… although neither the Fintene by itself nor the combination of the Rextor and Tamphen together would have killed Josh, …	… the Fintene by itself would have killed Josh, the Rextor by itself would have killed Josh, and the Tamphen by itself would have killed Josh. Taking any two of the three drugs together would also have killed Josh. And…	… although neither the Fintene by itself, nor the Rextor by itself, nor the Tamphen by itself would kill Josh, taking any two of the three drugs together would kill Josh. And…

legal drugs a person distributed.[32] The Supreme Court held that in the mine run of cases, liability for the sentence-increasing provision is limited to conduct involving drug use that is a "but-for cause of the death or injury,"[33] i.e., a necessary condition of death.[34] Fintene use was necessary to the result of death in arms [1] & [2] of Macleod's vignette, so it was a but-for cause there. The opposite holds for arms [3] & [4]. Macleod asked respondents "whether Josh's death 'resulted from the use of' the Fintene,"[35] as well as whether Josh would have remained alive had he not used Fintene.[36] If ordinary people understand "Josh's death resulted from the use of Fintene" to be the same as "Josh would not have died had he not used Fintene," then their answers to these two questions should be the same.

32 21 U.S.C. § 841(b)(1)(C).

33 571 U.S. 218–219.

34 The Court left open the question of how to interpret the statutory text in cases "where use of the drug distributed by the defendant is … an independently sufficient cause of the victim's death or serious bodily injury." *Id.* at 219.

35 *Id.* at 997.

36 *Id.* at 998.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I. Macleod's results

Figure 3 plots Macleod's results for statutory causation – the share of survey respondents who answered "Yes" when asked whether Josh's death did "result from the use of the Fintene". Even among those assigned the "unnecessary" causation conditions US and UI, where but-for causation is absent, 92 % and 55 % (combined, 73 %) ascribed statutory causation; these numbers are barely distinguishable from the 87 % and 57 % figures (combined, 72 %) for those assigned the "necessary" conditions NS and NI. In Macleod's words: "This finding directly contradicts the courts' pronouncements about ordinary meaning and common understanding."[37] Additionally, "the presence or absence of but-for causation appeared to make far less difference in statutory causation ascription than did the presence or absence of independent sufficiency,"[38] with the sufficiency categories (NS, US) having a combined statutory causation ascription rate of 90 % compared to 56 % for the insufficiency categories (NI, UI).

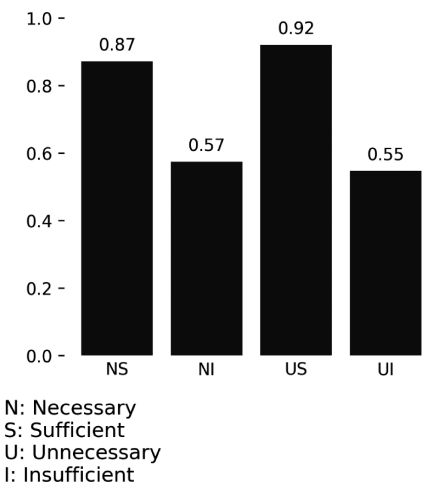


Figure 3: Macleod Results for Percentage. Ascribing Causation Using Statutory Language

37 *Id.* at 1000.
38 *Id.*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Demographics and LLM results

Turning to the demographic variables, here are basic facts about respondents in Macleod's study who were assigned the *Burrage*-based vignette:

- To one decimal point, 50 % were female, 50 % male.
- 23 % had a high school diploma, 9 % had less than that; 22 % had some college but no degree; 12 % had a 2-year degree; 22 % had a 4-year degree; 9 % had a master's degree; and 3 % a doctorate or professional degree.
- 15 % were Hispanic; 85 % not.
- 76 % were White; 12 % Black; 5 % Asian; 1 % American Indian, Alaska Native, Native Hawaiian or Pacific Islander; and 6 % some other racial category.
- Age averaged 48 years; ranged from 18 to 84; and had standard deviation 17.

Each combination of these variables that was present in Macleod's data constitutes a *persona*. Prompts sent to LLMs began by instructing the LLM of the persona, e.g., "You are a 63-year-old White male with a master's degree." The prompts then included Macleod's vignette text for the assigned causation-condition arm of the *Burrage* case,[39] followed by Macleod's questions about causation and respondents' confidence in their own answers.[40] I sent the prompt to LLMs once for each persona; results below are averages weighted to account for the number of times each persona appears in the sample.

Figure 4 reports the share of LLM responses ascribing statutory causation to Fintene for each of the NS, NI, US, and UI vignette arms, and each of 7 LLMs. In each plot, the black bars are the corresponding proportions from Macleod's sample (those reported in Figure 3), and the gray bars are the LLM's proportions.[41] Contrary to Macleod's human survey results, virtually all the LLM personas found the statutorily required causal condition met for arms [1], [2], and [3]. LLM responses also in-

39 Macleod provided me with the full survey wording; 508 respondents were assigned to the *Burrage*-like vignette.

40 Including these questions was necessary to replicate the structure of Macleod's survey instrument.

41 By row/column of the figure, the LLM models are: GPT 4o 2024-08-06 (1/2); GPT 4o mini 2024-07-18 (2/1); o3-mini (2/2); Claude 3.5 Haiku 20241022 (3/1); Claude 3.7 Sonnet 20250219 (3/2); Gemini 1.5 flash (4/1); Gemini 1.5 Pro (4/2).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

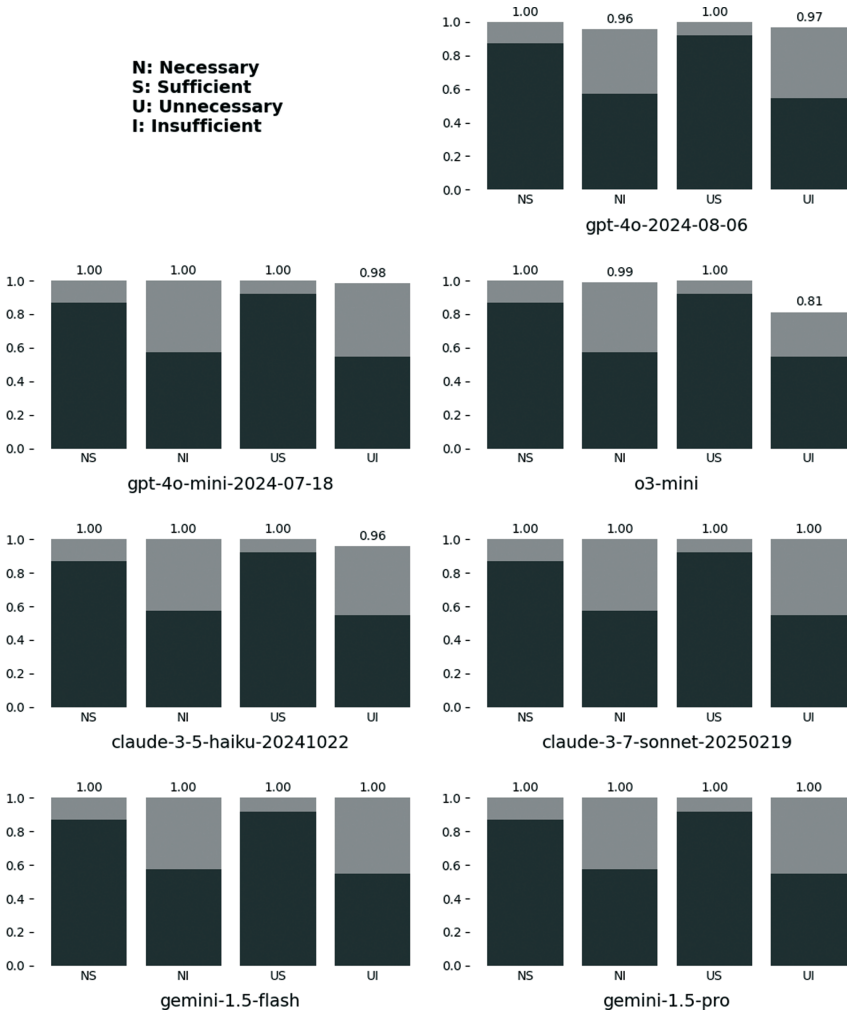


Figure 4: Percentage of LLM Responses Ascribing Statutory Causation

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

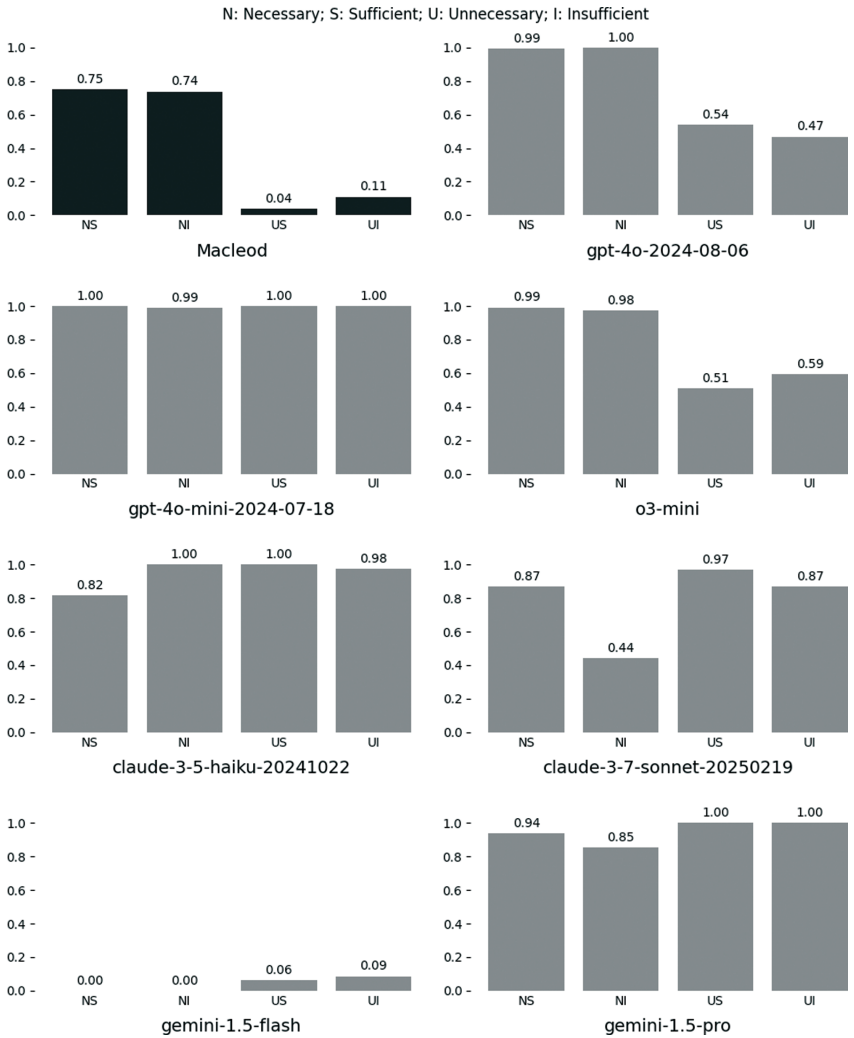


Figure 5: But-For Causation Ascription in *Burrage* Vignettes

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dicated statutory causation was met in arm [4] much more frequently than the human responses did: whereas about 60 % of Macleod's human respondents said Josh's death "result[ed] from" his use of Fintene, the corresponding fraction exceeded 80 % for all 7 LLMs and 95 % for 6 of the 7. In sum, across all four of Macleod's causation conditions, LLM personas were much more likely than humans to respond that statutory causation was met.[42]

Figure 5 reports results for Macleod's question about but-for causation.[43] The top-left chart shows that about three-fourths of respondents in the NS & NI conditions, for whom but-for causation was present, did ascribe but-for causation to Josh's Fintene use. Almost none of those assigned to the US and UI conditions ascribed but-for causation. Thus, most of Macleod's humans correctly responded to the but-for causation question.

Results were again quite different for the LLMs. For the NS & NI conditions, the Macleod sample proportions were about 75 %, but 7 of the 14 LLM proportions were 98 % or greater, 2 were exactly zero, and one was less than 50 %. For the US & UI conditions, the Macleod sample proportions were close to zero, but 7 of the LLM proportions were 97 % or greater, and another 5 were at least 47 % The one LLM that does well for the US & UI conditions, gemini-1.5-flash, fails miserably at the NS & NI conditions.[44]

In sum, results in Figure 4 and Figure 5 indicate that even when demographic personas are used, the 7 LLMs considered here do a poor job of replicating not only Macleod's human-respondent statutory causation results, but also his but-for causation results.[45]

42 For the LLMs that allowed control of the temperature parameter, I found that setting it to 0 yielded virtually no variation in statutory causation ascription. I therefore set it to 1 to increase the possibility that the LLMs would replicate the variation across causation conditions among Macleod's human respondents.

43 The question text was: "Would Josh still have died on Tuesday if he had used the Rextor and Tamphen, but not the Fintene?" Note that this phrasing uses only ordinary terms, avoiding the term of art "but-for".

44 It seems this model just answers the but-for causation question negatively, regardless of the vignette arm used.

45 Macleod's survey included questions about how confident respondents were about their causation-related answers. Human respondents were generally quite confident, and the LLM responses to the same questions mostly met or exceeded the stated levels of confidence.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

D. Discussion

EM were the first legal scholars to investigate whether LLMs replicate human survey results about legal interpretation questions. At least for GPT 3.5 as applied to versions of the classic no-vehicles ordinance, their results suggested the answer was no. Using newer LLMs, the present paper investigates whether providing demographic information about survey respondents can reverse that result. As discussed in section 0, LLMs aggregate information about language-use patterns into mathematical functions. Though LLM-based chat apps can be stunningly good at some tasks, their nature gives no reason to think they replicate human thought. Thus, it is non-obvious why prompting an LLM with words to the effect, "You are a Type-X person" would cause it to generate text that mirrors the answers actual Type-X people give. Nevertheless, the fundamental question of interest here is empirical, so it must be addressed with data.

My empirical findings indicate that the 7 LLMs I queried aren't reasonable alternatives to human surveys about the key aspects of Macleod's survey on causation and legal interpretation, even when I use the available demographic information about Macleod's respondents. To be sure, my study was not designed to be comprehensive. I didn't, for example, query all available LLMs. Still, my results throw more cold water on the notion that LLMs reliably yield answers about legal interpretation questions similar to what representative samples of humans would provide.

Still, perhaps it is too early to give up. First, as LLMs' parameter size grows, performance might rise. Second, studies show LLM queries may do better at replicating human-survey responses when prompts include transcripts from detailed human interviews rather than simple demographic facts like the ones available for this paper. For example, one study used interviews including life story questions such as "Tell me the story of your life – from your childhood, to education, to family and relationships, and to any major life events you may have had" and current-events questions such as "How have you responded to the increased focus on race and/or racism and policing?", yielding interview transcripts averaging 6,491 words; these were included in LLM prompts.[46] This study reported substantial accuracy improvements relative to using demographics only when

46 *See, e.g.*, Park et al., *Generative Agent Simulations of 1,000 People*, *supra* note 6 at 3.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

administering the General Social Survey (GSS) to LLMs.[47] That suggests people's textual answers to detailed interview questions are statistically associated with the answers they give to surveys such as the GSS. In the formal notation *supra*, it suggests that L may be an adequate reduced form representation provided that $P^*(P, H)$ is constructed as the concatenation of person H's detailed interview transcript and the underlying survey prompt of interest, P.

Perhaps this association extends to the ways people answer questions relevant to legal interpretation as well. Assessing that conjecture would be a natural next inquiry.

47 *Id.* at 5–6; the GSS is "widely used across … social sciences to assess respondents' demographic backgrounds, behaviors, attitudes, and beliefs on a broad range of topics, including public policy, race relations, gender roles, and religion," *id.* at 5.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Verfahrensfairness ohne Menschen? Evidenz aus Human- und LLM-Experimenten

Yoan Hermstrüwer & Alexander Stremitzer

1. Einleitung[1]

Algorithmische Entscheidungssysteme gewinnen in der Rechtsprechung zunehmend an Bedeutung (Stevenson & Doleac 2019; Kleinberg et al. 2018; Kehl et al. 2017; Barocas et al. 2019; Pasquale 2015). Trotz ihres Potenzials zur Verbesserung richterlicher Entscheidungen bleibt ihr Einsatz hochumstritten (Mayson 2018; Završnik 2020). Kritiker verweisen auf das Risiko, dass Algorithmen aufgrund geschützter Merkmale diskriminieren könnten (Dressel & Farid 2018; Corbett-Davies et al. 2017), und mahnen mangelnde Verfahrensfairness an (Re & Solow-Niederman 2019; Binns 2022). Diese Einwände sind sowohl für die Rechtsanwendung als auch für die Gewährleistung des Gleichbehandlungsgrundsatzes relevant (Hermstrüwer 2020), denn Verfahrensfairness fördert nachweislich die Gesetzesbefolgung (Tyler 1997; Tyler & Huo 2002).

Bereits heute kann KI in bestimmten juristischen Aufgaben mit menschlichen Experten mithalten oder sie übertreffen (Jung et al. 2020; Aletras et al. 2016; Katz et al. 2017). Gleichwohl zeigen Studien, dass KI-basierte Entscheidungen oft als weniger fair empfunden werden (Burton et al. 2020; Bigman & Gray 2018). Dieses Phänomen – die sogenannte Mensch-KI-Fairness-Lücke – entsteht im direkten Vergleich rein menschlicher mit rein algorithmischer Entscheidungsfindung (Chen et al. 2025; Dietvorst et al. 2015; Mahmud et al. 2022).

In der Praxis agieren Algorithmen allerdings selten ohne menschliche Aufsicht (Selbst 2021; Engstrom et al. 2020). Bestimmungen wie Art. 22 DSGVO und Art. 14(1) der EU-KI-Verordnung schreiben eine menschliche Beteiligung oder Überwachung von Hochrisiko-KI-Systemen vor (Wachter et al. 2017; Veale & Borgesius 2021). Die menschliche

1 Für hervorragende Assistenz bei der Durchführung des LLM-Experiments danken wir Cara Cerny.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Aufsicht dient dabei mehreren Zwecken: der Korrektur algorithmischer Fehler (Lehr & Ohm 2017), der Gewährleistung der Integrität des Begründungsprozesses (Liu & Li 2025) sowie der Wahrung von Autonomie und Würde der Betroffenen (Danaher 2016; Wischmeyer 2020).

Empirische Befunde zeigen, dass die Bereitstellung von Anhörungsrechten und Begründungen die Fairness-Lücke verringern kann (Chen et al. 2022; Henning & Langenbach 2024). Eine weitere Studie legt nahe, dass hybride Systeme, die algorithmische Unterstützung mit menschlicher Bewertung kombinieren, als fairer wahrgenommen werden als rein menschliche Entscheidungen (Hermstrüwer & Langenbach 2023). Unklar bleibt jedoch, wie intensiv die menschliche Überprüfung sein muss, um diesen Effekt zu erzielen.

In diesem Beitrag kombinieren wir die Befunde eines mit Menschen durchgeführten Vignetten-Experiments (Chen et al. 2025) mit zwei Replikationsstudien, in denen wir große Sprachmodelle (LLMs) als Probanden einsetzen. Im Humanexperiment untersuchen wir mit 4.817 US-Teilnehmern die Robustheit der Mensch-KI-Fairness-Lücke und analysieren, wie unterschiedliche Intensitäten menschlicher Aufsicht die Fairnesswahrnehmungen beeinflussen. In den LLM-Experimenten replizieren wir die Studie mit vier Modellen: Claude Sonnet 4, DeepSeek-V3, Gemini 2.5-Flash und GPT-4o.

Die Ergebnisse unseres Humanexperiments bestätigen die Existenz der Mensch-KI-Fairness-Lücke: Rein algorithmische Entscheidungen werden als weniger fair wahrgenommen als rein menschliche Verfahren. Zugleich stellen wir fest, dass hybride Prozesse – selbst bei minimaler menschlicher Beteiligung – die Lücke vollständig beseitigen können. Sie werden als ebenso fair angesehen wie rein menschliche Prozesse. Überraschenderweise erweist sich die bloße Tatsache menschlicher Überprüfung als wichtiger als deren Intensität.

Die Ergebnisse unserer beiden LLM-Experimente scheinen diesen Befund auf den ersten Blick zu bestätigen. Aggregiert man die Daten über alle Modelle und Entscheidungssituationen, ergibt sich qualitativ ein ähnliches Gesamtbild. Bei genauerer Betrachtung treten jedoch fundamentale Unterschiede zutage, die bedeutende methodische Schwierigkeiten offenbaren.

Der Gang der Untersuchung ist folgender: Zunächst beschreiben wir das experimentelle Design unseres Humanexperiments und unserer LLM-Experimente (Abschnitt 2). Es folgen eine Auswertung und Diskussion unserer Ergebnisse (Abschnitt 3). In der Schlussbetrachtung diskutieren

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

wir die rechtlichen und methodischen Implikationen unserer Befunde (Abschnitt 4).

2. Experimentelles Design

Wir führen drei Vignetten-Experimente durch: eines mit Menschen und zwei mit LLMs. Alle drei Experimente zielen darauf ab, die wahrgenommene Verfahrensfairness in drei verschiedenen Szenarien zu untersuchen.

Humanexperiment. Die Teilnehmer wurden über Prolific in den USA (N = 4.817) rekrutiert und zufällig einer von fünf Bedingungen zugeordnet, die mit einem von drei Szenarien gekreuzt wurden (5x3 between-subjects design). Aufbauend auf einer unserer Vorgängerstudien (Chen et al. 2022) las jeder Teilnehmer eine stilisierte Vignette über ein Szenario betreffend die Verhängung des Strafmaßes, den Erlass einer strafprozessualen Haftentscheidung oder einen Fall zur Verbraucherschiedsgerichtsbarkeit. Im Strafmaßszenario bewertet ein Richter die Schuld des Angeklagten John und entscheidet, ob eine härtere oder mildere Strafe zu verhängen ist. Im Haftszenario schätzt ein Richter die Flucht- und Wiederholungsgefahr des Beschuldigten John ein und entscheidet über die Freilassung gegen Kaution. Im Verbraucherschiedsgerichtsszenario stellt der Schiedsrichter fest, ob es einen Schmierfleck auf der Linse einer Kamera gibt, die der Verbraucher John auf einer Online-Plattform gekauft hat, und entscheidet, ob der Kaufpreis zu erstatten ist.

Wir testen fünf Bedingungen. In den beiden Bedingungen mit rein menschlicher Entscheidungsfindung, Human High und Human Low, wird die Entscheidung immer ausschließlich von einem Menschen getroffen. Die Bedingungen Human High und Human Low unterscheiden sich in der Intensität, mit welcher der menschliche Entscheidungsträger den Fall prüft und die Beweise sichtet. Diese Treatmentvariation berücksichtigt die begrenzte Zeit, die (Schieds-)Richter haben, um die Beweise zu bewerten. Asylanhörungen zum Beispiel dauern in der Regel nicht länger als einige Minuten (Political Asylum/Immigration Representation Project 2016); dasselbe gilt für Anhörungen im Haftverfahren (Stevenson 2018). Während der Mensch in der Human High-Bedingung immer „eine gründliche Prüfung des Falles" und eine „gründliche" Überprüfung der Beweise durchführt, führt er in der Human Low-Bedingung nur „eine kurze Prüfung des Falles" durch und überprüft die Beweise nur „kurz".

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Am entgegengesetzten Ende des Mensch-KI-Spektrums implementieren wir eine Robot-Bedingung, in der die Entscheidung immer von einem Algorithmus getroffen wird. In dieser Bedingung führt ein Algorithmus eine „gründliche Prüfung des Falles" und eine „gründliche" Überprüfung der Beweise durch. In den zwei hybriden Bedingungen, Hybrid High und Hybrid Low, interagiert ein menschlicher Entscheidungsträger mit einem algorithmischen Entscheidungsassistenzsystem, das eine Empfehlung abgibt. Auch hier besteht der Unterschied zwischen den Hybrid High- und den Hybrid Low-Bedingungen in der Intensität, mit der der Mensch den Fall prüft und die Beweise sichtet (gründlich oder kurz). In jedem Szenario und jeder Bedingung messen wir die wahrgenommene Verfahrensfairness auf einer 7-Punkte-Likert-Skala mit sieben Labels („stimme überhaupt nicht zu" = 1, „stimme voll und ganz zu" = 7). Zusätzlich erheben wir Daten zur wahrgenommenen Ergebnisfairness, Gründlichkeit, Transparenz, Empathie, Genauigkeit und Effizienz.

LLM-Experimente. Können große Sprachmodelle menschliches Verhalten vorhersagen? Und inwiefern lassen sich große Sprachmodelle als Substitut menschlicher Probanden einsetzen? Zur Beantwortung dieser beiden Fragen replizieren wir unser Vignetten-Experiment mit mehreren LLMs (Claude Sonnet 4, DeepSeek-V3, Gemini 2.5-Flash und Open AI GPT-4o).[2] Über die jeweilige Schnittstelle (API) werden jede Bedingung und jedes Szenario in je 25 Iterationen in jedes der vier Modelle eingespeist. Alle Iterationen werden, soweit verfügbar, bei einer Temperatur von 0 und ohne festen Seed implementiert. Pro Experiment erheben wir 375 individuelle Beobachtungen je LLM; insgesamt umfasst unsere Stichprobe damit über beide Experimente hinweg 3.000 individuelle Beobachtungen.

Das erste LLM-Experiment zielt auf die Vorhersage (prediction) menschlicher Entscheidungen. Der entsprechende Prompt enthält neben der Aufforderung zur Prognose auch Informationen zur soziodemographischen Struktur des Probandenpools aus unserem Humanexperiment. Das zweite LLM-Experiment ist auf die Generierung eigener Fairnessbewertungen (decision) der unterschiedlichen Verfahren durch das jeweilige Modell gerichtet. Wir führen also ein Prediction-Experiment und ein Decision-Experiment durch.

2 Die Prompts und Python-Skripten stellen wir auf Anfrage gerne zur Verfügung.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

3. Ergebnisse

Treatment-Effekte. Die Treatment-Effekte unseres Humanexperiments bekräftigen unseren früheren Befund: Roboter-Richter werden als weniger fair wahrgenommen als menschliche Richter (Abb. 1a). Es zeigt sich eine deutliche KI-Mensch-Fairness-Lücke. Paarweise Vergleiche der Bedingungen bestätigen, dass die Roboter-Bedingung zu signifikant niedrigeren Fairness-Bewertungen führt als die menschlichen Bedingungen (Wilcoxon-Rangsummentests, $p < 0.001$).

Kann diese Mensch-KI-Fairness-Lücke im richterlichen Kontext durch die Einführung menschlicher Aufsicht über den Roboter-Richter gemindert werden? Zunächst stellen wir fest, dass menschliche Beteiligung an der algorithmischen Entscheidungsfindung für die Verfahrensfairness wichtig ist. Diese ist in beiden Hybrid-Bedingungen höher als in der Robot-Bedingung (Wilcoxon-Rangsummentests, $p < 0.001$). Wir beobachten indes kaum Unterschiede in den Fairnessbewertungen zwischen den menschlichen und hybriden Bedingungen. Dies stimmt mit der Annahme überein, dass die Menschen empfindlich darauf reagieren, ob die richterliche Entscheidung ausschließlich von einem Algorithmus getroffen wird, nicht aber auf unterschiedliche Intensitäten menschlicher Beteiligung. Bereits ein hybrides Verfahren – also ein Verfahren, in dem

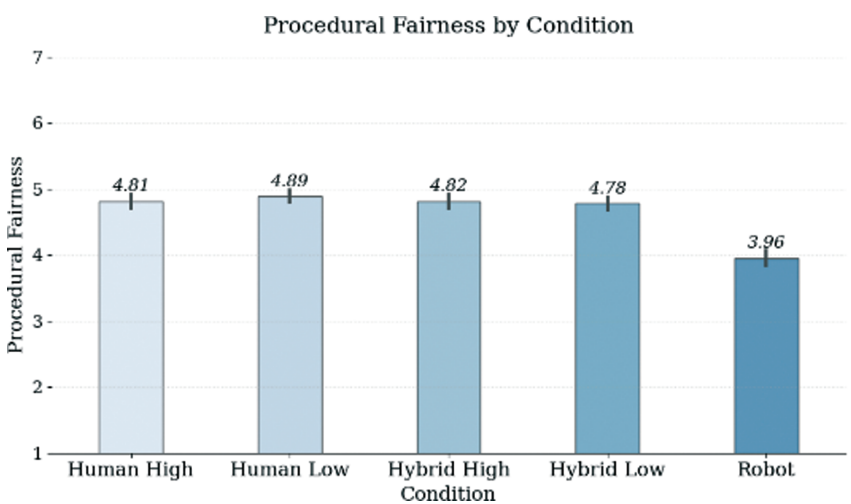


Abb. 1a: Prozedurale Fairnesswahrnehmungen von Menschen

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

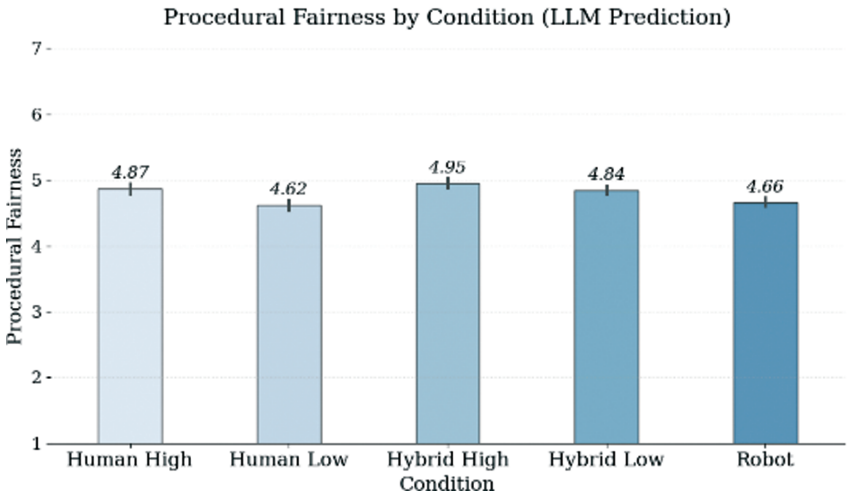


Abb. 1b: LLM-Vorhersagen

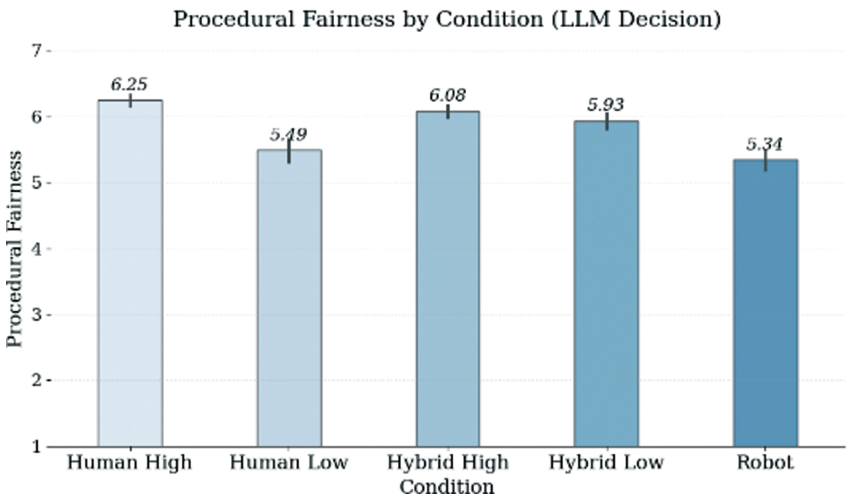


Abb. 1c: LLM-Entscheidungen

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

menschliche Aufsicht über den Algorithmus ausgeübt wird – erscheint ausreichend, um die Mensch-KI-Fairness-Lücke weitgehend zu schließen.

Zur Untersuchung dieser Hypothese analysieren wir, ob die Bewertungen der Verfahrensfairness zwischen den menschlichen und hybriden Bedingungen statistisch äquivalent sind, indem wir das TOST-Verfahren (Two One-Sided Tests) verwenden (Schuirman 1987). Wir setzen die Äquivalenzmarge auf $\Delta = 0.25$; unter der Annahme einer gepoolten Standardabweichung von 1.0 entspricht dies einer geringen Effektgröße von Cohen's $d = 0.25$ (vgl. Cohen 1988). Die Ergebnisse legen nahe, dass die wahrgenommene Verfahrensfairness für alle paarweisen Vergleiche der menschlichen und hybriden Bedingungen äquivalent ist. Dazu gehört der Befund, dass menschliche Verfahren mit hoher menschlicher Beteiligung (Human High) nicht höher bewertet werden als hybride Verfahren mit geringer menschlicher Beteiligung (Hybrid Low), was auf einen Ausgleichseffekt der KI-Fairness hindeutet. Der Übergang von rein menschlichen Verfahren mit einem hohen Maß an menschlicher Beteiligung zu hybriden Verfahren mit einem geringen Maß an menschlicher Beteiligung verringert die wahrgenommene Verfahrensfairness folglich nicht. Umgekehrt erhöht der Übergang von rein menschlichen zu hybriden Entscheidungsverfahren die wahrgenommene Verfahrensfairness aber auch nicht.

Schließlich bleibt die wahrgenommene Verfahrensfairness auch weitgehend unbeeinflusst davon, ob die menschliche Kontrollintensität hoch oder gering ausfällt. Diese Befunde lassen Zweifel an der Vorstellung aufkommen, eine Erhöhung der Intensität menschlicher Aufsicht wirke sich positiv auf die wahrgenommene Verfahrensfairness aus.

Die Ratio experimenteller Befunde ist kausale Inferenz. Doch auch wenn wir den Effekt unserer Bedingungen kausal identifizieren können, bleiben die dem Treatment-Effekt zugrundeliegenden psychologischen Mechanismen weitgehend im Dunkeln. Mithilfe einer Mediationsanalyse lässt sich der Gesamteffekt – definiert als Differenz der Fairnessbewertungen zwischen der Robot-Bedingung und den menschlichen Bedingungen – in direkte und kausal vermittelte Effekte zerlegen (Pearl 2001). Unter der Annahme kausaler Abhängigkeit zwischen Mediatoren schätzen wir ein lineares Strukturgleichungsmodell mit variablen Koeffizienten; Standardfehler werden per Bootstrap ermittelt (Imai und Yamamoto 2013).

In Übereinstimmung mit früheren Befunden (Chen et al. 2022) legt unsere Analyse nahe, dass die Mensch-KI-Fairness-Lücke primär auf Be-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

denken hinsichtlich der Genauigkeit algorithmischer Entscheidungsfindung zurückzuführen ist. Die wahrgenommene Genauigkeit vermittelt etwa ein Viertel des Gesamteffekts, die Gründlichkeit etwa 15 %. Demgegenüber spielen „weiche" Faktoren eine untergeordnete Rolle: Die wahrgenommene Empathie macht nur etwa 8 % aus, die wahrgenommene Transparenz etwa 2 %. Die wahrgenommene Effizienz erweist sich als irrelevant für die Fairnessbewertungen. Die Annahmen der Mediationsanalyse sind streng, und ihre Befunde lassen kausale Identifikation nicht ohne Weiteres zu. Dennoch stützt unsere Analyse die These, dass die Mensch-KI-Fairness-Lücke besser durch „harte" Faktoren wie Genauigkeit als durch „weiche" Faktoren wie Empathie zu erklären ist.

Anders liegen die Dinge in unseren beiden LLM-Experimenten (Abb. 1b und 1c). Zwar beobachten wir auch hier die niedrigsten Fairnessbewertungen in der Robot-Bedingung. Allerdings erweist sich die Mensch-KI-Fairness-Lücke als deutlich geringer – und das Muster unterscheidet sich quantitativ und qualitativ von den Humandaten.

Im Prediction-Experiment (Abb. 1b) zeigen die LLMs ein bemerkenswert flaches Antwortprofil: Die Mittelwerte variieren lediglich zwischen 4.62 (Human Low) und 4.95 (Hybrid High) auf der 7-Punkte-Skala. Die Robot-Bedingung (M = 4.66) wird dabei sogar geringfügig fairer bewertet als Human Low (M = 4.62). Die höchste Fairnessbewertung erhält nicht etwa die Human High-Bedingung (M = 4.87), sondern Hybrid High (M = 4.95). Insgesamt scheinen LLMs bei der Vorhersage menschlicher Urteile die Intensität der Mensch-KI-Fairnesslücke systematisch zu unterschätzen.

Im Decision-Experiment (Abb. 1c) zeigt sich ein differenzierteres Bild. Hier bewerten die LLMs rein menschliche Verfahren mit hoher Prüfungsintensität am fairsten (Human High: M = 6.25), gefolgt von den hybriden Bedingungen (Hybrid High: M = 6.08; Hybrid Low: M = 5.93). Auffällig ist die Nähe zwischen Human Low (M = 5.49) und Robot (M = 5.34): Die Differenz beträgt lediglich 0.15 Skalenpunkte. Dieser Befund deckt sich mit den Ergebnissen einer unserer Vorgängerstudien, in der vollautomatisierte Verwaltungsverfahren nahezu gleichermaßen fair bewertet wurden wie Verfahren mit nur geringer menschlicher Beteiligung (Hermstrüwer & Langenbach 2023). Aus Sicht der LLMs scheint nicht die bloße Anwesenheit eines Menschen entscheidend, sondern die Intensität seiner Beteiligung – ein Muster, das sich in den Humandaten gerade nicht zeigt.

Zusammenfassend lässt sich festhalten: Während menschliche Pro-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

banden bereits durch minimale menschliche Beteiligung (Hybrid Low) vollständig zufriedengestellt werden und nicht zwischen unterschiedlichen Intensitätsstufen differenzieren, reagieren LLMs sensibler auf Variationen in der Prüfungsintensität, unterschätzen aber zugleich die kategoriale Bedeutung, die Menschen der Mensch-Maschine-Unterscheidung beimessen.

Modellkonsistenz. Rechtsprobleme sind selten wohldefiniert (Engel 2001). Insbesondere Entscheidungen über das Strafmaß zeichnen sich durch hohe Komplexität und erhebliche Unsicherheit über die normativ angemessene Strafe aus. Vor diesem Hintergrund überrascht es nicht, dass sich Unterschiede zwischen den Entscheidungsverfahren gerade dort manifestieren, wo subjektive Bewertungen rechtlich relevanter Tatsachen eine zentrale Rolle spielen. Die modellspezifische Analyse offenbart allerdings erhebliche Heterogenität zwischen den untersuchten LLMs (Abb. 2a und 2b).

Claude Sonnet 4 zeigt im Prediction-Experiment (Abb. 2a) ein nahezu flaches Profil: Die Werte variieren lediglich zwischen 3.78 (Robot, Strafmaß) und 5.12 (Human High, Schiedsgericht). Das Modell prognostiziert nur minimale Treatment-Effekte – die vorhergesagte Mensch-KI-Fairness-Lücke bleibt marginal. Im Decision-Experiment (Abb. 2b) differenziert Claude ebenfalls kaum: Im Schiedsgerichts- und Haftszenario bewertet das Modell alle Bedingungen identisch mit 6.0. Treatment-Effekte zeigen sich ausschließlich im Strafmaßszenario, wo Human High und die hybriden Bedingungen Werte von 5.0–6.0 erzielen, während Human Low (M = 4.0) und Robot (M = 4.0) deutlich abfallen. Dieses Muster entspricht den in Chen et al. (2022) präregistrierten Hypothesen, die sich allerdings im damaligen Humanexperiment nicht bestätigt hatten.

DeepSeek-V3 erzeugt das differenzierteste Antwortmuster – und zugleich den auffälligsten Einzelbefund der gesamten Studie. Im Prediction-Experiment liegen die Werte durchweg im mittleren Bereich (ca. 3.5–5.0), wobei das Modell im Strafmaßszenario die niedrigsten Werte prognostiziert (Human High: 3.54, Robot: 3.90). Im Decision-Experiment zeigt sich eine ausgeprägte Mensch-KI-Fairness-Lücke: Im Schiedsgerichtsszenario werden Human High, Human Low und Hybrid High jeweils mit 7.0 bewertet, während Robot auf 5.0 abfällt. Im Haftszenario ergibt sich ein ähnliches Bild mit einer Lücke von etwa einem Skalenpunkt zwischen den menschlichen/hybriden Bedingungen und der Robot-Bedingung.

Besonders markant ist das Strafmaßszenario: Human High erzielt hier 6.64, während Robot bei 3.0 liegt – eine Differenz von über 3.5 Ska-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

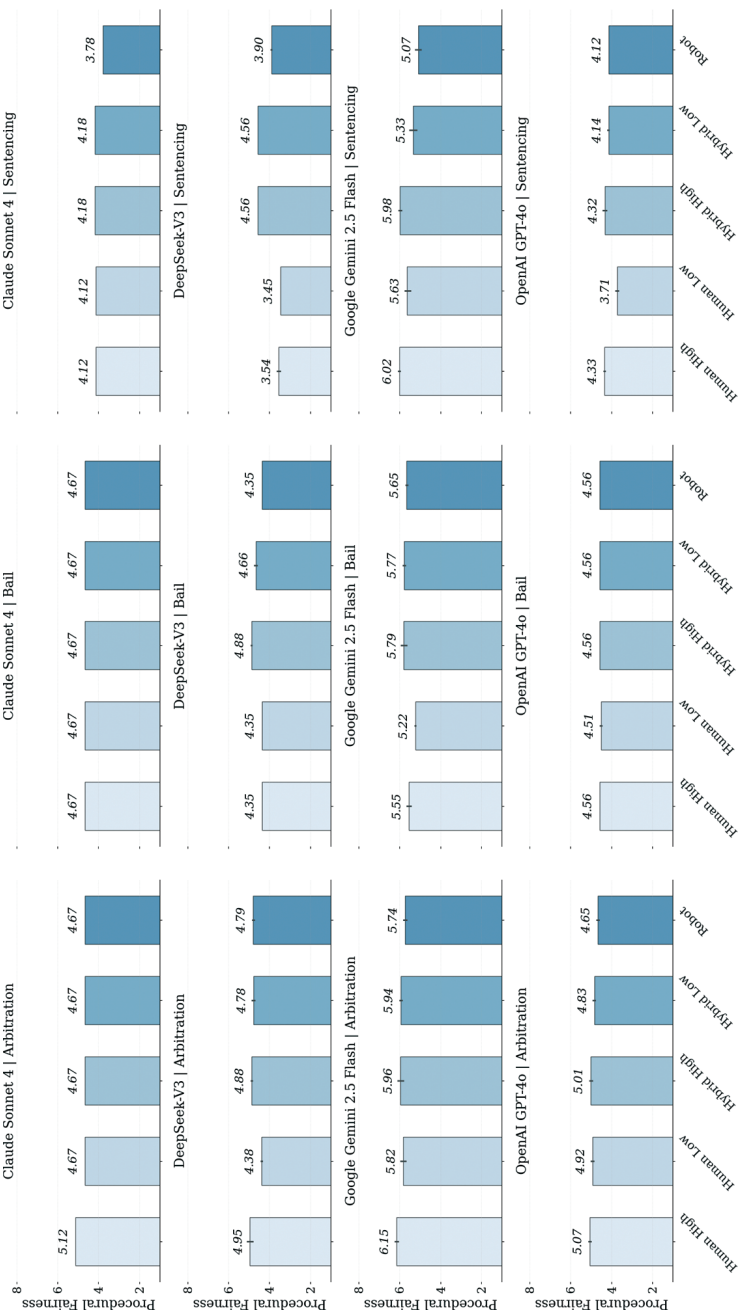


Abb. 2a: LLM-Vorhersagen

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

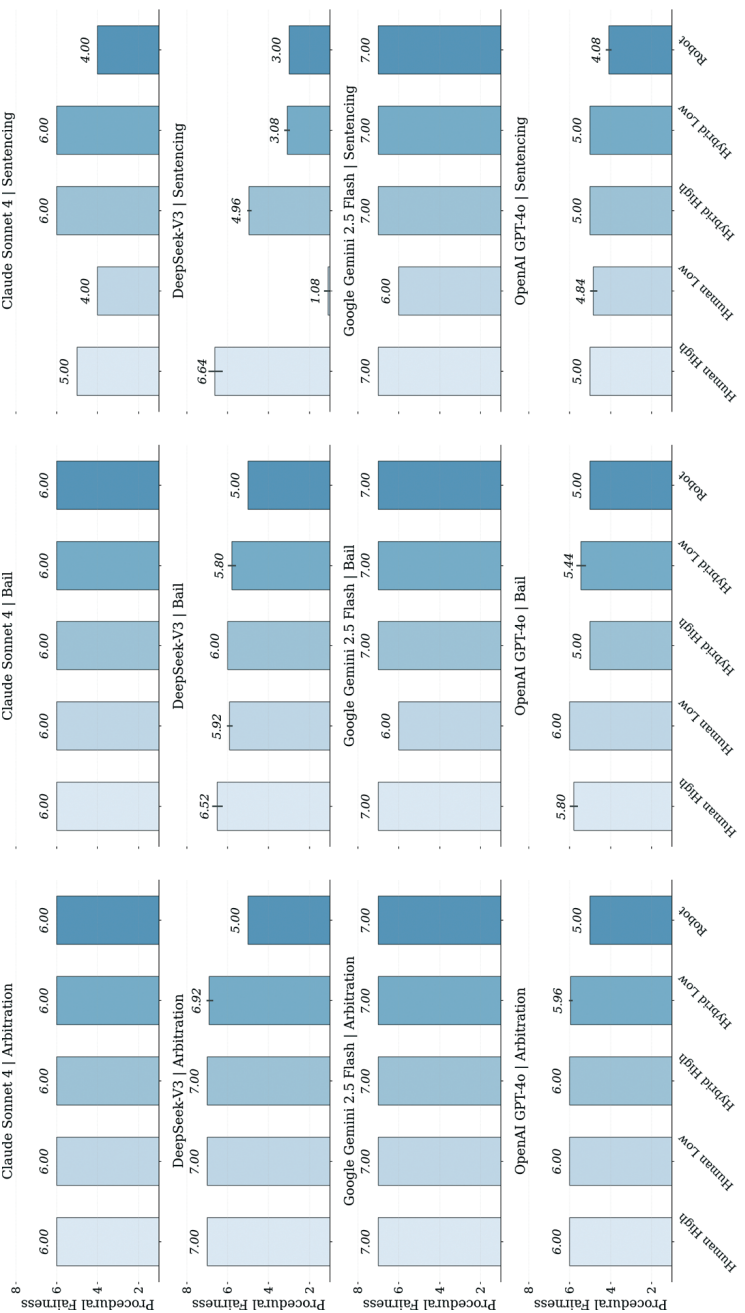


Abb. 2b: LLM-Entscheidungen

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

lenpunkten. Der extremste Befund betrifft jedoch Human Low: Mit einem Wert von nur 1.08 erreicht diese Bedingung den niedrigsten Wert aller Modell-Szenario-Kombinationen in der gesamten Studie. DeepSeek bewertet demnach eine flüchtige menschliche Prüfung bei schwerwiegenden Entscheidungen als nahezu maximal unfair – unfairer noch als eine rein algorithmische Entscheidung. Dies legt nahe, dass das Modell nicht die bloße Anwesenheit eines Menschen als fairnessrelevant erachtet, sondern die Ernsthaftigkeit seiner Befassung. Eine kursorische menschliche Beteiligung erscheint aus dieser Perspektive als Scheinlegitimation, die das Verfahren stärker delegitimiert als vollständige Automatisierung.

Bemerkenswerterweise spiegelt DeepSeek die zentralen Prinzipien des europäischen Regulierungsrahmens damit qualitativ wider. Die risikobasierte Klassifikation ebenso wie das Human-in-the-Loop-Prinzip fungieren als normative Referenzpunkte, die sich in den Fairnessurteilen des Modells niederschlagen. Dabei wird deutlich, dass nicht jede Form menschlicher Beteiligung genügt, sondern vielmehr eine substanzielle menschliche Aufsicht (meaningful human oversight) erforderlich ist.

Google Gemini 2.5 Flash weicht grundlegend von den übrigen Modellen ab. Im Prediction-Experiment zeigt Gemini moderate Differenzierung mit Werten zwischen etwa 4.4 und 6.0, wobei im Strafmaßszenario die niedrigsten Werte auftreten (Human High: 6.02, Robot: 5.07). Im Decision-Experiment hingegen tritt ein Ceiling-Effekt auf: Das Modell bewertet nahezu alle Bedingungen in allen Szenarien mit dem Maximalwert 7.0 – selbst die Robot-Bedingung im Strafmaßszenario. Die einzige Ausnahme bildet Human Low im Haftszenario (M = 6.0). Die ansonsten robuste Mensch-KI-Fairness-Lücke lässt sich mit diesem Modell nicht reproduzieren. Methodisch wirft dieser Befund die Frage auf, ob Gemini bei normativen Bewertungsaufgaben zu affirmativen Antworten neigt oder ob das Modell genuine Indifferenz gegenüber der Mensch-Maschine-Unterscheidung aufweist.

OpenAI GPT-4o reproduziert die Mensch-KI-Fairness-Lücke in moderater, aber konsistenter Form. Im Prediction-Experiment zeigt das Modell ein relativ flaches Profil mit Werten zwischen 3.71 (Human Low, Strafmaß) und 5.07 (Human High, Schiedsgericht). Die vorhergesagten Treatment-Effekte fallen gering aus, wobei die Robot-Bedingung tendenziell etwas niedriger bewertet wird als die menschlichen Bedingungen. Im Decision-Experiment differenziert GPT-4o stärker: Im Strafmaßszenario zeigt sich eine klare Abstufung mit Human High und den hybriden Bedingungen bei etwa 5.0 und der Robot-Bedingung bei 4.08. Im Schiedsgerichtsszenario

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

liegen Human High, Human Low und Hybrid High bei 6.0, während Robot auf 5.0 abfällt. Im Haftszenario fallen die Unterschiede am geringsten aus. Insgesamt zeigt GPT-4o ein Muster, das dem des Humanexperiments am nächsten kommt, wenngleich die absoluten Niveaus variieren.

Die vier Modelle lassen sich hinsichtlich ihrer Sensitivität für die Mensch-KI-Unterscheidung auf einem Spektrum anordnen: DeepSeek reagiert am stärksten auf Treatmentvariationen, gefolgt von GPT-4o und Claude, während Gemini kaum differenziert. Bemerkenswert ist zudem die Diskrepanz zwischen Prediction und Decision: Alle Modelle zeigen im Decision-Experiment stärkere Treatment-Effekte als im Prediction-Experiment – sie unterschätzen also systematisch die Bedeutung, die sie selbst der Mensch-KI-Unterscheidung beimessen, wenn sie menschliches Verhalten vorhersagen sollen.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

4. Schlussbetrachtung

Für die rechts- und justizpolitische Diskussion ergeben sich mehrere Schlussfolgerungen. Erstens bestätigen unsere Humandaten, dass der Einsatz algorithmischer Entscheidungssysteme ohne menschliche Aufsicht unabhängig vom Entscheidungskontext mit erheblichen Fairnesseinbußen verbunden ist. Die Human-in-the-Loop-Anforderungen der DSGVO und der EU-KI-Verordnung finden damit empirische Unterstützung.

Zweitens ist der Befund, dass bereits minimale menschliche Beteiligung ausreicht, um die Mensch-KI-Fairness-Lücke zu schließen, ambivalent. Einerseits deutet er auf erhebliche Kosteneinsparungen hin, ohne Einbußen in der wahrgenommenen Fairness. Hybride Systeme könnten so Vorteile wie höhere Genauigkeit, größere Konsistenz oder eine Reduktion systematischer Verzerrungen bieten, ohne ihre gesellschaftliche Legitimität zu gefährden. Andererseits legt derselbe Befund das Risiko von Humanwashing nahe: Wenn rein formale, kursorische Prüfungen dieselbe Legitimationswirkung entfalten wie substanzielle menschliche Kontrolle, entsteht ein Anreiz für ressourcenschonende, aber faktisch wirkungslose Aufsichtsmechanismen. Die zentrale regulatorische Herausforderung besteht daher darin, substanzielle menschliche Aufsicht von bloß symbolischer Beteiligung zu unterscheiden – eine Differenzierung, die sich in den Fairnesswahrnehmungen der Rechtsunterworfenen offenbar nicht zuverlässig widerspiegelt.

Das zentrale methodische Ergebnis der Studie ist die ausgeprägte Heterogenität zwischen den untersuchten LLMs. Die Modelle weichen nicht nur deutlich von den Befunden der Humanexperimente ab; sie erzeugen auch untereinander stark divergierende und teils gegensätzliche Resultate, die je nach Modell unterschiedliche Narrative über Fairnesswahrnehmungen stützen. Während DeepSeek substanzielle menschliche Beteiligung schätzt, implizieren die Befunde von Gemini, dass die Unterscheidung zwischen menschlichen und maschinellen Entscheidern weitgehend irrelevant ist; GPT-4o reproduziert das Muster der Humanexperimente lediglich abgeschwächt.

Das zentrale Problem liegt dabei weniger in der Abweichung von den Urteilen menschlicher Probanden, die über gängige Rekrutierungsplattformen wie Prolific gewonnen wurden. Deren Stichproben sind nicht notwendig repräsentativ; entsprechend kann auch das Humanexperiment nicht ohne Weiteres als ground truth gelten.

Das eigentliche methodische Kernproblem beim Einsatz von LLMs

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

als Ergänzung oder Substitut zu Humanexperimenten besteht vielmehr in der mangelnden Modellstabilität. Wenn identische Stimuli zu qualitativ divergierenden Ergebnissen führen, ist zweifelhaft, inwiefern LLM-Experimente derzeit eine belastbare Grundlage für empirische oder normative Schlussfolgerungen bieten. Die Resultate hängen in hohem Maße vom gewählten Modell und – häufig implizit – von der Aggregationsmethode ab. Da sowohl die Modellauswahl als auch die Aggregationsstrategie erhebliche Freiheitsgrade bieten, lassen sich mit LLMs nahezu beliebige Ergebnisse generieren.

Auch wenn LLMs ein faszinierendes und hilfreiches Instrumentarium bieten, werfen unsere Befunde zumindest Zweifel daran auf, dass LLMs sich bereits heute als kostengünstige Alternative zu klassischen Humanexperimenten nutzen lassen. Human-in-the-Loop bleibt daher in doppelter Hinsicht zentral: zur Sicherung wahrgenommener Verfahrensfairness und zur Gewinnung belastbarer empirischer Erkenntnisse über die Wirkungen des Rechts. Auch wenn eine Konvergenz menschlicher und maschineller Bewertungen denkbar ist, lässt sich mit Blick auf die Generierung von Hypothesen und von Verhaltensbefunden folgern: Human-Experiment und LLM-Experimente sind keine Substitute, sondern Komplemente. Für die experimentelle und empirische Rechtsforschung ist das möglicherweise ein Vorteil. Das Arsenal von Methoden im Antlitz fundamentaler Unsicherheit wird lediglich reichhaltiger.

Literatur

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PLoS ONE, 11(10), e0164856.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning. MIT Press.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. Cognition, 181, 21–34.

Binns, R. (2022). Human judgment in algorithmic loops: Individual justice and automated decision-making. AI & Society, 37(3), 1169–1179.

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making, 33(2), 220–239.

Chen, B., Hermstrüwer, Y., Langenbach, P., Stremitzer, A., & Tobia, K. (2025). Mitigating the Judicial Human-AI Fairness Gap. Center for Law & Economics Working Paper Series, Nr. 7.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Chen, B., Stremitzer, A., & Tobia, K. (2022). Having Your Day in Robot Court. Harvard Journal of Law & Technology, 36(1), 127–169.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2. Aufl.). Lawrence Erlbaum Associates.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (S. 797–806). ACM.

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. Philosophy & Technology, 29(3), 245–268.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580.

Engel, C. (2001). Offene Gemeinwohldefinitionen. Rechtstheorie, 32(1), 23–52.

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. Report for the Administrative Conference of the United States.

Henning, M., & Langenbach, P. (2024). Procedural fairness in algorithmic decision-making: The role of explanations and human oversight. Journal of Behavioral Public Administration, 7(1), 1–28.

Hermstrüwer, Y. (2020). Fairnessprinzipien der algorithmischen Verwaltung: Diskriminierungsprävention beim staatlichen Einsatz von Machine Learning. Archiv des öffentlichen Rechts, 145(4), 479–521.

Hermstrüwer, Y., & Langenbach, P. (2023). Human-AI cooperation in legal decision-making: Evidence from algorithmic assistance in the courtroom. Law and Human Behavior, 47(3), 234–267.

Imai, K., & Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. Political Analysis, 21(2), 141–171.

Jung, J., Concannon, C., Shroff, R., Goel, S., & Goldstein, D. G. (2020). Simple rules to guide expert classifications. Journal of the Royal Statistical Society Series A: Statistics in Society, 183(3), 771–800.

Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE, 12(4), e0174698.

Kehl, D., Guo, P., & Kessler, S. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. The Quarterly Journal of Economics, 133(1), 237–293.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Lehr, D., & Ohm, P. (2017). Playing with the data: What legal scholars should learn about machine learning. University of California Davis Law Review, 51(2), 653–717.

Liu, Y., & Li, Z. (2025). AI-assisted judicial reasoning: Opportunities and challenges for legal decision-making. Journal of Legal Technology, 8(1), 45–62.

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Telematics and Informatics, 67, 101751.

Mayson, S. G. (2018). Bias in, bias out. Yale Law Journal, 128(8), 2218–2300.

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (S. 411–420). Morgan Kaufmann.

Political Asylum/Immigration Representation Project (2016). Pro Bono Asylum Manual. Boston College Law School.

Re, R. M., & Solow-Niederman, A. (2019). Developing artificially intelligent justice. Stanford Technology Law Review, 22(2), 242–289.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics, 15(6), 657–680.

Selbst, A. D. (2021). An institutional view of algorithmic impact assessments. Harvard Journal of Law & Technology, 35(1), 117–186.

Stevenson, M. T. (2018). Distortion of justice: How the inability to pay bail affects case outcomes. The Journal of Law, Economics, and Organization, 34(4), 511–542.

Stevenson, M. T., & Doleac, J. L. (2019). Algorithmic Risk Assessment in the Hands of Humans. NBER Working Paper Nr. 26448.

Tyler, T. R. (1997). The psychology of legitimacy: A relational perspective on voluntary deference to authorities. Personality and Social Psychology Review, 1(4), 323–344.

Tyler, T. R., & Huo, Y. J. (2002). Trust in the Law: Encouraging Public Cooperation with the Police and Courts. Russell Sage Foundation.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97–112.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76–99.

Wischmeyer, T. (2020). Artificial intelligence and transparency: Opening the black box. In T. Wischmeyer & T. Rademacher (Hrsg.), Regulating Artificial Intelligence (S. 75–101). Springer.

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. ERA Forum, 20(4), 567–583.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

VI.
Theory

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The Coase Theorem and Divorce

Eric Helland

In honor of Christoph Engels's career long habit of revisiting, reevaluating, and often extending classic papers in law and economics and empirical legal studies, this paper revisits Mnookin and Kornhauser's 1978 paper's insight that divorce bargaining takes place "in the shadow of the law" and evaluates it through the lens of the recent advances in our understanding of the Coase theorem and bargaining in the shadow of the law more generally. The Coasian prediction that private bargaining yields efficient and durable outcomes is at odds with the observed fragility of divorce settlements, where custody, support, and visitation orders are frequently reopened for modification. Using evidence from the Los Angeles Superior Court and extending the divergent expectations model, I examine how the limits of the Coase theorem map onto family law. Bargaining theory limits, interdependent utilities (altruism and spite), transaction costs, commitment problems, legal limits rooted in third-party rights, and agency problems from lawyer incentives all narrow the settlement range and, in some cases, make durable agreements less likely. These extensions suggest why divorce settlements often unravel and why family courts devote substantial resources to post-judgment modification.

A. Introduction*

It has been almost fifty years since the publication of Mnookin and Kornhauser's seminal article, *Bargaining in the Shadow of the Law* (Mnookin and Kornhauser, 1978). Their essay was the first to apply what has come to be known as the divergent expectations model to divorce and family law. Building on the high rate of stipulated judgments (settlements) and the divergent expectations model of pretrial bargaining, the paper's central

* The author wishes to thank Professor Christoph Engel, distinguished colleague, outstanding scholar, wonderful coauthor, German food connoisseur, and friend, for his insightful comments over the years, and looks forward to many more.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

insight was that divorce law does not simply impose order from above. Instead, legal rules provide the framework within which divorcing couples bargain to determine post-dissolution rights and responsibilities. This stood in contrast to the prevailing view of the time that the family court's primary function was to protect the interests of third parties, particularly children, or to protect weaker spouses. Mnookin and Kornhauser wrote at the dawn of the no-fault revolution in divorce.[1] Their article argued a very Coasian point: if transaction costs are low and parties are reasonably well informed, bargaining should achieve efficient outcomes in divorce. The law would matter primarily in determining the distribution (who has bargaining power), not efficiency or, by extension, durability.[2]

In some ways, Mnookin and Kornhauser's article is different and more forward-thinking than the conventional formulation of the Coase theorem prevalent when the article was written. The conventional version

1 As I discuss further below, much of Professor Engel's research bears directly on understanding the strengths and limitations of Mnookin and Kornhauser's approach. Perhaps unsurprisingly, Christoph has also contributed directly to our understanding of how legal change unfolds. Writing with Keren Weinshall, he shows that innovations in law typically diffuse along an S-shaped path: a long period of stability, a phase of rapid adoption once reform takes off, and finally a plateau as the innovation becomes institutionalized (Engel and Weinshall, 2022). This pattern closely matches the trajectory of divorce law: decades of doctrinal stability, a wave of rapid change in the 1970s, as many jurisdictions around the world adopted no-fault divorce, and the eventual emergence of a new equilibrium of private ordering modeled by Mnookin and Kornhauser (1978).

2 An instructive example of what Medema (2020) describes as a "Coasian critique" is Donohue III (1990). Posner and Shavell had argued that shifting from the American rule (each party bears its own fees) to the British rule (loser pays) would affect settlement rates, since the parties' incentives to settle differ under the two regimes. Donohue responds by invoking the Coase theorem in its conventional "invariance" form. If transaction costs are low, litigants should be able to contract around the default fee-shifting rule, so efficiency and settlement rates should not depend on whether the law adopts the American or British rule. Donohue emphasizes that Posner and Shavell's prediction arises from overlooking this Coasean benchmark. Yet he also notes the puzzle: in practice, parties rarely contract around fee-shifting defaults, which suggests that either transaction costs are higher than theorists assume, or litigants fail to perceive and exploit the possibility of contractual adjustment. This highlights the broader function of a Coasian critique: to challenge theories that attribute substantive effects to legal rules by asking why private bargaining does not undo those effects. In the same spirit, my argument is that the divergent expectations model of divorce bargaining takes for granted that parties negotiate "in the shadow of the law," but fails to explain why divorcing spouses so often cannot or do not bargain their way around default rules to reach efficient and durable settlements.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

was a proposition concerning externalities, which argued that if transaction costs are zero and property rights are well defined, bargaining between affected parties will lead to a Pareto-efficient allocation of resources regardless of the initial assignment of rights. Two claims follow: an efficiency proposition (bargaining leads to Pareto efficiency) and an invariance proposition (the efficient outcome does not depend on the initial legal entitlement).[3]

Almost 50 years after the publication of Mnookin and Kornhauser, it's worth asking their central question again. Why do the outcomes in divorce cases diverge from the irrelevance proposition of the Coase theorem? Consider the evidence discussed below: as was true 50 years ago, a large share of divorcing parties arrive in court with some matters already settled, and stipulated judgments remain the most common outcome in most courts. Table 1 shows that for divorce (dissolution) cases filed in Los Angeles Superior Court (LASC) between 2017 and 2024, the trial rate is only about 15 %, 40 % are stipulated judgments (settlements), while in 38 % one party does not participate (defaults).[4] Mnookin and Kornhauser were interested in how legal rules and procedures influence this process of private ordering: to what extent do courts authorize and enforce privately negotiated commitments at divorce, and to what extent do legal or institutional limits constrain such commitments? Already in 1979, they emphasized that some areas of family law remained outside Coasian bargaining, such as community property rules, the "best interests of the child" standard, and statutory child support formulas. Since then, the trend has been toward greater acceptance of stipulated judgments with minimal judicial review, consistent with the limited investigatory capacity of courts.

3 In fact, in his original essay, Coase emphasized the opposite: in the real world, transaction costs are significant, and therefore legal rules matter for efficiency. As Medema and others point out, the "Coase theorem" as a formal proposition was really Stigler's creation, distilled from Coase's examples. Coase's original aim was to highlight the centrality of transaction costs rather than to assert a clean theorem. Medema describes the Coase theorem as a "blackboard result," a baseline for thinking about real-world institutions. The neat predictions of efficiency and invariance under zero transaction costs serve as a counterfactual against which we can evaluate the impact of legal rules, bargaining frictions, and institutional design.

4 Because California requires separate filing fees for a divorce petition and a response, some "defaults" may reflect settlements in which only one party pays the filing fee. In such cases, sometimes called a "default with agreement," the respondent risks losing procedural protections, since failing to respond means the petitioner may still seek a different division of assets or custody than the parties agreed to informally.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

For example, judges in the Los Angeles Superior Court routinely accept stipulated arrangements unless there is clear evidence that a child's safety is at risk. In practice, the court's function has become to process disputes and enforce settlements rather than to engage in independent fact-finding. So far, so good for the Mnookin and Kornhauser model.

However, the durability of these settlements is far less clear. As shown in Table 1 and Figure 2, evidence from the LASC shows that agreements often break down, particularly with respect to custody, visitation, and support. Unlike torts or contracts, family cases do not truly end at judgment. Even after settlement, either party may petition for modification on the grounds that circumstances have changed or that the best interests of the child require it. Mediation is common and often produces stipulated judgments, but these agreements appear to be fragile as modification requests consume a large share of judicial resources. This pattern is puzzling from a Coasian perspective. If transaction costs are low and parties could bargain efficiently at divorce, as would appear to be the case, they should be able to draft contingent contracts covering foreseeable contingencies, and there should be little need for repeated modification requests. The persistence of post-divorce requests for modification of previous judgments, even stipulated judgments (i.e., settlements), suggests that the Coase theorem's predictions do not map cleanly onto family law.

One way to motivate the puzzle is by comparison to civil litigation at LASC. In civil cases, fewer than 10 percent go to trial, and only about 2 percent are appealed. By contrast, although only 15 percent go to trial in divorce, roughly 8 percent of cases return to court for a request for modification within five years. Measured against the civil appeal rate, this persistence is striking: divorce judgements are four times more likely to be revisited than civil judgments, and unlike appeals, modification proceedings are not limited to error correction but reflect fresh bargaining failures. If we applied the Coase theorem in its most stringent form, we would expect post-divorce bargaining to reduce the need for judicial intervention to levels closer to the civil benchmark, not this elevated rate.

In principle, divorcing spouses could bargain over custody, visitation, alimony, or asset division once a marriage dissolves. The Coase theorem suggests that if transaction costs are low, the initial legal rule, such as the presumptive custody arrangement or the default property division, should not affect efficiency or durability. Parents could, in theory, trade or adjust rights to reach efficient arrangements, which theoretically should be durable since they were privately arrived at. The law would matter pri-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

marily for distribution (i.e., which spouse has bargaining power), but not for efficiency.

Yet if divorce bargaining regularly defies these predictions, then we should consider the features of family disputes that make the Coase theorem's invariance proposition seem questionable. In one sense, all of this fits in the original spirit of Coase's article, if not the invariance proposition Stigler attributed to Coase. Transaction costs are high: lawyers' fees and asymmetric information about assets or parenting ability make bargaining costly. Commitment problems are endemic, particularly in custody disputes where cooperation must persist over years. Asymmetric information is pervasive: spouses may hide income, exaggerate needs, or misrepresent parenting quality. Children introduce third-party externalities, since their welfare is central, but they are not parties to the bargain. Finally, legal constraints limit bargaining freedom: custody must conform to the "best interests of the child," and child support is non-waivable by statute in most jurisdictions. However, some critiques go beyond Coase's original "transaction costs matter" limitation on the invariance theorem he may never have actually articulated. Mnookin and Kornhauser applied bargaining theory to divorce, arguing that default legal rules shape the bargaining set even when cases settle. But consider Farrell (1987)'s early critique of Coase, which contends that efficiency does not automatically follow from private negotiation, even absent transaction costs. This paper considers several possible problems with the strong Coasian invariance proposition as it applies to divorce and suggests several modifications to the standard divergent expectations model that might capture these limitations in ways that suggest possible avenues for future research.

This puzzle resonates with the research program of Christoph Engel, much of whose scholarship has been devoted to uncovering the limits of Coasian optimism and rational choice models, and to demonstrating the importance of institutions in remedying these limits. His contributions on bounded rationality and heuristics, the fragility of cooperation in social dilemmas, the role of institutional design in shaping behavior, and the externalities of collective goods all highlight mechanisms that derail efficient private ordering. In this spirit, the present essay revisits *Bargaining in the Shadow of the Law* and asks: why might divorce bargaining fail? Some potential answers are, of course, found in Engel's experimental work on the precariousness of cooperation (Engel and Rockenbach, 2024) and bounded cognition in property settings (Engel and Bar- Gill, 2024), in which he shows how limited rationality and behavioral biases systemat-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ically deter efficient trades once bargaining shifts from competitive markets into bilateral monopolies. For example, Mnookin and Kornhauser's original model relied on a market analogy, but several of Engel's experiments are suggestive that divorce bargaining may live in what Engel and Helland call the "fundamental transformation" zone, where Coasian optimism breaks down due to bilateral monopolies and information problems (Engel and Helland, 2023).[5]

B. Evidence from the Los Angeles Superior Court

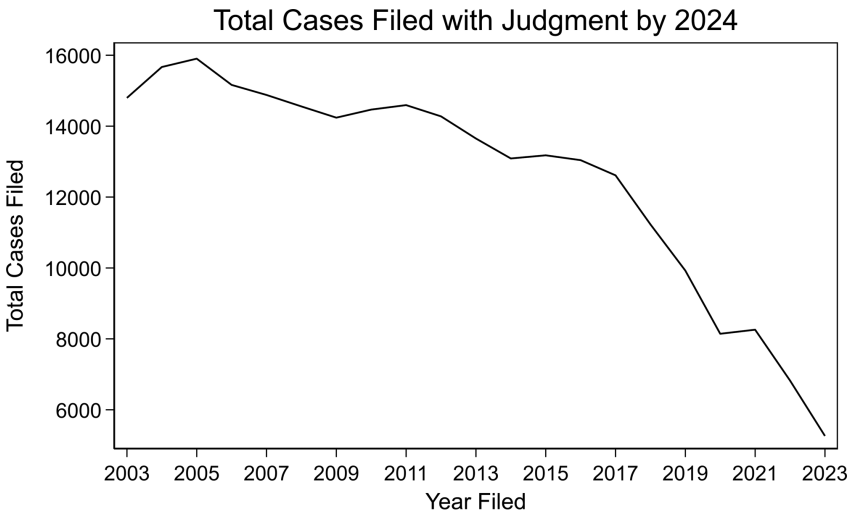
To motivate the puzzle at the heart of this paper, I begin with some statistics from the Los Angeles Superior Court (LASC). LASC is the largest unified trial court in the United States, handling more cases than any other single jurisdiction. Given the size of Los Angeles County, it is unsurprising that the court processes tens of thousands of divorce petitions each year. As shown in Figure 1, the number of divorces involving minor children has declined from a peak in 2003. In recent years, part of the apparent decline may reflect delays in case resolution, since my sample includes only completed cases. Nevertheless, the broader pattern is consistent with state and national trends: both in Los Angeles and California as a whole, the divorce rate has fallen since 2000, with roughly 8.9 % of the adult population reporting that they are divorced.[6]

More interesting for this paper is the outcome and durability of divorce cases. As noted above, there is no apparent reason, in a pure Coasian invariance sense, that the outcomes in family court should be less likely to settle or that the resolutions that occur should be less durable than other civil cases. Nothing in the legal rules should prevent parties from, on average, forecasting future conditions and writing a full contingent,

5 Over Engel's career, he has paid particular attention to the role of judges, whether in defining collective goods, satisficing under cognitive limits (Engel and Güth, 2018), or balancing normative concerns (Engel and Rahal, 2019). Although not addressed in this essay, his work reminds us that the judge is not merely a cipher but an active institutional actor shaping bargaining outcomes. Although not directly addressed in the divergent expectations model, the role of judges in divorce cases is certainly important to understanding parties' behavior.

6 [6]See U.S. Census Bureau, American Community Survey, 2018–2022 5-Year Estimates. The ACS measures the marital status of the population age 15 and older; the proportion currently divorced in California is approximately 8.9 %.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -



Note: This graph shows data from LASC on dissolution petitions involving minor children.

Figure 1: Number of Divorce Petitions LASC

and hence durable, contract regardless of legal rules. As noted in the introduction, civil litigation at LASC is by some measures less contentious: fewer than 10 percent of cases go to trial, and only about 2 percent are appealed–typically larger high-stakes cases in the complex courts.

Outcomes in divorce cases in LASC fall into three broad categories: contested trials, stipulated judgments (settlements), and defaults (where one party does not appear). Not all litigants are represented by counsel, and judgments are frequently modified even when settlements are reached as custody, visitation, child support, and spousal support are repeatedly relitigated. Mediation is common and often produces stipulated judgments, but these agreements are not necessarily durable.

For context, consider the breakdown of outcomes for petitions filed between 2017 and 2024 presented in Table 1. In divorce cases with minor children, which judges suggest are the most contentious, 15 percent go to trial (a contested judgment), and roughly 8 percent of cases return to court for a request for modification within five years. Measured against the civil appeal rate, this lack of durability is interesting: divorce judgments are four times more likely to be revisited than civil judgments. Unlike appeals, modification proceedings are not limited to error correction but often reflect fresh bargaining failures.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 1: Requests for Modification by Judgment Type Dissolution Cases Involving Minor Children (2017–2024)

Outcome	Request for Modification in 2 years	Request for Modification in 5 years	Total Cases
Default Judgment	1,216	1,713	24,674
Stipulated Judgment	1,614	2,343	26,481
Contested Judgment	700	897	4,213
Other	164	222	9,545
Total	**3,694**	**5,175**	**64,913**

Unlike most civil cases, family cases rarely end with finality: appeals are infrequent, but either party can return to court to argue that circumstances have changed and seek modification of orders. Attorney involvement often generates substantial fees, further raising transaction costs. Consider Figure 2. The top panel shows the fraction of judgments requesting a modification within the next two years, while the bottom panel shows the fraction of judgments with modification requests within five years. Interestingly, while the proportion of those judgments seeking a modification grows, the increase is not dramatic, suggesting that modifications may not be driven entirely by changing circumstances.

From a Coasian invariance perspective, this pattern is puzzling. If transaction costs were low, parties should bargain efficiently over custody, support, and property division. Once an agreement is reached, especially in a stipulated judgment, parties should have internalized foreseeable contingencies, reducing the need for court intervention. Coasean invariance suggests that once a bargain is struck, the efficient frontier has been reached, and further modification petitions should be unnecessary.

What we observe in LASC seems at odds with the strongest predictions of a Coasian invariance proposition. A high proportion of custody and support orders are later modified, and often relatively quickly, whether initially stipulated by the parties or imposed by a judge. This raises a series of questions: Why do divorcing spouses not write contingent contracts anticipating future changes? Why do they return to court rather than adjust privately? Why are settlements not substantially more durable than judgments?

In the next section, we explore several explanations, including the high transaction costs of renegotiation, the difficulty of enforcing long-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

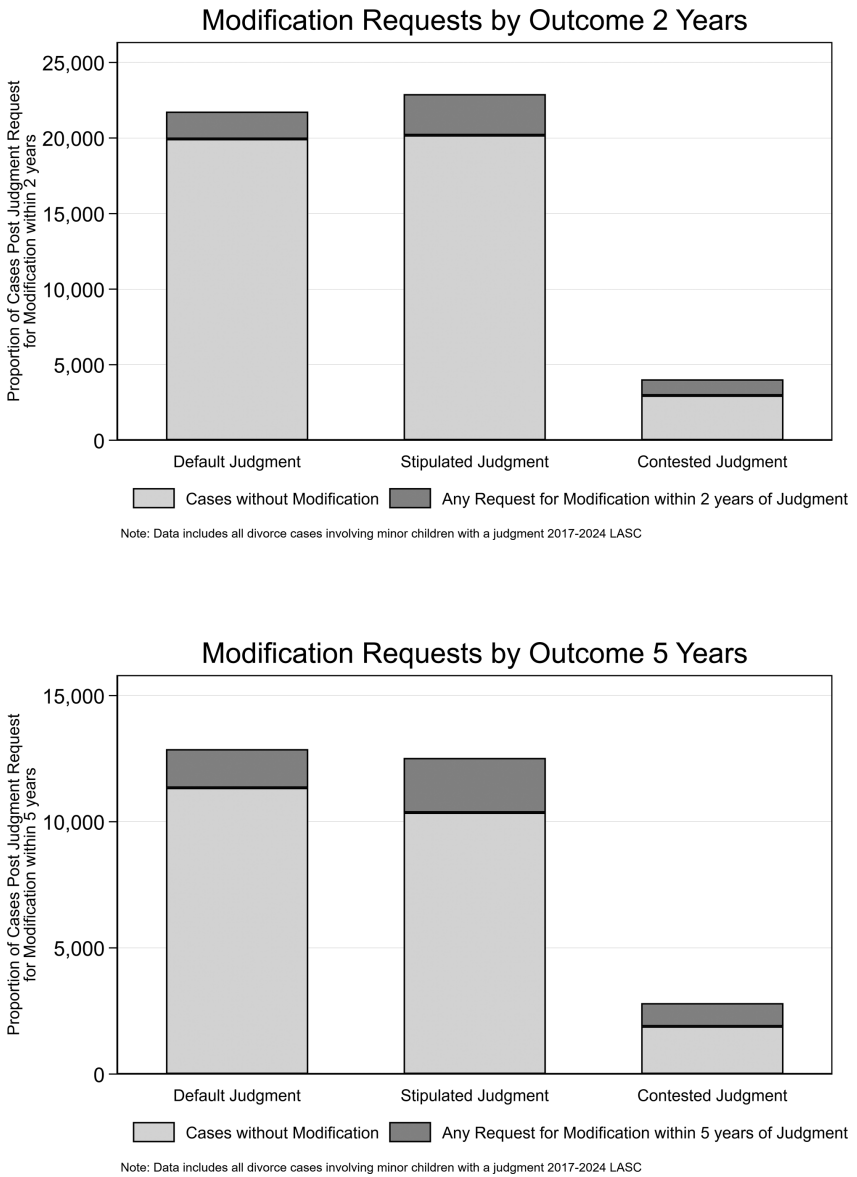


Figure 2: Requests for Modification

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

term commitments in parenting arrangements, pervasive private information, externalities to children, and legal limits on private ordering. Altruism and its strategic variants may also distort bargaining, as parents seek to present themselves as cooperative or caring even at efficiency's expense. Family law's statutory limits, especially the mandatory modifiability of child support, further guarantee that settlement is never truly final.

C. The Divergent Expectations Model in Divorce

I. Standard Model

A central contribution of Mnookin and Kornhauser (1978) was to show how divorce negotiations take place not in a legal vacuum but in the shadow of default rules and judicial procedures. Even when cases settle, the law shapes bargaining outcomes by providing threat points, structuring private commitments, and constraining opportunism. Subsequent work by Cooter, Marks, and Mnookin extended this insight into a broader bargaining theory of contracts, highlighting how legal rules influence not only efficiency but also the durability of negotiated outcomes.[7]

As typically presented, the model does not entirely capture why divorce cases are often among the most contentious cases the courts face. All divorces require resolution of two to four distributional issues: division of property and debts, possible claims on future earnings (alimony), child support, and custody/visitation arrangements. While property division and alimony can often be modeled as transfers, custody and visitation are indivisible rights at the core of family bargaining. In California, child support is determined by a statutory formula based on income and custody

7 There is a related literature on the economics of the family that reinforces these themes. Household bargaining models of the 1980s and 1990s (e.g. Manser and Brown, McElroy and Horney, and Lundberg and Pollak) treated divorce as a credible threat that shaped intra-marital allocations. Spouses' fallback positions, determined by what they could expect in divorce, influenced resource division within marriage and at separation. Later "divorce-threat" and lifecycle models explicitly incorporated limited commitment and dynamic interaction, showing that divorce laws, custody presumptions, and support rules shift bargaining power and can destabilize agreements. Empirical work confirms that legal reforms altering threat points, such as changes in child benefit allocation or custody presumptions, have measurable effects on intra-household outcomes and settlement behavior.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

shares, but the remaining issues are subject to negotiation and judicial discretion. Litigants routinely report that outcomes are uncertain and highly fact-dependent, especially for custody and visitation. This background of uncertainty, combined with commitment problems and asymmetries of information, creates room for strategic bargaining and repeated requests for modification. To capture this dynamic, I adapt Cooter et al. (1982)'s "differing perceptions" model of litigation to the custody context. Instead of binary trial outcomes (plaintiff wins/loses), custody litigation may result in sole custody for one parent or joint custody, with associated asset allocations. Settlement occurs if there exists a transfer T that makes both parties weakly better off than their subjective trial values.

Let the total divisible assets be J. If custody outcome $c \in \{W, H, J\}$ occurs (wife custody, husband custody, joint custody), the husband receives $\alpha_c J$ and the wife $(1 - \alpha_c)J$. In addition, each parent values custody directly. Let

$$\begin{aligned} u_w(W) &= V + (1 - \alpha_W)J, \quad & u_h(W) &= \alpha_W J, \\ u_w(H) &= (1 - \alpha_H)J, & u_h(H) &= V + \alpha_H J, \\ u_w(J) &= v + (1 - \alpha_J)J, & u_h(J) &= v + \alpha_J J, \end{aligned}$$

where $V > v \geq 0$ measures the utility of sole versus joint custody.[8]

Each party $i \in \{w, h\}$ incurs litigation cost C_i. Each has subjective beliefs about trial outcome probabilities:

$$\hat{p}^i_W,\ \hat{p}^i_H,\ \hat{p}^i_J, \qquad \hat{p}^i_W + \hat{p}^i_H + \hat{p}^i_J = 1.$$

The wife's and husband's expected utilities from a trial are:

$$\begin{aligned} E^T_w &= \hat{p}^w_W\, u_w(W) + \hat{p}^w_H\, u_w(H) + \hat{p}^w_J\, u_w(J) - C_w, \\ E^T_h &= \hat{p}^h_W\, u_h(W) + \hat{p}^h_H\, u_h(H) + \hat{p}^h_J\, u_h(J) - C_h. \end{aligned}$$

Suppose the parties settle on a custody arrangement c and a net transfer T (from husband to wife). Then

$$U^S_w(c, T) = u_w(c) + T, \qquad U^S_h(c, T) = u_h(c) - T.$$

8 Here, custody is modeled not as part of the divisible asset pool J, but as an indivisible entitlement that directly contributes to each parent's utility. The parameters V and v capture the nonpecuniary value of sole versus joint custody, reflecting that parenting rights cannot be reduced to purely monetary transfers.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The wife accepts (c, T) iff $T \geq T^W_{\min}(c) \equiv E^T_w - u_w(c)$; the husband accepts iff $T \leq T^H_{\max}(c) \equiv u_h(c) - E^T_h$.

Hence, settlement under custody outcome c is feasible iff

$$T^W_{\min}(c) \leq T \leq T^H_{\max}(c),$$

or equivalently,

$$E^T_w + E^T_h \leq u_w(c) + u_h(c). \tag{8}$$

If joint custody is excluded, only $c \in \{W, H\}$ remain. Assume symmetric valuations and asset splits so that the only stake is the custody value V. Let $\pi \equiv \hat{p}^w_W$ (be the wife's belief she will win custody) and $q \equiv \hat{p}^h_W$ (the husband's belief that his ex-wife wins). Then (8) reduces to

$$(\pi - q)V \leq C_w + C_h. \tag{9}$$

Settlement exists when (9) holds; trial occurs when it is violated.[9]

With $c \in \{W, H, J\}$, the condition (8) applies to each custody type. Define relative "stake differences" for custody k relative to c:

$$\Delta_w(k|c) = u_w(k) - u_w(c), \qquad \Delta_h(k|c) = u_h(k) - u_h(c).$$

Then the feasibility under custody c is equivalent to

$$\sum_{k \in \{W,H,J\}} \left(\hat{p}^w_k \Delta_w(k|c) + \hat{p}^h_k \Delta_h(k|c)\right) \leq C_w + C_h.$$

A settlement exists if at least one custody choice c satisfies this inequality. The trial condition is the negation:

$$E^T_w + E^T_h > \max_{c \in \{W,H,J\}} \{u_w(c) + u_h(c)\}.$$

9 This is the custody analogue of the standard model's $(P_p - P_d)J \leq C_p + C_d$.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 2: Symbols and Definitions for the Custody Adaptation of the Differing Perceptions Model

Symbol	Definition
J	Total divisible assets
α_c	Husband's share under custody outcome c
$u_i(c)$	Utility of party i under custody c (assets + custody value)
V, v	Nonpecuniary values of sole vs. joint custody
C_i	Litigation cost for party i
$\hat{p}_c^{\,i}$	Party i's subjective belief about outcome c
E_i^T	Party i's subjective expected trial utility
T	Transfer from husband to wife
$T^W_{\min}(c)$	Minimum transfer for wife to accept custody c
$T^H_{\max}(c)$	Maximum transfer for husband to accept custody c
Feasibility	$E_w^T + E_h^T \leq u_w(c) + u_h(c)$
Trial condition	$E_w^T + E_h^T > \max_c\{u_w(c) + u_h(c)\}$

Intuitively, settlement is possible whenever the parties' subjective trial values are not too far apart. Each side compares what it expects to receive at trial (net of costs) with what it could obtain through a negotiated custody arrangement plus a transfer. If the sum of their subjective trial values does not exceed the total surplus available under some custody outcome, then there exists an overlap, an interval of transfers $[T^W_{\min}(c),\ T^H_{\max}(c)]$, that both parties prefer to trial. This overlap is the settlement range. Conversely, if subjective trial values are so optimistic that they exceed any custody surplus together, no such overlap exists, and the case proceeds to trial. So far, the model is broadly consistent with Mnookin and Kornhauser (1978) and Cooter et al. (1982)'s prediction that the settlement rate should be relatively high, which we find in the LASC data. At the same time, custody differs from purely monetary disputes in one crucial respect: the subjective value of custody outcomes evolves as children age and circumstances change. This makes initial settlements more vulnerable to later contestation, helping to explain why requests for modification persist long after judgment in family law, in contrast to most other civil cases. In the following subsection, we modify the model to examine requests for modification after judgment.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In Table 2 we provide the definitions for each of the components of the modified model.

II. Requests for Modification

One issue not covered in the standard divergent expectations model is the possibility of a "request for modification," in which one or both of the parties ask the court to modify the judgment after it has been agreed to in a settlement (stipulated judgment) or trial (contested judgment). To extend the model, I divide it into two stages.

Stage 1: pre-filing bargaining over custody $c \{W, H, J\}$ and transfer T ; if no agreement, trial 1 occurs.

Stage 2: either party may petition the court for reconsideration (a second adjudication). Before trial 2 (technically, a hearing on the request for modification), the parties may bargain again.

Let $u_i(c)$, beliefs $\hat{p}_k^i$, and costs C_i be as defined above. Stage-1 settlement under custody c is feasible iff

$$E_w^{T,1} + E_h^{T,1} \;\le\; u_w(c) + u_h(c), \tag{10}$$

where

$$E_w^{T,1} = \sum_{k\in\{W,H,J\}} \hat{p}_k^{\,w}\, u_w(k) - C_w, \qquad E_h^{T,1} = \sum_k \hat{p}_k^{\,h}\, u_h(k) - C_h.$$

After trial 1, suppose parties receive a signal s (e.g., the observed decision, judge's reasoning, new evidence) that updates beliefs to $\hat{p}_k^{\,i,2} = \Pr(k \mid s)$ and induces incremental litigation costs $K_i \ge 0$ for a second trial.[10] The Stage-2 expected trial utilities are

$$E_w^{T,2} = \sum_k \hat{p}_k^{\,w,2}\, u_w(k) - (C_w + K_w), \qquad E_h^{T,2} = \sum_k \hat{p}_k^{\,h,2}\, u_h(k) - (C_h + K_h).$$

A Stage-2 settlement under custody c is feasible iff

$$E_w^{T,2} + E_h^{T,2} \;\le\; u_w(c) + u_h(c). \tag{11}$$

Proposition 1 (Replication without changes). If $\hat{p}_k^{\,i,2} = \hat{p}_k^{\,i}$ for all i, k and $K_i = 0$, then the Stage-2 feasibility set (11) coincides with Stage 1 (10).

10 If no new information or cost change occurs, take $\hat{p}_k^{\,i,2} = \hat{p}_k^{\,i}$ and $K_i = 0$.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Hence, absent new information or cost changes, the second-stage bargaining problem is unchanged, and no new request for modification should arise. Broadly, Proposition 1 implies that when legal rules are transparent and circumstances remain stable, the Stage 2 problem simply reproduces Stage 1. In such environments, requests for modification should be rare, since the equilibrium is already renegotiation-proof.

Proposition 2 (Why post-trial settlement can emerge). Suppose at least one of the following holds: (i) belief dispersion shrinks after s (e.g., $\hat{p}^{\,w,2}$ and $\hat{p}^{\,h,2}$ move closer); (ii) incremental costs $K_w + K_h > 0$; (iii) the court's first decision $y \in \{W, H, J\}$ changes the custody/asset baseline relevant to $u_i(c)$ (e.g., status quo bias or enforcement frictions). If the first judgment y shifts the effective baseline, then the surplus term $u_w(c) + u_h(c)$ is evaluated relative to y, which can raise the RHS of the Δ-form inequality and create overlap at Stage 2.[11] Then (11) can hold even when (10) failed, creating a Stage-2 settlement interval

$$[T^{W,2}_{\min}(c),\ T^{H,2}_{\max}(c)] = [E^{T,2}_w - u_w(c),\ u_h(c) - E^{T,2}_h].$$

Equivalently, in the $\{W, H, J\}$ model, any combination of (i) smaller $\sum_k |\hat{p}^{\,w,2}_k - \hat{p}^{\,h,2}_k|$, (ii) larger $K_w + K_h$, or (iii) a baseline shift from y that increases $\max_c\{u_w(c) + u_h(c)\}$ can turn infeasibility at Stage 1 into feasibility at Stage 2.

Lemma (Stage-2 feasibility without Stage-1 feasibility). If $(\pi^{(1)} - q^{(1)})V > C_w + C_h$ but $(\pi^{(2)} - q^{(2)})V \le C_w + C_h + K_w + K_h$, then no settlement overlap exists at Stage 1 while a (nonempty) settlement interval exists at Stage 2.[12]

Intuitively, updated beliefs, additional expected costs, or a changed outside option expand the overlap of reservation values and make settle-

11 Put more intuitively: once the court enters a custody order as part of the judgment, it essentially becomes the default arrangement—hence the terminology "request for modification." Parents now face legal requirements that make the existing order the starting point for future bargaining. A first judgment, therefore, shifts the baseline so that even if the original settlement range was empty, a new overlap of interests may appear when the order itself anchors the parties' expectations.

12 Formally, if only $\{W, H\}$ are available and the stake is V, the Stage-2 analogue is
$$(\pi^{(2)} - q^{(2)})V \ \le\ C_w + C_h + K_w + K_h,$$
where $\pi^{(2)} = \hat{p}^{\,w,2}_W$ and $q^{(2)} = \hat{p}^{\,h,2}_W$. Settlement post-trial becomes strictly more likely when $K_w + K_h > 0$ or when $|\pi^{(2)} - q^{(2)}|$ shrinks.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ment more likely. Put differently, if the judge does something the parties did not expect, the request for modification can be a way to get a judgment closer to the parties' expectations.

In custody bargaining, the judge's prospective assignment (and its uncertainty) forms the outside option that shapes negotiated outcomes; adding a reconsideration stage simply re-runs the same overlap test with updated beliefs and costs. Without such updates, there is no reason to expect a different negotiated arrangement before a second trial. With updates (new information from the first proceeding, altered risk, or added costs), a second-stage settlement can arise even when the first stage failed.[13]

What role does anticipation play, and when is a judgment in this model renegotiation-proof? Suppose parties foresee a reconsideration stage (Stage 2) after an initial adjudication. If there is no new information and no cost change (beliefs $\hat{p}_k^{i,2} = \hat{p}_k^i$ and incremental costs $K_i = 0$), then the Stage 2 overlap test

$$E_w^{T,2} + E_h^{T,2} \;\leq\; u_w(c) + u_h(c)$$

coincides with the Stage 1 test. Hence, the Stage 1 bargaining problem already internalizes the possibility of reconsideration, and any Stage 1 equilibrium that is subgame-perfect is renegotiation-proof: off-path petitions for reconsideration should not be observed. By contrast, if Stage 2 brings either (i) belief convergence (updated $\hat{p}_k^{i,2}$ after observing the first proceeding) or (ii) added second-trial costs $K_w + K_h > 0$, then

$$E_w^{T,2} + E_h^{T,2} \;=\; \sum_k \hat{p}_k^{w,2} u_w(k) + \sum_k \hat{p}_k^{h,2} u_h(k) \;-\; (C_w + C_h + K_w + K_h)$$

can fall relative to Stage 1, so the overlap condition may newly hold even if it previously failed. In short, absent updates, the reconsideration game replicates Stage 1; with updates, a post-trial settlement interval $[T_{\min}^{W,2}(c),\, T_{\max}^{H,2}(c)] = [E_w^{T,2} - u_w(c),\; u_h(c) - E_h^{T,2}]$ can open where none existed ex ante.

Broadly, the lesson of extending the divergent expectations model is that requests for modification arise only when the underlying belief

13 For example, the high turnover in family courts, where judges typically last only a few years before being transferred to areas such as criminal law or general civil, inadvertently leads to more requests for modification because it results in new information to the parties.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

structure or the cost environment changes, thereby reopening a settlement range. Since large shifts in litigation costs or judicial consistency are relatively rare, these mechanisms alone cannot explain the frequency of modification requests observed in the LASC data. In the next section, we turn to several alternative explanations for why modifications persist despite the availability of private ordering.

D. Possible Explanations for the Modifiation Puzzle

Taking the divergent expectations model as their starting point, Mnookin and Kornhauser (1978) asked a simple question: if negotiated settlements are private ordering, why would divorcing spouses ever proceed to trial, and why, we could add in the spirit of their model, would they ever seek to modify an existing settlement or judgment? The obvious answer captured by their model is that circumstances change. However, if bargaining is efficient, parties should be able to contract around foreseeable contingencies and reach durable agreements. Yet, in practice, custody, visitation, and support orders are frequently reopened and often for reasons that would appear entirely foreseeable. This suggests a broader set of frictions on the efficiency of private ordering.

In fact, Mnookin and Kornhauser (1978) identified several reasons why bargaining may fail: spite between the parties, a general dislike of negotiation that raises transaction costs, fundamental disagreement over the likely trial outcome, or the absence of a meaningful middle ground when the resources at issue are too small relative to the value each parent places on custody. In many ways, these insights suggest several pathways to modifying their model in ways that would explain why some cases are litigated rather than settled.

Over the last 50 years, captured in much of Professor Engels' work, our understanding of how people behave in bargaining negotiations has provided several other explanations for the puzzle of why so many divorce cases are contested and further why so many cases that are settled or adjudicated return to court for modification. The possible explanations explored in this article include: private information and misrepresentation, which distorts bargaining outcomes; negative altruism, which can undermine hard bargaining while still producinginefficiency; externalities to children, whose evolving needs are not internalized by parents; transaction costs of renegotiation and court enforcement; commitment

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

problems that make long-term promises in custody or visitation unstable; legal limits that forbid parties from contracting around certain obligations such as child support; and finally, agency problems stemming from the incentives of lawyers in the family law system. Each of these explanations offers a different perspective on why settlements in divorce are often fragile and why requests for modification are so common. In what follows, I consider each in turn.

I. Bargaining Theory Limits

In its strong form, the Coase theorem predicts that if transaction costs are absent and property rights are clearly defined, parties will bargain their way to efficient outcomes regardless of the initial assignment of rights. This "market-like" intuition underlies much of the early law-and-economics work on divorce: spouses should be able to exchange entitlements to custody, visitation, or support until the allocation maximizes joint welfare, and judgments or settlements should be durable because parties cannot do any better – an extension of the first welfare theorem. In this view, the law matters primarily for distribution, not efficiency. Yet the more recent literature on bargaining theory has suggested problems with this market analogy. For example, in perhaps the most famous critique of the Coasian invariance proposition, Farrell (1987) stresses that bargaining is typically conducted under conditions of incomplete information. In contrast, the Coasian invariance theorem, in its strong form, predicts that if transaction costs are absent and property rights are clearly defined, parties will have enough information to bargain their way to efficient outcomes regardless of the initial assignment of rights. This minimization of information's impact produces the "market-like" intuition that underlies much of the early law-and-economics work on divorce and its predication that spouses should be able to exchange entitlements to custody, visitation, or support until the allocation maximizes joint welfare. In this view, the law matters primarily for distribution, not efficiency, because information problems can be overcome.

Farrell doubts this market analogy, stressing that bargaining is typically conducted under conditions of incomplete information. Divorcing spouses often have private information about income, assets, future plans, or even their capacity to comply with parenting schedules. When parties strategically misrepresent that information, bargaining becomes

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

distorted. Screening and signaling strategies lead to delay, costly litigation postures, or inefficient agreements that later unravel. The implication is that even if Coasian "transaction costs," defined as the cost of enforcing property rights, are minimal, informational frictions prevent bargaining from reliably delivering efficient outcomes.

Binmore et al. (1998) presents a slightly different critique. He argues that bargaining is not a market at all. In markets, competitive forces push resources toward efficient allocations; in bilateral bargaining, outcomes depend heavily on timing, threats, focal points, and institutional context. Bargaining games often have multiple equilibria, and efficiency is not guaranteed. In divorce, this means that legal rules, custody presumptions, support formulas, and domestic violence protections do more than shift bargaining power. They structure the game itself, determining the available strategies and which equilibria are plausible.

More broadly, these critiques suggest that the Coasian optimism embedded in Mnookin and Kornhauser's "shadow of the law" model may be overstated. Divorce bargaining is a bilateral monopoly, not a competitive market. Private information, strategic misrepresentation, and the institutional levers supplied by family law make breakdowns and inefficient outcomes common. If efficiency is not the natural prediction of bargaining theory, then it is unsurprising that divorce settlements so often give rise to post-judgment modification requests.[14]

Formally, in the divergent expectations model, these critiques would involve allowing bargaining to fail not because of differing beliefs alone but because of strategic risk and delay. Let $\theta \in (0, 1)$ denote the probability that bargaining breaks down and proceeds directly to trial, and let $\delta i \in (0, 1]$ be party i's discount factor for delay.

If the custody outcome c is agreed with the transfer T, the expected discounted utility is

$$U_w^S(c, T) = (1 - \theta)\, \delta_w\, [u_w(c) + T] + \theta\, E_w^T,$$

$$U_h^S(c, T) = (1 - \theta)\, \delta_h\, [u_h(c) - T] + \theta\, E_h^T.$$

14 In principle, these informational frictions could also be framed as raising the cost of bargaining, since concealment and misrepresentation often require discovery or enforcement proceedings to resolve. For clarity, in this section, we treat them primarily as distortions to subjective beliefs and return to their transaction-cost dimension below.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Settlement under custody c is feasible iff

$$(1-\theta)(\delta_w+\delta_h)\left[u_w(c)+u_h(c)\right]+\theta\left(E_w^T+E_h^T\right) \;\geq\; E_w^T+E_h^T. \quad (12)$$

Compared to the baseline overlap condition $E_w^T+E_h^T \leq u_w(c)+u_h(c)$, the presence of $\theta > 0$ and $\delta_i < 1$ reduces the effective surplus from settlement. Intuitively, breakdown risk and delay shrink the settlement range, making trial more likely. A reconsideration stage may improve prospects only if the first trial reduces belief dispersion or increases perceived θ, thus shifting the inequality in favor of settlement at Stage 2.

Like much litigation, the LASC family law data reflects this dynamic in discovery disputes, repeated continuances, and evidentiary hearings, all of which suggest that private information and strategic misrepresentation routinely derail efficient settlement.

II. Interdependent Utilities and Spite

The two-stage model showed how updates to beliefs or costs can generate requests for modification. Yet that analysis preserved a central assumption of the Coase theorem: each party's utility depends only on their own consumption or payoffs. In this separable-utility world, inefficiencies arise only from informational frictions or enforcement costs. However, household bargaining models, beginning with Becker, long recognized that spouses' utilities are interdependent: one person's well-being directly enters the other's utility function. In the divorce context, those interdependencies can run in both directions, sometimes positively (altruism), but, in light of the fact that parties are divorcing, more often negatively (spite). Importantly, these interdependencies are not transaction costs but preferences: even if bargaining is costless, utilities that incorporate spite can block efficient trades and destabilize settlements.[15]

15 Recent work by Gilbert and Hayashi (2025)) shows how altruism fundamentally changes the predictions of the Coase theorem. Under the strong version of Coase's claim, two results follow if transaction costs are zero: (a) bargaining produces Pareto-efficient outcomes, and (b) the efficient outcome is invariant to the initial allocation of rights. Gilbert and Hayashi demonstrate that these predictions no longer hold once utilities include others' welfare. With pure altruism, where one party values the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

If one party's utility decreases when the other does well, then bargaining need not converge on efficient or even mutually acceptable outcomes. Divorce disputes often feature elements of spite: parties may be willing to absorb personal costs in order to reduce the other's welfare. This kind of interdependent utility is outside the Coasian framework, yet it is exactly what we observe in practice, and it helps explain why settlements often fail or unravel.

Children complicate the picture further by serving as implicit third parties in the bargaining game. In theory, divorcing parents could negotiate custody and support. However, children are not at the table, even though their welfare is at stake. Parents may internalize some of their children's utility, but altruism toward children is uneven; one parent may weigh the child's needs more heavily than the other, or value the child's welfare only when it reflects back on their own custody or reputation. The result, as Gilbert and Hayashi show in their general analysis, is that efficiency and invariance break down when third-party welfare is only partially embedded in bargaining parties' utilities. Parents may trade away long-term child welfare for short-term gains in custody or support, leaving agreements fragile. The "best interests of the child" standard and the non-waivable character of child support can be seen as institutional correctives: they force the child's welfare into the bargain in a way that private negotiation alone cannot guarantee.

To formalize these intuitions, suppose each party experiences disutility from the other party's utility.[16] Let $\lambda_w, \lambda_h \in [0, 1)$ measure the intensity of spite for the wife and husband, respectively. Under custody c and transfer T,

---

other's outcomes regardless of cause, Pareto efficiency and invariance can survive in a limited sense, but allocative efficiency can fail, because altruistic parties may refuse payments or concessions that make them feel complicit in the other's harm. With impure altruism, "warm glow," where utility depends on having personally caused a benefit, both efficiency and invariance collapse: outcomes depend heavily on initial entitlements, and efficient bargains may be blocked altogether. Importantly, altruism is not a "transaction cost" but a preference. Even with zero transaction costs, altruistic or impurely altruistic utilities can prevent efficiency.

16 In the category of a movie quote being worth several thousand datapoints: "Please, Jess, Marie. Do me a favor, for your own good, put your name in your books right now before they get mixed up and you won't know whose is whose. 'Cause someday, believe it or not, you'll go fifteen rounds over who's gonna get this coffee table. This stupid wagon wheel Roy Rogers garage sale coffee table!" *When Harry Met Sally* (1989). However, for evidence with over 15,000 data points, see Chang et al. (2024).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Under custody c and transfer T,

$$U_w^S(c, T) = u_w(c) + T - \lambda_w [u_h(c) - T],$$

$$U_h^S(c, T) = u_h(c) - T - \lambda_h [u_w(c) + T].$$

With subjective beliefs $\hat{p}_k^i$,

$$E_w^T = \sum_k \hat{p}_k^w \big(u_w(k) - \lambda_w u_h(k)\big) - C_w, \qquad E_h^T = \sum_k \hat{p}_k^h \big(u_h(k) - \lambda_h u_w(k)\big) - C_h.$$

A settlement under custody c is feasible iff

$$E_w^T + E_h^T \;\leq\; u_w(c) + u_h(c) - \big(\lambda_w u_h(c) + \lambda_h u_w(c)\big). \tag{13}$$

Relative to the baseline condition $E_w^T + E_h^T \leq u_w(c) + u_h(c)$, the effective surplus on the right-hand side is reduced by spite terms. Thus, spite shrinks the settlement range and makes trial more likely. At a reconsideration stage, unless spite parameters λ_i change, the overlap condition (13) is unaffected, so new settlements arise only if beliefs or costs change after trial.

In practice, the role of spite is visible in hostile filings, cross-motions, and requests that appear more punitive than instrumental and a refusal to settle when the outcome could, according to judges, be easily predicted, patterns common in the LASC docket and consistent with utilities that include negative interdependence.

III. Transaction Costs

Unlike spite or altruism, which alter preferences, transaction costs operate directly on the feasibility of settlement itself. Even if parties are willing to settle, the additional resources required to implement an agreement may tip the balance toward trial or later modification. In some ways, my approach differs from Mnookin and Kornhauser (1978), who argue that in their divorce model, bargaining is shaped by a broader definition of transaction costs, ranging from financial expense to emotional strain. By contrast, in this paper, I follow Gilbert and Hayashi (2025) and Medema

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

(2020) and adopt a narrower definition than is often found in the Coasian literature: transaction costs are the resources used to establish, defend, and enforce property rights. Under this framing, preferences such as spite or altruism belong to the parties' utility functions, as was done above, and are not the costs of transacting. Thus, transaction costs include the legal fees, enforcement expenses, discovery costs, and procedural hurdles that stand between divorcing spouses and a stable allocation of custody, support, or property.[17]

Even under this narrower definition, divorce is marked by unusually high transaction costs relative to other civil cases. The involvement of minor children amplifies the number of issues to be negotiated, and higher levels of income or property raise the stakes of bargaining. Emotional and psychological conflict further inflate transaction costs by prolonging mediation and increasing reliance on attorneys, even when utilities are otherwise independent. Enforcement poses another challenge: stipulations approved by the court are fragile because parents may renege on visitation, resist relocation, or fail to pay support. Since private enforcement is impossible, each breakdown requires renewed litigation, raising the effective cost of settlement.

Information problems compound these difficulties. Parties often conceal income, misrepresent assets, or exaggerate parenting ability. Establishing the truth requires discovery, expert testimony, and evidentiary hearings, all of which consume resources. In some sense, this is captured by Farrell's critique that bargaining with private information produces inefficiency even when direct negotiation appears low-cost. Children's welfare introduces another structural cost. Because children are not parties to the bargain, the court must represent their interests. As those interests evolve with age and circumstances, agreements that seemed workable at one point in time may quickly become obsolete, requiring further legal intervention.

Finally, legal rules themselves operate as sources of transaction costs. Although courts often rubber-stamp stipulated judgments, family law im-

17 Information problems compound these difficulties. Parties often conceal income, misrepresent assets, or exaggerate parenting ability. Establishing the truth requires discovery, expert testimony, and evidentiary hearings, all of which consume resources. Although in principle such informational frictions could be modeled as shifts in subjective beliefs (as in the previous section), in practice they also generate substantial enforcement and evidentiary costs, so here they can be treated as transaction costs in their own right.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

poses mandatory mediation, statutory formulas for child support, and procedural requirements constraining bargaining. These measures reflect policy judgments about child protection and fairness, but they also act as frictions: they limit the bargaining set, delay resolution, and increase the resources required to reach a durable settlement. In short, even when parties are motivated to settle, the cost of defining, defending, and enforcing property and custody rights in divorce remains high, which may go a long way to explaining why settlements are more fragile in family law and, hence, why modification requests are so common.

To formalize this intuition in the divergent expectations model, imagine that in addition to trial costs C_i, each party incurs a transaction cost $S_i \geq 0$ when negotiating and implementing a settlement (e.g. mediation fees, drafting, enforcement filings). If custody c is agreed with transfer T, then

$$U_w^S(c, T) = u_w(c) + T - S_w, \qquad U_h^S(c, T) = u_h(c) - T - S_h.$$

Trial payoffs remain as in the baseline model:

$$E_w^T = \sum_k \hat{p}_k^{\,w} u_w(k) - C_w, \qquad E_h^T = \sum_k \hat{p}_k^{\,h} u_h(k) - C_h.$$

A settlement under custody c is feasible iff

$$E_w^T + E_h^T \;\leq\; u_w(c) + u_h(c) - (S_w + S_h). \tag{14}$$

This condition is the settlement analogue to the divergent expectations model's baseline inequality, adjusted for settlement costs. Relative to the original condition $E_w^T + E_h^T \leq u_w(c) + u_h(c)$, the effective surplus is reduced by the combined transaction costs $S_w + S_h$. This shrinks the settlement range and makes trial more likely at Stage 1. At a reconsideration stage, if the second trials involve additional incremental costs $K_w + K_h$ exceeding $S_w + S_h$, then (14) may newly hold, making post-trial settlement more attractive.

Empirically, transaction costs show up in enforcement hearings, legal and filing costs, and procedural delays in LASC, where even stipulated judgments frequently return to court because the cost of enforcing and sustaining them appears to be high.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

IV. Commitment Problems

Unlike transaction costs, which reduce surplus through resource expenditures, commitment problems reduce surplus through anticipated noncompliance: parties discount the value of agreements because they expect them to unravel. The conventional Coasian framework assumes that once a bargain is struck, it can be costlessly enforced. Enforcement is taken for granted: in Coase's canonical examples, the farmer and railroad can write a contract, and if one side breaches, courts (or private mechanisms) perfectly ensure compliance. Under this assumption, commitment problems are irrelevant. Once agreement is reached, the bargain is stable, and the efficient allocation is realized.

Divorce bargaining looks very different. Agreements about custody schedules, child support, schooling choices, or relocations are ongoing, testing the limits of already frayed relationships. They are not one-off transfers. They are especially fragile because enforcement is weak. Child support is technically enforceable, but requires continuous monitoring and repeated motions to collect; arrears are common. Custody and visitation orders are even harder to police: courts and law enforcement are reluctant to intervene in day-to- day disputes over pick-ups, extracurricular activities, or holiday schedules, and doing so can be traumatic for children. Unlike contracts for goods or money, there is no private enforcement option – no "repo man" for visitation rights.

As a result, parents strike bargains in the shadow of anticipated defection. Each side knows an agreement may not stick, so they hold back, overinvest in litigation, or design rigid rules that quickly become inefficient. The instability of bargains generates recurrent inefficiency: parties anticipate breakdowns, return to court, or fail to reach agreement in the first place. This is why modification petitions are so common: they substitute for reliable enforcement mechanisms. Ex ante, parties distort their bargains by shading toward self-protection (e.g., requesting sole custody to reduce reliance on the other parent). Ex post, agreements often collapse, leading to renewed litigation.

To capture this dynamic, suppose settlements are imperfectly enforceable: with probability $\lambda \in [0, 1]$, a party reneges or the arrangement fails, reducing the realized surplus. Equivalently, let λ denote the expected "leakage rate" of settlement surplus. If custody c is agreed with the transfer T, then each party anticipates only a fraction $(1 - \lambda)$ of the promised

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

payoffs:

$$U_w^S(c, T) = (1 - \lambda)\,[u_w(c) + T], \qquad U_h^S(c, T) = (1 - \lambda)\,[u_h(c) - T].$$

Trial payoffs remain as in the divergent expectations model:

$$E_w^T = \sum_k \hat{p}_k^{\,w} u_w(k) - C_w, \qquad E_h^T = \sum_k \hat{p}_k^{\,h} u_h(k) - C_h.$$

Settlement under custody c is feasible iff

$$E_w^T + E_h^T \;\leq\; (1 - \lambda)\,[u_w(c) + u_h(c)]. \tag{15}$$

Relative to the baseline condition in the divergent expectations model, $E_w^T + E_h^T \leq u_w(c) + u_h(c)$, the enforceable surplus is now scaled down by $(1 - \lambda)$. This shrinks the settlement range and makes a trial relatively more attractive, since court-imposed orders may be easier to enforce. At a reconsideration stage, if the court's involvement lowers λ, then (15) may newly hold, opening the door to post-trial settlement or renegotiation. Empirically, high rates of child support arrears and recurrent custody modification petitions in LASC are consistent with a high λ: parties discount settlements heavily because they anticipate defection or weak enforcement, which helps explain the volume of post-judgment litigation.

V. Legal Limits (Non-Contractibility, Third-Party Rights)

Unlike transaction costs or commitment problems, which reduce the effective surplus from settlement, legal limits reshape the bargaining set itself by ruling out certain outcomes ex ante. In practice, not all custody arrangements or transfers are legally enforceable. Some outcomes are barred outright by statute or by the embedded rights of third parties, such as children. Family law deliberately restricts the bargaining domain by forbidding parents from contracting away certain obligations. For example, child support is, by statute, modifiable regardless of agreement, and custody stipulations are always subject to judicial review under the "best interests of the child" standard. These limits are not simply transaction costs or bargaining breakdowns; they are institutional veto points that constrain what private ordering can achieve.

The strong invariance version of Coase's theorem presumes a well-defined legal entitlement: the farmer has the right to clean air, or the rail-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

road has the right to pollute. Once such a right exists, parties can bargain around it at zero transaction cost to reach an efficient allocation. But if the law refuses to recognize a right in the first place, whether because of moral concerns (as in organ sales) or deliberate vagueness (as in airplane seat recline), then there is no entitlement to trade, and hence no, or at least less, Coasian bargaining space. Divorce has these issues. Sometimes rights do not exist at all in alienable form, as when parents cannot agree to waive child support. Other times, rights exist but are held "in trust" for children, who cannot bargain for themselves. The bargaining set is narrower than Coase's framework assumes in both cases.

These legal limits directly affect the durability of divorce settlements. Property division and spousal support, which are more contractible, tend to be relatively stable. Custody and child support, which are legally non-waivable, are inherently less durable because the law reserves the power to reopen or override agreements. The observed instability of settlements in family court is not just a function of transaction costs, information asymmetries, or commitment problems. It reflects that the law prohibits fully private ordering in domains where children's welfare is at stake.

More broadly, family law highlights a feature Coase never addressed: the deliberate withholding of alienable rights when trades would harm third parties. In divorce, this is not a marginal issue but the central feature of custody and child support law. Parents may strike agreements, but those bargains are always provisional, subject to judicial oversight and the possibility of modification. Durability arises only when the law resolves the third-party problem directly, for example, by declaring one parent unfit or allowing a child of sufficient age to express their custodial preference. In all other cases, bargaining outcomes remain legally contingent, illustrating how legal rules not only shape distribution but also define the scope of bargaining itself.

Formally, let the feasible custody set be $\mathcal{C} \subseteq \{W, H, J\}$, and let each custody outcome $c \in \mathcal{C}$ allow only transfers in $[\underline{T}_c, \overline{T}_c] \subseteq \mathbb{R}$. A settlement (c, T) is feasible iff

$$T^W_{\min}(c) \;\le\; T \;\le\; T^H_{\max}(c), \quad \text{with} \quad T \in [\underline{T}_c, \overline{T}_c],\ c \in \mathcal{C}. \tag{16}$$

Equivalently, feasibility requires

$$E^T_w + E^T_h \;\le\; u_w(c) + u_h(c),$$

and the existence of a transfer T in the legally permitted interval.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Legal limits reduce the feasible custody set and narrow the transfer interval relative to the baseline divergent expectations model. This shrinks the settlement range and makes trials more likely since the court may impose outcomes that parties cannot privately contract. At a reconsideration stage, if judicial authority expands the set of admissible outcomes (e.g. by authorizing joint custody when parties could not), condition 9) may newly hold and enable post-trial settlements.[18]

In LASC, this legal structure is reflected in the fact that custody and support orders are the categories most frequently subject to modification. In contrast, property and spousal support, having far fewer limits on the terms parties are allowed to negotiate, are far less likely to return to court.

VI. Agency Problems (Lawyer Incentives)

Another possible source of bargaining failure in divorce lies not in the parties' own preferences or transaction costs, but in the incentives of their legal representatives. In many areas of civil litigation, lawyer and client interests are closely aligned. As Spier (2007) observes, contingent fees in tort cases give both lawyer and client the same objective: maximize the expected settlement quickly, since delay or wasteful litigation reduces their joint surplus. By contrast, family law overwhelmingly relies on hourly billing or flat retainers. In divorce cases, by contrast, the absence of contingent fees means lawyers face no direct cost from delay and may even benefit from prolonged conflict. This fee structure creates a classic principal–agent problem. While divorcing spouses often want

18 These dynamics are more complex when legal protections themselves reshape bargaining incentives. Consider Engel and Bar-Gill, whose paper shows experimentally that the strength of legal remedies shapes willingness to trade entitlements in heterogeneous ways (Bar-Gill and Engel, 2018). They find that some participants become more willing to transact when rights are strongly protected ("positive types"), others are less willing ("negative types"), while some are unaffected. This heterogeneity implies that legal protection can either expand or collapse the settlement range depending on the parties' fairness norms. In terms of the taxonomy here, this is less a matter of spite than of commitment and legal limits: the way rights are defined and enforced influences the credibility of bargained-for transfers and whether the parties perceive settlement as fair or legitimate. In divorce, the strength of custody or property entitlements may similarly entrench one party. At the same time, stronger rights can make the other party more willing to compromise, which helps explain bargaining failure even when efficient trade is possible.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

to resolve their disputes efficiently and durably, their lawyers may have financial incentives to prolong litigation, encourage discovery disputes, or craft settlements likely to produce future disputes that generate additional fees.[19]

Attorneys also shape bargaining outcomes in subtler ways. By controlling the flow of legal and factual information, they influence how clients perceive risks and how the "shadow of the law" is framed. Aggressive lawyers may downplay litigation risks to justify a trial, while risk-averse ones may exaggerate them to encourage a settlement. Retaining a high-priced or aggressive attorney can even serve as a costly signal of willingness to litigate, distorting bargaining in ways consistent with Farrell's costly signaling problems. Moreover, inequalities in lawyer quality and the fact that many parties are self-represented further amplify asymmetries. A weaker party with no representation or an outmatched lawyer may agree to terms that are unstable or unfair, leading to later modification petitions.[20]

Formally, in family court, lawyers may have objectives that diverge from their clients' interests. Let $\gamma_i \geq 0$ denote the extra utility counsel derives from trial (e.g. billing, reputation) and $\sigma_i \geq 0$ denote the disutility from settlement (e.g. lower fees, less visibility).[21]

---

19 What is less clear is the importance of the repeat-play nature of family law, which magnifies these agency problems. Unlike tort cases, divorce need not end with the final judgment. Custody, visitation, and support arrangements frequently return to court for enforcement or modification. Lawyers, therefore, may anticipate recurring disputes and, consciously or not, could design fragile settlements, for failing to advise their clients about the fragile nature of the settlement the parties think they want, which ensures future involvement.

20 These incentive problems have inspired new experiments in legal practice, such as Hello Divorce, which operates on a flat-fee basis and combines document automation with attorney or mediator support. The goal is to reduce reliance on hourly billing and to realign lawyer incentives with client welfare by lowering costs, improving predictability, and facilitating cooperative resolution. Hello Divorce has been able to experiment with this model in Utah's regulatory "sandbox," which permits nonlawyer ownership and innovative delivery structures. See Stanford Law School, *Legal Innovation After Reform* (Sept. 29, 2022), https://law.stanford.edu/2022/09/29/stanford-law-report-on-legal-innovation-after-reform. However, in most states, such models remain constrained by unauthorized practice of law (UPL) rules and restrictions on nonlawyer ownership of law firms. These structural legal limits mean that the efficiencies promised by flat-fee, tech-enabled platforms cannot be fully realized outside of experimental jurisdictions like Utah.

21 Unlike the contingent fee arrangements often studied in civil litigation, family law and divorce attorneys are typically retained on a flat fee basis with an upfront retainer.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

For a settlement (c, T):

$$U_w^{S,\text{agent}}(c, T) = u_w(c) + T - \sigma_w, \qquad U_h^{S,\text{agent}}(c, T) = u_h(c) - T - \sigma_h.$$

For trial:

$$E_w^{T,\text{agent}} = E_w^T + \gamma_w, \qquad E_h^{T,\text{agent}} = E_h^T + \gamma_h.$$

A settlement under custody c is feasible iff

$$E_w^{T,\text{agent}} + E_h^{T,\text{agent}} \leq u_w(c) + u_h(c) - (\sigma_w + \sigma_h). \tag{17}$$

Relative to the baseline condition, lawyer incentives both inflate the perceived trial values (via $+\gamma_i$) and reduce the effective settlement surplus (via $-\sigma_i$). This shrinks the settlement range and makes trial more likely than under client preferences alone. At a reconsideration stage, if fee budgets are depleted or reputational motives change, the agency distortions may diminish, expanding the overlap and making post-trial settlement more likely.

Broadly, this aligns with what we observe at LASC: cases with higher attorney involvement are more likely to generate repeated motions and post-judgment hearings, consistent with either selection effects, lawyers are involved in more contentious cases, or lawyer incentives, that is, lawyers continuing the case when the outcome is not entirely predictable to increase fees. Several court personnel have suggested that one of the primary causes of a stipulated judgment request in a contentious divorce case is that the parties have exhausted their retainers and can no longer afford their lawyers.

VII. Model Summary

The six deviations from the Coasian benchmark each shrink the settlement range, but they do so through different mechanisms. Bargaining theory limits discount the value of settlement through delay and breakdown risk; spite directly reduces the joint surplus by embedding negative interdependencies; transaction costs deduct resources required to

This changes the incentive structure: lawyers have limited upside from settlement efficiency but secure compensation regardless of outcome.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Table 3: Comparative Statics of Settlement Range: Baseline vs. Six Modifications

Modification	**Change to Feasibility Condition (vs. baseline $E_w^T+E_h^T \leq u_w(c)+u_h(c)$)**	**Effect on Settlement Range**	**Request for Modification (after settlement)**	**Request for Modification (after trial)**
Baseline	$E_w^T+E_h^T \leq u_w(c)+u_h(c)$	Full range of feasible custody/transfer combinations	Typically unnecessary unless beliefs/costs or enforceability change ex post.	Possible if beliefs converge and/or incremental second-trial costs appear.
#1 Bargaining theory limits	Effective surplus scaled/shifted by breakdown risk θ and discounting δ_i: $(1-\theta)(\delta_w+\delta_h)\,[u_w+u_h] \;+\; \theta(E_w^T+E_h^T) \;\geq\; E_w^T+E_h^T$	Range shrinks due to risk of breakdown (θ) and discounting (δ_i)	Unlikely unless θ falls (e.g., better mediation) or timing improves.	More likely if first trial reduces belief dispersion or raises K_w+K_h.
#2 Spite (negative altruism)	$E_w^T+E_h^T \leq u_w+u_h \;-\; (\lambda_w u_h + \lambda_h u_w)$	Range shrinks by subtraction of λ_i terms from total surplus	Unlikely unless spite parameters λ_i change (counseling, cooling-off).	Only via updated beliefs/costs; spite alone does not relax the condition.
#3 Transaction costs	$E_w^T+E_h^T \leq u_w+u_h \;-\; (S_w+S_h)$	Range reduced by explicit deduction of settlement costs $(S_w + S_h)$	Possible if court can approve terms that reduce S_w+S_h.	Likely if incremental modification costs, $K_w+K_h > S_w+S_h$.
#4 Commitment problems	$E_w^T+E_h^T \leq (1-\lambda)\,[u_w+u_h]$	Range scaled down proportionally by factor $(1-\lambda)$	Common: convert private deal into enforceable order to lower λ.	Likely if court can impose monitoring or enforcement that lowers λ.
#5 Legal limits (non-contractibility, third-party rights)	Require $c \in \mathcal{C}$ and $T \in [\underline{T}_c, \overline{T}_c]$ and $E_w^T+E_h^T \leq u_w+u_h$	Range restricted by narrowing custody set $\mathcal{C}$ and transfer interval $[\underline{T}_c, \overline{T}_c]$	Yes if court approval expands feasible transfers/outcomes beyond private $\mathcal{C}$.	Yes—court may choose from a broader menu than parties can contract over.
#6 Agency problems (lawyer incentives)	$(E_w^T+\gamma_w)+(E_h^T+\gamma_h) \leq u_w+u_h-(\sigma_w+\sigma_h)$	Range shrinks due to inflated trial payoffs ($+\gamma_i$) and reduced settlement surplus ($-\sigma_i$)	Possible if clients override counsel or fee terms change (reducing σ_i).	Likely if fee budgets deplete or reputational motives fade (lower γ_i).

https://doi.org/10.2271/9783748847287 https://www.inlibra.com/de/agb - Open Access - (cc) BY

negotiate and enforce agreements; commitment problems scale down the enforce- able surplus via the risk of defection; legal limits narrow the bargaining set by withholding alienable rights; and agency problems distort both trial and settlement payoffs through lawyer incentives. These distinctions matter for prediction: each deviation implies different conditions under which post-judgment modification requests are likely to arise. Table 3 summarizes the comparative statics across the six modifications to the divergent expectations model.

E. Conclusion

The central promise of the Coase theorem is that, when transaction costs are low and entitlements are well defined, private bargaining should yield efficient and durable outcomes regardless of the initial allocation of rights. In family law, Mnookin and Kornhauser's *Bargaining in the Shadow of the Law* framework adapted this insight to divorce, suggesting that default rules merely structure the bargaining set while efficiency is achieved through settlement. Yet the experience of family courts, and our evidence from the Los Angeles Superior Court, suggests that this Coasian optimism maybe overstated. Settlements in divorce are fragile, and requests for modification are routine.

This paper provides a thought experiment on extending the divergent expectations model to capture these deviations from the Coasian benchmark. Each friction, including informational asymmetries, interdependent utilities such as altruism or spite, transaction costs of enforcement and renegotiation, commitment problems in ongoing parental relationships, legal limits rooted in third-party rights, and the misaligned incentives of lawyers, can be modeled as a modification to the feasibility condition that determines whether settlement occurs. The common effect is to shrink the range of transfers and custody arrangements both parties prefer to trial. Narrower settlement ranges make trial more likely in the first instance and ensure that they are less durable even when settlements are reached. Once beliefs shift, enforcement costs accumulate, or statutory limits are invoked, the feasibility condition can fail, triggering petitions for modification.

The resulting picture is consistent with the observed high rate of renegotiation after judgment. Divorce bargaining is not a one-shot contract over divisible goods, but an ongoing bilateral monopoly over the fu-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ture of children, property, and financial support. Once expanded along the lines suggested by these Coasian limits, the divergent expectations model provides a broader explanation of why settlements unravel and why family courts devote so much of their docket to modification hearings. In this domain, the law matters not only for distribution but also for efficiency, and the Coase theorem serves less as a prediction than as a counterfactual benchmark against which the pervasive instability of family bargaining can be understood.

References

Bar-Gill, O. and C. Engel (2018). How to protect entitlements: An experiment. *The Journal of Law and Economics 61* (3), 525–553.

Binmore, K., C. Proulx, L. Samuelson, and J. Swierzbinski (1998). Hard bargains and lost opportunities. The Economic Journal 108 (450), 1279–1298.

Chang, Y.-c., D. T.-w. Hung, C.-C. Lin, and J. T.-y. Wang (2024). Emotional bargaining after litigation: An experimental study of the coase theorem. *Journal of Empirical Legal Studies 21* (4), 786–825.

Cooter, R., S. Marks, and R. Mnookin (1982). Bargaining in the shadow of the law: A testable model of strategic behavior. *The Journal of Legal Studies 11* (2), 225–251.

Donohue III, J. J. (1990). Opting for the british rule, or if posner and shavell can't remember the coase theorem, who will. *Harv. L. Rev. 104*, 1093.

Engel, C. and O. Bar-Gill (2024). Bounded cognition and the taking of personal property: An experiment. *American Law and Economics Review*.

Engel, C. and W. Güth (2018). Modeling a satisficing judge. Rationality and Society 30 (2), 220–246.

Engel, C. and E. Helland (2023). Does the fundamental transformation deter trade? an experiment. American Law and Economics Review 25 (1), 338–380.

Engel, C. and R.-M. Rahal (2019). Justice is in the eyes of the beholder: Eye tracking evidence on bal- ancing normative concerns in torts cases. Unpublished manuscript, Max Planck Institute for Research on Collective Goods, Bonn.

Engel, C. and B. Rockenbach (2024). What makes cooperation precarious? *Journal of Economic Psychol- ogy 101*, 102712.

Engel, C. and K. Weinshall (2022). Diffusion of legal innovations. *Annual Review of Law and Social Science 18* (1), 139–153.

Farrell, J. (1987). Information and the coase theorem. Journal of Economic Perspectives 1 (2), 113–129.

Gilbert, M. D. and A. T. Hayashi (2025). Altruism and coasean bargaining. *Virginia Public Law and Legal Theory Research Paper* (2025–33).

Medema, S. G. (2020). The coase theorem at sixty. Journal of Economic Literature 58 (4), 1045–1128.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Mnookin, R. H. and L. Kornhauser (1978). Bargaining in the shadow of the law: The case of divorce. *Yale lJ 88*, 950.

Spier, K. E. (2007). Litigation. In A. M. Polinsky and S. Shavell (Eds.), *Handbook of Law and Economics*, Volume 1, pp. 259–342. Amsterdam: Elsevier.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Establishing the Individualistic Viewpoint within Instrumental Administrative Analysis

Laurence O'Hara

A. Introduction

Behavioral law today assembles under its roof a great variety of research interests and approaches (there could be no better proof than the present volume) – and Christoph Engel has played a pivotal role in shaping many of them. One main task of the field is clearly the immediate production of empirical observations. But it also has the task to conceptualize behavior and digest empirical knowledge for the broader legal and regulatory discourses. My contribution focuses on this second task. It does so not just because the occasion of an influential mind's jubilee calls for review and reflection,[1] but also because the future influence of the approach rests on its ability to shape conceptual thinking in the legal and policy world beyond singular findings from individual experimental or observational studies. A central recent development in this regard has been the gradual consolidation of a literary canon of behavioral theory (both empirical and normative) and methodology: While the first overarching accounts around the year 2000 built largely on instructive examples (though ones paradigmatic to this day),[2] the last decade has increasingly brought about comprehensive conceptions, which – despite all necessary selectivity re-

1 You could say that this conceptual level of transfer and reconstruction was the starting point of Christoph Engel's journey as a Max Planck leader, if you compare first the book series at Nomos "Common Goods: Law, Politics and Economics", including influential volumes such as *Engel et al.* (eds.), Wissen – Nichtwissen – Unsicheres Wissen, 2002; and then meta-analytical accounts as for example in *Engel/Morlok* (eds.), Öffentliches Recht als ein Gegenstand ökonomischer Forschung, 1998; and *Engel/Schön* (eds.), Das Proprium der Rechtswissenschaft, 2007.

2 *Jolls/Sunstein/Thaler*, A Behavioral Approach to Law and Economics, Stanford Law Rev. 50 (1998), 1471; *Korobkin/Ulen*, Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics, California Law Rev. 88 (2000), 1051; *Sunstein* (ed.), Behavioral Law and Economics, 2003.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

garding individual components – allow for an exhaustive overview over the field and its implications for legal practice and scholarship.[3]

A particularly promising setting for a discussion of this conceptual dimension is the administrative landscape, in which our research institution at Bonn was situated (and specifically tailored to which – again under Christoph Engel's leadership – the institute brought forth several of such overarching treatments of the field[4]). Not only does the German systematic tradition of legal thinking raise special demands for organizing knowledge according to practical needs,[5] which at the same time makes it conducive to such mappings, e.g. by providing connecting factors and points for comparison.[6] German administrative law over the past decades has also become home to a specifically juridical form of policy-thinking, namely the "regulatory" or "instrumental" perspective nowadays most prominently discussed as a core component of *Neue Verwaltungsrechtswissenschaft*.[7] The more legal concepts are construed as tools for prospective influence rather than posterior review, the more do they offer numerous *nexus* for behavioral concepts to connect.

Sure enough, gaps remain between the various disciplinary traditions and practical activities. Law and regulatory landscape do not simply raise

3 Especially noteworthy regarding this comprehensive perspective are the handbooks authored/edited by *Zamir/Teichmann*: Behavioral Law and Economics, 2018; and earlier The Oxford Handbook of Behavioral Economics and the Law, 2014, with a core methodological contribution by *Engel*, Behavioral Law and Economics: Empirical Methods, 125 et seqq.

4 *Engel et al.* (eds.), Recht und Verhalten, 2007; *Petersen/Towfigh*, Economic Methods for Lawyers, 2015; *Engel et al.*, Öffentliches Recht als Verhaltensordung, 2025.

5 An exhaustive account of the different approaches is offered by *Hilbert*, Systemdenken in Verwaltungsrecht und Verwaltungsrechtswissenschaft, 2015. Closely connected is the debate on characteristics and functions of the German doctrinal method, to which, again, Christoph Engel's group has repeatedly contributed, cf. Engel/Schön (n. 1); and *Kirchhof/Magen/Schneider* (eds.), Was weiß Dogmatik?, 2012.

6 Especially on connections with other subjects *Stark*, Interdisziplinarität der Rechtsdogmatik, 2020.

7 An account for an English-speaking audience can be found at *Eifert*, Conceptualizing Administrative Law: Legal Protection versus Regulatory Approach, in Pünder/Waldhoff (eds.), Debates in German Public Law, 2014, 203 et seqq. Otherwise cf. on the development of the perspective *Appel* and *Eifert*, Verwaltungsrecht zwischen klassischem dogmatischen Verständnis und steuerungswissenschaftlichem Anspruch, VVDStRL 67 (2008), 226 et seqq., 286 et seqq.; and on its role in Neue Verwaltungsrechtswissenschaft *Voßkuhle*, Neue Verwaltungsrechtswissenschaft, in Voßkuhle/Eifert/Möllers (eds.), GVwR I, 3rd ed. 2022, § 1 para. 2 et seqq.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

questions that social sciences could answer in straightforward manners. To enhance the instrumental perspective by behavioral insights is a conceptual task in itself. In what follows, I want to shed light on this task, its properties and preconditions of success. I do so largely by discussing three conceptual levels (2.–4.). They can be read as both: Broad idea-historical reconstructions of what had to be undertaken in the bridging of the gaps, and programmatic tasks or quality standards for such undertakings in the future.

B. Stringent individualism as an enhancement of instrumental analysis

Behavioral analysis of administrative law and policy is first and foremost an expansion of the regulatory paradigm from an individualistic perspective – that is, of course, from a perspective of methodological, not of normative individualism.[8] Hence, the first step in its development had to consist in a focus shift towards effects of law in the first place, towards exploring its ability to make real differences in the real world, and also to establish the mentioned specific form of a regulatory theory for policy advise.

For jurisdictions where economic thinking was prevalent and the according methodology accessible, it is obvious why economic analysis of law quickly gained dominant influence within such regulatory perspectives:[9] it provides an analytical tool with which assumptions about behavior can be made for all possible life contexts – sometimes better, sometimes worse, but never not at all. Ultimately, people are more likely to do things that promise to lead to the outcomes they want to achieve. On this basis, particularly with different forms of expected utility analyses, it

8 In that sense, the root is first and foremost *Weber* as moderated by *Schumpeter*, see the helpful summary of the idea-development by *Heath*, Methodological Individualism, in Zalta/Nodelman (eds.), Stanford Encyclopedia of Philosophy (Summer 2024 ed.), https://plato.stanford.edu/archives/sum2024/entries/methodological-individualism (accessed 15.09.25). See also the quintessential elaboration put forward by *Elster*, The Case for Methodological Individualism, Theory and Society 11 (1982), 453.

9 Already in the first generation of law and economics literature, a regulatory orientation clearly surfaces in the quest to identify certain designs of legal institutions to bring about certain results, cf. *Calabresi*, Some Thoughts on Risk Distribution and the Law of Torts, Yale Law J. 70 (1961), 499; or even *Coase*, The Problem of Social Cost, J. of Law & Economics 3 (1960), 1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

became possible to theorize more and more regulatory contexts, including behaviors that are not *per se* market-based, such as crime[10] or the use of natural resources as collective goods[11].

It is equally obvious that sooner or later a legal theory of regulation had to be developed for a jurisdiction like the Federal Republic of Germany. In the US, law is traditionally perhaps more understood as something static that is practiced by lawyers in the courts; that leaves the making of public policy, especially in government agencies (outside cleaved and donation-driven elected politics), as a naturally separated domain for the experts, above all from economics. In Germany, on the other hand, rule of law means *rule by law*. Parliamentary law is to great extents a form of administrative action. Its overarching purposes are certainly determined politically, but its design is the work of specialized civil servants in the executive – anything else would do a disservice to elected officials and the basic law's conception of the democratic will. There is no doubt that juridical science, which was increasingly becoming a vehicle for shaping policy, needed a theory of governance in order to do justice to the political advisory function of lawyers.[12]

Importantly in terms of disciplinary development, the German regulatory perspective, as compared to the Anglo-American policy scheme, remained more diffuse in methodological terms. This can be seen archetypically in the concept of actor-centered institutionalism according to *Mayntz* and *Scharpf*,[13] which has had perhaps the strongest influence on the tradition here in terms of social sciences. The concept is by no means "uneconomic"; on the contrary, the institutions set the framework and the actors act based on their best interests. But the line of analysis never amounted to a tradition of consummate utility or preference analyses, neither positive nor normative; the actors themselves mostly remain vaguely construed collective actors. Of course, a lot of the time, this is heuristically sound, not least in view of our corporatist political tradition and our industrial economy – our entire understanding of politics in the

10 *Becker*, Crime and Punishment: An Economic Approach, J. of Political Economy 76 (1968), 169.

11 *Ostrom*, Governing the Commons: The Evolution of Institutions for Collective Action, 1991.

12 That is the background of the developments summarized in sources in n. 7.

13 *Mayntz/Scharpf*, Der Ansatz des akteurzentrierten Institutionalismus, in Mayntz/Scharpf (eds.), Steuerung und Selbstorganisation in staatsnahen Sektoren, 1995, 39; *Scharpf*, Games Real Actors Play, 1997.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Federal Republic and our political language presuppose that collectives act: not just the Bundestag, the SPD, the Federal Ministry of Finance, also the car industry or its component suppliers. Simplification by personalization of vague groups with vague interests/preferences is often efficient given the need to grasp the issue cognitively. When analyzing effects of administrative law, we regularly encounter excessive complexity even when we ignore the individual level, or parts of it. If I assess, for example, an emissions standard for diesel cars or a speed limit for certain roads as a regulatory output, I am sometimes better off assuming that my impact[14] is perfectly successful, i.e., that everyone affected complies with my regulations. Even then, whether I can expect to achieve my desired outcome – that particulate matter or carbon emissions fall below a certain threshold – may depend on highly complex considerations. From physical-meteorological conditions to possible avoidance behavior, this can depend on countless factors.

None of this changes the fundamental insight, though: if the law aims at real world effects, it usually has to persuade individuals to decide and act in certain ways. Even in highly institutionalized, legally-driven processes such as adapting a company's organization or business practices to regulatory requirements, ultimately individuals need to do certain things. The moment when the decision is made to behave in a certain way is the point at which instrumental effects actually take place – it is the key moment of regulation. Without such moments, administration cannot function, including when managing its own activities. Regulation based on individual motivation and cognition is therefore not a specific field of government activity, but a level of operation whose influence cannot be avoided anywhere.

The fact that – despite this unavoidably crucial rule of the behavioral level – German legal administrative thinking has traditionally not made greater efforts towards additions from economic analysis is – apart from divergent educational backgrounds and faculty constitutions – primarily due to three factors: First, legal-economic literature often has a norma-

14 On the many different facets and potential behavioral channels cf. the summarizing effort under a joint concept of impact studies *Friedman*, Impact: How Law Affects Behavior, 2017.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

tive-individualistic tendency;[15] if an argument rests on welfare analysis,[16] it easily loses its validity in discourse surroundings that take normative rationality seriously, in particular under the "value order" of the Basic Law. Second, expected utility theory and, as a result, the neoclassical model with each further theoretical development became increasingly detached from reality; what every economist is used to as daily business is easily rejected by any lawyer as unrealistic.[17] Thirdly – and most importantly – the ability of economic models to accurately represent decision-making problems, which are actually prevalent in political practice, remains limited because they have to process all aspects that are important to individual actors as value-relevant aspects of consequences. They therefore fail at the latest when individuals are intrinsically motivated by normative obligations. Put simply, if I steal my colleague's cell phone, *Gary Becker* can process all the material advantages and disadvantages of my crime, including the fact that I have to deal with her anger and contempt, and possibly even that I feel a little ashamed of my actions. But my breach of norms itself plays no role in his scheme. Or, to put it in technical terms: to date, no method has been found that could directly process cardinal preferences about norms (ultimately, not even an internally meaningful one with ordinal preferences about norms).[18] However, empirical research provides very clear evidence that norms do play an important, even decisive role for many people in many situations.[19]

15 It is all over the standard material, cf. e.g. the influential accounts by *Cooter/Ulen*, Law and Economics, 6th ed. 2016; *Cooter/Gilbert*, Public Law and Economics, 2022.

16 On the positive-normative divide also *Towfigh/Petersen* (n. 4), 2 et seqq.

17 As said, that usually requires lawyers to resort to more coarse levels of analysis in the first place and explains how they digest basic economic concepts such as incentives differently, cf. *Sacksofski*, Anreize, in Voßkuhle/Eifert/Möllers (eds.), GVwR II, 3rd ed. 2022, § 39, para. 19 et seqq., 27 et seqq.

18 Cf. first the incommensurability-related line of argument in the balancing-literature as summarized by *Petersen*, How to Compare the Length of Lines to the Weight of Stones: Balancing and the Resolution of Value Conflicts in Constitutional Law, German Law J. 14 (2013), 1387. Cf. further the impressive efforts undertaken by *Spohn*, Defeasible normative reasoning, Synthese 197 (2020), 1391, where the developed "ranks" do amount to a preference order, but none gradually-numerically driven.

19 Cf. here the summary by *Hoeft*, The Force of Norms? The Internal Point of View in Light of Experimental Economics, Ratio Juris 32 (2019), 339.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. Disciplinary expansions and canonical consolidations

With that, we are already deep into the second step, namely the revision of individualistic theory formation. The basic idea behind this is generally familiar these days, namely that the picture of individual behavior is gradually being sharpened by the inclusion of empirical findings. It can also be assumed that everyone is familiar with prominent examples of such findings: the framing dependence of judgments and decisions, tendencies of behavior under risk and uncertainty, loss aversion, the fact that sources of dangers are systematically perceived differently than according to the extents and probabilities of possible damages, or even that we react to descriptive and injunctive norms, real or merely perceived.[20] An important facet of this revision process is – despite all the remaining uncertainty – the consolidation of a common set of terms within the canon, which can also be used for interdisciplinary communication. This is something that can now be clearly seen in retrospect, precisely because of developments over the past twenty years.

It is worthwhile in this context to linger for a moment on the idea of norms, because it allows to grasp another facet of the revision process, namely differentiation and diversification of analytical approaches. It concerns especially the difference between a behavioral economic approach and a behavioral approach in the broader sense. Contrary to what one could assume from what I just said, there absolutely are situations where norm-complying behavior can be captured in economic theory. If we play a dictator game with cake, and I, as the dictator, give you a piece because that corresponds to my sense of justice, then my choice is immediately reflected in the consequences. The game is designed in such a way that it only depends on my sense of justice with regard to the cake, and my only way to comply with the justice-norm is by sharing some part of the cake. So it is not meaningless to say: The norm to me is worth so much, 100 g for you, for example, as compared to 1 kg for me. And it goes on: If the designer of the game gives us money instead of cake as the object of moral obligation, she can suddenly make the game much more complex. For example, she can specify that under certain circumstances, an entire group of players wins or loses money, or that players can punish each other. Now my social preference – or rather, my perceived attachment

20 Most have been around since the early accounts in n. 2.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

to the norm of justice – can suddenly be traded off against many other aspects of the game and, in their comparative currency, quantified in the result. So preferences over norms actually do work? Anyone sensitized to the underlying analytical dimension will, of course, have noticed that our designer has created a very strange world. In fact, she has – at least approximately – defined the problem of incommensurability out of this world. She has simply decreed that all behaviors in the game can possibly cause changes only on the same scale. And that's not how the world out there works, which is why many subjects of behavioral analysis cannot be satisfactorily captured in an economic theory scheme, not even in a behavioral economic one.[21] Of course, incentives remain a central factor, and it is often realistic to assume that individuals also develop elaborate concepts of expected utility; think of a team of bankers considering complex rules for their investment behavior. But very often today, analytical concepts must be applied in isolation from genuine theory in the sense of utility functions. Take *Tyler's* concept of legitimacy based on procedural justice.[22] In this theory, deterrence-incentives naturally remain a decision-making criterion – but they have irrevocably transcended the notion of an actual trade-off, of gradual validity as weighed against each other.[23] In this context, also the influential concept of heuristic thinking becomes relevant. Heuristics,[24] as far as I can see, are like norms necessarily coded in a binary-conditional manner; otherwise, the idea of a rule-of-thumb would not work: to simply choose a product according to price as the subjectively most important factor (instead of conducting a comprehensive trade-off given all value-relevant factors) exactly means to think in a non-maximizing way.[25]

Against the backdrop of these considerations, it is easy to see how the field of behavioral analysis could diversify also in terms of other disciplinary and methodological perspectives. After its economics-oriented beginnings, it was initially expanded to include cognitive and social

21 This point is particularly stressed by *Berg/Gigerenzer*, As-if behavioral economics: Neoclassical economics in disguise?, History of Economic Ideas 18 (2010), 133 (142 et seqq. and throughout).

22 *Tyler*, Why people obey the law, 2006.

23 Cf. *Tyler* (n. 22), 22 et seqq., 42 et seqq., 64 et seqq.

24 *Gigerenzer/Hertwig/Pachur* (eds.), Heuristics, The foundations of adaptive behavior, 2011.

25 *Berg/Gigerenzer* (n. 21), especially at 137 et seqq.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

psychological perspectives. Then further empirical approaches were incorporated, particularly from sociology and administrative sciences, but increasingly also from philosophy. Last but not least, approaches from the independent field of empirical legal research, especially experimental legal research, have also been incorporated. It brought with it a methodological opening away from the purity of the incentivized economic lab experiment towards more practice-aligned ways of data collection, for example vignette studies, and recently machine learning/large language models.

D. Re-conceptualizations of the legal material and regulatory practice

The third step concerns the canon in the second order; it primarily relates to the legal system and our concepts of it, where the aim is to incorporate the canon. The concept of nudging, understood as an intended (and essentially exclusively) behavioral intervention, is now well established; we can consider it a distinctive form of instrument.[26] Remaining within the realm of the universal concepts of administrative law (in German: Allgemeines Verwaltungsrecht), however, the greater part of the task still lies ahead. It consists of realizing behavioral effects as a natural part of the regulatory analysis for any part of the instrumentarium.[27] Exactly that is the flip side of the ubiquity of behavioral effects: almost every government measure produces them. Even where the main debate on the surface is about legal or real-world consequences, it may be that the psychological effects are what actually make the difference. Take for example a financial disincentive a government agency imposes in order to coerce an individual into a certain behavior. From a behavioral perspective, it can make a big difference on several levels whether we have a penalty payment to enforce an environmental ban or a reduction in welfare payments to make the receiver look for work. Not only does money also have a diminishing marginal utility; those affected are likely to depend on the money

26 German public law has seen numerous dissertations in the past years, e.g. *Gerg*, Nudging, 2019; *Kronenberger*, Nudging als Steuerungsinstrument des Rechts, 2019; *Kunzendorf*, Gelenkter Wille, 2021; *Gebhardi*, Verhaltensökonomisch informierte Steuerungsinstrumente, 2022; *Baer*, Staatliche Steuerung durch Nudging im Lichte der Grundrechte, 2023.

27 Cf. as first steps along this path the various conceptualizations in *Engel et al.*, Öffentliches Recht als Verhaltensordnung (n 4).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

to varying degrees, and therefore we will tailor the levels of sanctions very differently in the first step: a reduction in citizen's income of a few hundred euros, an environmental violation by a corporate actor of many thousands. Also, due to inertia tendencies and omission bias, it may depend on whether obligations to act or refrain from acting are enforced. It may depend on whether they are presented as taking away or reducing what is given (i.e., income reduction), because of loss aversion and greater risk aversion in the profit frame (tendency toward sure gain). It may depend on whether the specific treatment and practice of the authority as a whole are perceived as fair, and on the language used by the law and administrative practice.

And it not just applies to money as an "economic" variable. We can see this form of parallel effects in all kinds of state legal and communication acts. Compliance with norms, especially compliance with the law, is a set of behaviors that individuals often engage in intuitively and habitually. In this case, an administrative order in the form of a binding command – which initially imposes an additional legal obligation – is not only a primary measure but also a secondary, i.e. an enforcement-measure.[28] The same applies – in normal regulatory operations – to a hearing letter, which, depending on its wording, can easily be perceived by the recipient as a warning threatening formal legal or physical coercive action.

What becomes apparent here, is the strong influence of context – which shows us that behavioral analyses must increasingly be carried out against the backdrop of specialized administrative law: behavior in social welfare law, behavior in trade supervision, behavior in civil service law, and so on. The analysis has the great advantage that it can draw on the respective area-specific rationalities and knowledge bases that are preserved everywhere in concepts of specialized law. This enables it to counteract, to a certain extent, the structural deficits of empirical social research, particularly limitations of internal or external validity, respectively.

A substantial influence on this third conceptual step is the fact that specialist subjects have begun to incorporate the behavioral level in their own foundations. In some cases, regulators themselves have made behav-

28 See in greater detail *O'Hara*, Ordnungsrecht und staatlicher Rechtsgüterschutz, in *Engel et al.*, Öffentliches Recht als Verhaltensordnung (n 4), 206 et seqq.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ioral assumptions,[29] while in others, there is extensive specialist discourse examining regulatory approaches based on behavior. Food and beverage law is certainly particularly prominent in this area, because its labeling requirements (including shock images on cigarettes) raise specific questions of salience, and also because Europe has a tradition of an information model.[30] We find explicit examples for behaviorally based regulatory approaches by the legislative in the law on digitalization.[31] Interestingly, as far as cookie notices are concerned, we can now observe a tendency toward re-revision on a behavioral basis. Initially, a mandatory choice design had been introduced, and now the "cookie pledge"[32] is moving toward a more substantive regulatory approach – precisely because of the behavioral phenomenon that consent behavior has become automated and users – in line with the heuristic approach – have not made any significant efforts to obtain information. A third recent example is the development of the electricity market design.[33] Already because an energy supply system that relies heavily on renewables requires considerably more fine-tuning at considerably more individual points than a system with conventional energy sources, the design must take more complex behavioral effects into account from the outset.

29 Of course, already classic fields of EU law like competetion depend on behavior influence including discussions of instrumental effects such as deterrence effects even by the court rulings themselves, cf. Case T-210/01 General Electric v. Commission, ECLI:EU:T:2005:456 paras 58, 64 et seqq., 75.

30 Most prominently in the regulation on food information, Regulation (EU) No 1169/2011, which naturally has to deal with the conditions of cognitive digestions by consumers, cf. e.g. Rec. 9, 26.

31 Apart from the old example of data protection, also newer levels of legislation rest on assumptions about individual behavior, cf. e.g. Rec. 67, Art. 25 DSA.

32 Summary of the process by the Commission including the opinion of the European Data Protection Board (13.12.23, OUT2023-0098) at https://commission.europa.eu/live-work-travel-eu/consumer-rights-and-complaints/enforcement-consumer-protection/cookie-pledge_en (accessed 15.09.25).

33 Cf. the summary by the Federal Ministry for Economic Affairs https://www.bundeswirtschaftsministerium.de/Redaktion/DE/Publikationen/Energie/20240801-strommarktdesign-der-zukunft.pdf (accessed 15.09.25).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

E. Outlook: Transforming administrative reasoning in transformative times

These last thoughts provide hints as to what we should expect for the future. Social science has potential to (further) transform the administrative world; at the same time, the regulatory body driven by the respective practical needs shapes the meaning any such literature might have. The most major recent shifts in administrative law throughout Europe circled around transformations, with climate law transformation and digital transformation at the center. This remains the case even as a phase of questioning has begun. Obviously, these megatrends hinge on the behavioral level in numerous ways, be it the introduction and enforcement or the repeal of according policies that is being discussed: Transformations cause massive cognitive effort. They require far-reaching decentralized behavior adjustments, often through measures, which have to be actively designed and implemented. Throughout there are prominent concepts from the behavioral literature at place: from defaults, inertia and omission bias to risk perception. Moreover, such shifts involve a large-scale change in norms. Previously established patterns of behavior must be changed; and – as in the digital transformation – the structure of more general norms, such as how we communicate and, not least, how we engage in politics, may change. No matter what one's stand on a need to regulate versus to deregulate: Whether any such measure is worthwhile, is only assessable against behavioral assumptions. Over the last twenty years we have seen significant shifts in what could be called state action paradigms. Given the grown awareness for individual behavior, any further paradigm shift will have to be developed cognizant of the micromechanisms at play.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Data as Facts

Frederike Zufall

A. Legal data sets

As digitalisation increases, the amount of legal resources in the form of data is growing exponentially. Not just legal scientific publications,[1] but also court decisions have been published for many years,[2] with the databases being constantly expanded. Nowadays, statutes, parlamentary documentation, regulatory guidelines or administrative proceedings are widely available as open source. Regulatory approaches towards increasing open data by the EU[3] have likewise accelerated this process, particularly given the economic value of data as training data for machine learning.[4]

At the same time, the availability of data in and on the legal domain, has always been a prerequisite for empirical legal research.[5] The existence and accessibility of digital representations of court decisions not only facilitates research on legal sources beyond the single document, but also

1 Compare databases for legal research such as, e.g., LexisNexis, HeinOnline, Westlaw or beck-online in Germany.

2 Compare the European Court of Human Rights' database of court decisions, available at https://hudoc.echr.coe.int, or of the European Court of Justice, at https://curia.europa.eu or the curated U.S. Supreme Court database by H. J. Spaeth, L. Epstein, A. D. Martin, J. A. Segal, T. J. Ruger, S. C. Benesh, and M. J. Nelson, 'Supreme Court Database (SCDB)'.

3 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, *OJ L 172,* 26.6.2019, pp. 56–83; Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), *OJ L 152*, 3.6.2022, pp. 1–44 .

4 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European Strategy for Data, COM/2020/66 final, p. 2–3.

5 M. Heise, 'The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism' (2002) 2002 *University of Illinois Law Review* 819 at 829–30; compare M. A. Livermore and D. N. Rockmore (eds), *Law as Data: Computation, Text, & the Future of Legal Analysis* (Santa Fe Institute Press, 2019).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

enables quantitative analysis on a whole data set. Investigating, how law works (or not), and affects reality (in the way it aims and claims), may have a broader base and a much greater impact if it can be demonstrated through quantitative methods.

However, empirical legal research did not stop at analysing metadata about judges[6] and their background,[7] or other circumstances, but has, with vector presentation from natural language processing allowed to move this analysis to the text[8] – the actual content of legal decisions, and the way how they apply law to the real world.

Besides this empirical perspective, the rennaissance of the field AI and law in recent years has yielded a number of contributions to the task of 'judgement prediction'.[9] Using huge corpora of court decisions, a machine learning classifier can be trained on the overall outcome decision, indicicating whether or not a new unknown case would be won, or on specific outcome parameters such as the content of a sentencing in criminal procedure. The underlying idea is that similar judgment texts will likely lead to a similar court decision. However, one problem with these approaches is that the training data may contain both the facts and the reasons of the judgments. Knowing already the normative verdict, and reasoning of the text, the prediction may not only be easier,[10] but it also

6 C. Engel, 'Lucky You: Your Case is Heard by a Seasoned Panel – Panel Effects in the German Constitutional Court' (2022) 19 *Journal of Empirical Legal Studies*.

7 Y. Lim, 'An Empirical Analysis of Supreme Court Justices' Decision Making' (2000) 29 *The Journal of Legal Studies* 721–52.

8 E. Ash and D. L. Chen, 'Case Vectors: Spatial Representations of the Law Using Document Embeddings' in M. A. Livermore, D. N. Rockmore (eds), Law as Data, (SFI Press, 2019), pp. 313–37.

9 Compare N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, 'Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective' (2016) 2 *PeerJ Computer Science* e93; I. Chalkidis, I. Androutsopoulos, and N. Aletras, 'Neural Legal Judgment Prediction in English' Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, (Florence, Italy: Association for Computational Linguistics, 2019), pp. 4317–23; L. Yuan, J. Wang, S. Fan, Y. Bian, B. Yang, Y. Wang, and X. Wang, 'Automatic Legal Judgment Prediction via Large Amounts of Criminal Cases' 2019 IEEE 5th International Conference on Computer and Communications (ICCC), (2019), pp. 2087–91; see for an overview: Y. Feng, C. Li, and V. Ng, 'Legal Judgment Prediction: A Survey of the State of the Art' (2022), pp. 5461–69; and J. Cui, X. Shen, and S. Wen, 'A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges' (2023) 11 *IEEE Access* 102050–71.

10 M. Medvedeva, M. Wieling, and M. Vols, 'Rethinking the field of automatic prediction of court decisions' (2022) *Artificial Intelligence and Law* at 3.2.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

confuses the actual process of subsumtion and reasoning that a human lawyer would perform.

B. Law and the real world

Judgment prediction tasks and the way in which they harness existing judgment corpora invite us to consider the extent to which these tasks have a parallel in judicial decision-making or more general in the process of legal decision-making.

The way in which judgments are written, with a clear distinction between the 'facts' and the 'reasons' of the case, is based on the legal theoretical distinction between normativity and reality, between 'is' and 'ought'.[11] This has also been reflected in general procedural principles and codifications of procedure law prescribing different sections for facts and reasons.[12] Distinguishing between the facts found and the legal, normative decision based on these facts corresponds to the different questions and dimensions that court proceedings need to address beforehand: what has actually happened? What can be proven? These questions of fact are subject to specific procedural rules and are addressed through the testimony of witnesses, judicial inspection or documentary evidence. By contrast, questions of law, and the application of law to these facts, are then situated in the normative realm.[13]

These theoretical grounds imply that predicting the outcome from judgment corpora containing the verdict or judgment texts containing the reasons does not equate to the judicial reasoning process. But neither does using the facts section of the texts, because this section was written *after* the parties had been heard and the evidence assessed in order to establish a procedural truth.[14] Moreover, the facts do not only comprise what has

11 D. Hume, *A Treatise of Human Nature* (1739) bk 3, part 1; H. Kelsen, *Reine Rechtslehre: Einleitung in die rechtswissenschaftliche Problematik* (F. Deuticke, 1934) pp. 37–38; F. Schauer, *Thinking like a lawyer: a new introduction to legal reasoning* (Harvard University Press, 2009) pp. 203–6.

12 Compare, for instance, § 313(1) of the German Civil Procedure Code and § 117(2) of the German Administrative Procedure Code. The Seventh Amendment (Amendment VII) of the U.S. Constitution also reiterates this distinction by preserving fact-finding to a jury.

13 Schauer, *Thinking like a lawyer: a new introduction to legal reasoning*, p. 204.

14 J. Bentham, *Rationale of Judicial Evidence: Specially Applied to English Practice : in Five Volumes* (Hunt and Clarke, 1827) bk 1, ch. 1.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

been deemed 'true', but only those facts that correspond to the specific underlying legal issue. The facts section mirrors the factual side of what is prescribed by the normativity of the law. Hence, it can only be written once the final decision and the assessment that led to it have been made. This is different from perceiving reality without the normative lens.

Still, the idea and endeavour of the judgment prediction task and the use of case law corpora for machine learning leads to another, more general underlying question: to what extent can data represent facts? Or to what extent can we treat existing data the same as we treat facts for a legal decision?

C. Reality as data

Beyond the facts that we find written down in existing judgement corpora, the amount of data about the real world is much greater. Think about the data collected by sensors in the context of autonomous vehicles, or CCTV settings. The potential sources here are not only much broader and bigger than the facts in case law data sets. In comparison to the facts section in judgment texts, they are also a rather direct representation of the real world –without the normative lens. It is more closely comparable to what humans perceive as reality. This is also similar in that it corresponds to the factual realm *before* it has reached the courtroom and been subject to evidence, procedure, and filtering to determine its relevance to the legal question at hand. In this light, we can say that the real task of judgment prediction still lies ahead of us: predicting a normative outcome from the actual reality as it presents itself to us.

When we, as humans, speak of reality – and I think this is important when referring to the theoretical legal distinction between is and ought – this is also very much tied to the cognitive dimension. The idea of the *is* has always been tied to what humans may perceive and cognitively process. This is also one reason why the Seventh Amendment to the US Constitution entrusts fact-finding to a jury of non-lawyers, detaching this dimension from the assessment by judges.

If legal decision-making is no longer performed by humans, the question not only arises of how normativity and law can be applied by machines. There is also the question of the associated reality, the questions of fact and how they are processed by machines. This is particularly pertinent given that the range of perception (if we may still call it that way)

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

through hardware is potentially broader than what humans may perceive through their senses. Ultimately, the range of what we consider to be the starting point of our 'is' becomes broader. Or in other words, as our reality becomes more digital, we could argue that data is then not only a potential representation of reality, but extends it. And this has not only consequences on the questions of facts, and how we establish what is true, but it equally extends the potential realm and scope of the *ought*.

D. Data as facts

In fact, we have already witnessed an extension of the *ought* with increasing regulation on data or digital phenomena over recent years. Data as a reality is already subject to various forms of regulation that address (partly or wholly) digital phenomena, such as data protection law, the regulation of online platform providers or illegal content on the internet.[15] Some of this regulation was designed for the offline world but has since been adapted for similar online situations, some has genuinely been adopted to regulate digital phenomena.

But beyond data being subject to data regulation or to digital regulatory frameworks, to what extent can data serve as facts in legal decision-making? This would most likely be the case in domains where the reality is already digital; not because the subject of regulation is data, but because the data is more readily available and more readily susceptible to processing. In criminal law, for instance, there are an increasing number of offences in the digital realm that manifest as data, such as various forms of illegal online content. Other use cases range from data processed by IoT devices to autonomous vehicles.

However, there are still and will always be factual scenarios that only appear non-digitally. For instance, many criminal offenses are committed in the real world and involve physical human activity. In these cases, data

15 Compare, for instance, the EU General Data Protection Regulation, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of such Data, and repealing Directive 95/46/EC; or the Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

has a different role. It can serve as potential evidence, for instance, if the real-world scenario has been recorded by a CCTV. The facts of the case are then not the CCTV recording, but the criminal conduct itself.

The existence of data may blur the lines between what is subject to law as facts and what is merely a data representation of it. When regulating data, the subject of normativity is data; but when hardware collects data on reality, does it only qualify as potential evidence? When we rely on our own human senses to assess what is real, we do not question its existence. However, we do so when we hear testimony from third parties, such as witnesses in court. Existing procedure and evidence law has been built on the trust in human cognition and perception. This does not mean, however, that heuristics and biases do not come into play when finding the facts. [16] But the question we ask when hearing witnesses, is whether we trust what they say, and trust in how they have perceived reality.

In contrast, if we treat data as facts, we would ask whether and under what conditions we believe and trust the data, and what it appears to stand for. As then the facts are not based on human perception, procedural rules for fact-finding would also need to adapt accordingly. And this would open up legal procedure to quantitative empirical methods like never before.

Ultimately, the increasing availability of data challenges our legal and dogmatic framework, not only as we try to adapt to digitalisation, and not only as its use as training data for machine learning increases. But it does so at its theoretical core, when we consider what it is that law addresses, and whether it does so at a procedural – or at the substantive level.

References

Aletras, N., D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, 'Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective' (2016) 2 *PeerJ Computer Science* e93.

Ash, E. and D. L. Chen, 'Case Vectors: Spatial Representations of the Law Using Document Embeddings' in M. A. Livermore, D. N. Rockmore (eds), Law as Data, (SFI Press, 2019), pp. 313–37.

Bentham, J., *Rationale of Judicial Evidence: Specially Applied to English Practice : in Five Volumes* (Hunt and Clarke, 1827).

16 C. Engel, 'Judicial decision-making a survey of the experimental evidence' (2022) p. 7.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Chalkidis, I., I. Androutsopoulos, and N. Aletras, 'Neural Legal Judgment Prediction in English' Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, (Florence, Italy: Association for Computational Linguistics, 2019), pp. 4317–23.

Cui, J., X. Shen, and S. Wen, 'A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges' (2023) 11 *IEEE Access* 102050–71.

Engel, C., 'Lucky You: Your Case is Heard by a Seasoned Panel – Panel Effects in the German Constitutional Court' (2022) 19 *Journal of Empirical Legal Studies*.

Engel, C., 'Judicial decision-making a survey of the experimental evidence' (2022).

Feng, Y., C. Li, and V. Ng, 'Legal Judgment Prediction: A Survey of the State of the Art' (2022), pp. 5461–69.

Heise, M., 'The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism' (2002) 2002 *University of Illinois Law Review* 819.

Hume, D., *A Treatise of Human Nature* (1739).

Kelsen, H., *Reine Rechtslehre: Einleitung in die rechtswissenschaftliche Problematik* (F. Deuticke, 1934).

Lim, Y., 'An Empirical Analysis of Supreme Court Justices' Decision Making' (2000) 29 *The Journal of Legal Studies* 721–52.

Livermore, M. A. and D. N. Rockmore (eds), *Law as Data: Computation, Text, & the Future of Legal Analysis* (Santa Fe Institute Press, 2019).

Medvedeva, M., M. Wieling, and M. Vols, 'Rethinking the field of automatic prediction of court decisions' (2022) *Artificial Intelligence and Law*.

Schauer, F., *Thinking like a lawyer: a new introduction to legal reasoning* (Harvard University Press, 2009).

Spaeth, H. J., L. Epstein, A. D. Martin, J. A. Segal, T. J. Ruger, S. C. Benesh, and M. J. Nelson, 'Supreme Court Database (SCDB)'.

Yuan, L., J. Wang, S. Fan, Y. Bian, B. Yang, Y. Wang, and X. Wang, 'Automatic Legal Judgment Prediction via Large Amounts of Criminal Cases' 2019 IEEE 5th International Conference on Computer and Communications (ICCC), (2019), pp. 2087–91.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Rational Choice Modeling of Sustainable Corporate Social Responsibility

Werner Güth & Hartmut Kliemt

The intuition that corporate social responsibility CSR can serve as a functional equivalent of the moral conscience ascribed to personal actors can be restated in elementary Rational Choice Modeling terms. To serve this function, engaging inner constitutional commitments to act in the sense of CSR must be a viable option for corporate actors. Moreover, these commitments must be detectable by other actors with sufficient certainty at sufficiently low costs. Only where these requirements are fulfilled there is a niche for implementing the CSR ideals of stakeholder conceptions within the constraints of ordinary market competition and with respect for stakeholder property.

A. Introduction and overview

We share Christoph Engel's long-standing interest in 'trust'.[1] But like him and most others who approach trust problems within a rational choice perspective, we have mostly sidelined the fact that a kind of community of trustors and trustees who share certain aims, ends and values is typically involved. When trustors and trustees interact repeatedly then the growth or decline of trust of the trustors is not simply a matter of Bayesian updating of probabilities concerning the behavior of trustees.[2]

We can properly rely on the stability of a bridge but hardly properly trust in the bridge like we might trust in our physician as willing us well in using her discretionary powers as our agent even if there is no direct material incentive for this. But we cannot properly trust a company in the

1 Engel's views on 'generalized trust' in relation to experimental results see (Engel 2010; 2018); our own interest in the related general commitment problem arose from reading Selten (Güth) respectively Schelling (Kliemt). Our interest in modeling the evolution of trustworthiness resp. reliability started with (Güth and Kliemt 1994). We strangely enough completely missed out on the possibility of running an experiment on trust as in (Berg u. a. 1995)).

2 See on this in detail (Lahno 2002; 1995a; 1995b)).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

same way as we trust a "significant other". Knowing this, we nevertheless speak and act in our relations with corporate actors as if we were continuously dealing with personal actors.

In this paper we allow for this way of speaking as well. Yet, we start from the tacit understanding that under the Rule of Law the reliability of corporations can be a functional equivalent of genuine trustworthiness. Where trustworthy natural persons are seen as intrinsically motivated by general moral rather than extrinsic material motives arising from the exigencies of separate choice-situations, corporations can be seen as being guided by internal constitutional rules. These rules are internal to the corporate actor and influence its behavior in external relations like commitments of personal actors.

We employ the 'lingua franca' of modern social theory or what we call "Rational Choice Modeling", RCM, which is mostly neutral with respect to the specific substantive Rational Choice Theory, RCT, employed, to illustrate how intrinsic can be represented besides extrinsic motivation explicitly in the same model of a corporate actor.[3] The ways and means of modeling intrinsic and extrinsic motivation in the case of personal actors within a "dual payoff function approach" that we have studied before can be transferred to the case of corporate actors. Thereby it becomes possible to provide a framework for studying the co-evolutionary process in which the population share of CSR actors develops in competition with non-moral profit-seeking personal and corporate actors. The upshot is that even in a world ultimately dominated by objective success – as measured in terms of profits and losses, spin offs and bankruptcies etc. – there can be a niche for decision-making that deviates from what measures of objective success would prima facie suggest. [4]

The indirect evolutionary approach models conditions under which CSR-oriented corporate actors can succeed in competition with purely profit-oriented actors even if we abstract of all personal relations by assuming random matching of interaction partners. This may be viable *provided* that internal corporate constitutional commitments to CSR can in principle be implemented and can be detected by potential interaction

3 For the relationship of RCM and, rational choice theory, RCT, see (Güth and Kliemt 2007).

4 To relate this to a rather popular management theory, we seek to represent an aspect of the general obliquity of (boundedly) rational human pursuits (Kay 2011) in precise RCM terms for the exemplary case of CSR.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

partners. We believe that contrary to commonsense views firms as corporate actors have – at least potentially – a comparative advantage over personal actors when it comes to developing a functional equivalent of a conscience and to signal their 'firm commitments' transparently.[5]

CSR can *ultimately* be a case of doing well by doing good. To the extent that this is the case it renders CSR compatible with the *normative* demands of pursuing shareholder value in a private law society.[6]

B. Sustainable *CSR*

Like a virtuous personal actor who is by some intrinsic motivational process kept from following proximate extrinsic incentives, the good corporate citizen adopts corporate social responsibilities that seem directly unfavorable to the corporate actor's bottom line. As in the case of natural selection of virtuous individuals who are restricted by inner commitments in their opportunity taking behavior it seems initially puzzling how it can ever be (even in a wide sense of the term) "profitable" for a fully rational corporate actor to lack or to forego the ability to realize opportunities. After all, a fully rational actor endowed with "foresight and understanding" (Toulmin 1961) takes into account all future causal consequences of actions as foreseen by her in the instance of choice making. As a rational actor she knows that she will never choose a dominated alternative in the future. So, what might she gain by committing to do just that?

By assumption the rational actor does not suffer from weakness of the will and other imperfections. Therefore, she will be better off with more rather than fewer options no matter what. Since she will always be free to choose in view of *all* future consequences of her choices, she seems to have no use for commitments that restrict her future option sets. [7]

In a so-called game against nature avoiding myopic behavior would suffice. But it does not suffice in the case we are interested in. The selection

5 Alluding to the nice pun of (Mayer 2013).

6 Private law society in the sense of (Böhm 1989; 1966); a defense of an obligation to increase shareholder value as a demand implied by promises given within the legal order of private law societies can be found in (Kliemt 2022).

7 In his popular "economics in one lesson" Hazlitt describes the focus on all – including the most remote consequences of action – as the hallmark of rational economic behavior (Hazlitt 1988).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of behavior we are interested in, emerges from the behavior of other individuals who respond to restrictions of opportunism that they perceive in the actor.[8] This is the route we take to provide a positive answer to Robert Frank's challenge: "If Homo Ecomomicus Could Choose his Own Utility Function, Would He Want One with a Conscience?" (Frank 1987)?

I. An elementary example of how inner commitments may pay off

Imagine a simple exchange of apples and bananas (as in Figure 1). Person A has apples, *a,* and person B has bananas, *b*.[9] A prefers *b* to *a* while person B prefers *a* to *b*. Let the apples be ripe before the bananas. So, A needs to deliver his apples, *a,* first in the hope that B will deliver her bananas, *b,* afterwards:

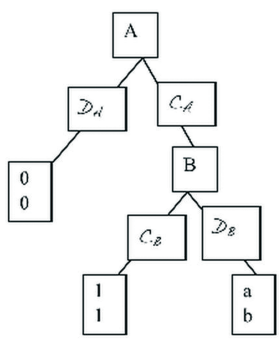


Figure 1 y>1>0>x represents players' preference rankings of plays of the game numerically
a, b, a&b, – in italics in the dotted boxes show substantial results of plays of the game

In Figure 1 a play of the game proceeds along a path from top to bottom. A chooses first, and B second. The numerical values at the end-nodes of the tree are *representations* of subjective preferences concerning the full path

8 See (Zahavi and Zahavi 1997) for instructive analogous examples from the animal kingdom.

9 One would have to think of a certain quality and quantity of apples and bananas at a certain time.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

"all things considered". Of the pairs of rankings corresponding to plays of the game the upper entry shows the ranking by A and the lower the ranking by B.[10] The entries in the dotted boxes show as first entry of a pair what A possesses, as second entry the substantive holdings of B after the corresponding play of the game.

We have – left to right – at the origin of the tree two alternatives open to A: *left, D_A, or non-delivery of apples by A* whose ranking of this alternative is represented by 'utility', "0" (representing the subjective preferences of A); right , C_A, *or delivery of apples a by the first mover A*. After C_A the second mover B can deliver *b* by choosing move C_B – leading to a "*subjective*" overall "payoff" or utility of "1" for, B, and of "1" for A. Alternatively, the second mover B can choose D_B leading to the play (C_A, D_B) and a "subjective" overall payoff of "*y(>1)*" for B – holding a & b after the play (C_A, D_B) – and of "*x(<0)*" for A – holding nothing after the play (C_A, D_B).

If the 'all-things-considered' subjective ordering u_i, $i = A, B$ of the substantive results is the same as the 'natural' ordering ξ_i, $i = A, B$ that considers only substantive results, then a rational actor A who foresees B's evaluation *y>1* would never deliver his apples in the first place. Due to 'backward induction', he would foresee ending up with nothing at all after choosing C_A and therefore would choose D_A to begin with.

Still, both could be better off in naturally ordered holdings than in the status quo. This would be the case if the actor in the second mover role were able to

(i) commit to behavior that goes against the extrinsic motives exerted by the substantive payoffs in the dotted boxes and
(ii) could signal this commitment. If for instance, due to a commitment, the all-things-considered preferences represented by u_B *would* change to 'new' all-things-considered preferences which *could* be represented by u_B^* with $u_B^*(C_A, D_B) < u_B^*(C_A, C_B)$ then this would indicate that actor B –

due to the additional commitment – would prefer the path (C_A, C_B) over (C_A, D_B) even though $\xi_B(C_A, D_B) > \xi_B(C_A, C_B)$.

10 The scale factor and zero points of the numerical preference-representations have been chosen according to some order-preserving positive linear transformation that yields (0, 0) for the status quo and (1, 1) to the status reached by a successful exchange.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Once B can commit and signal her commitment and A can detect the presence of suitable intrinsic motives, backward induction will induce A to anticipate the co-operative play (C_A, C_B) as resulting from his choice. This result will be better than the status quo according to the evaluations of both actors.

So, if homo economicus could in a situation like that of Figure 1 choose his own utility function (s)he would choose one with a conscience.

This seems to answer the "Frank-question". Yet, though interesting it does not show that individuals endowed with a conscience would survive in evolutionary competition. In particular, the familiar problem of mimicry would have to be overcome since it could be better to appear as if having a conscience without having it 'in truth'.

We do not think that it is convincing to think in stark alternatives of non-detectable commitments and fully detectable commitments here. So, we will first make a few remarks about the range of parameter values for which a positive share of conscientious actors might conceivably be evolutionarily stable. The arguments concerning the sustainability of a positive share of committed actors seem precarious since in case of personal actors the mimicry problem is lurking in the background. However, as far as this problem is concerned CSR as a functional equivalent of a conscience seems different and less problematic because its presence may be observable. So, setting aside the problem of mimicry let us first turn to a brief sketch of the basic comparative statics of the adaptive process in which the disposition to "do good" can be sustainable. After that we shall return to mimicry and the sustainability of CSR.

II. A graphical representation of pre-requisites of CSR sustainability

When called upon to act in the second-mover or B-role in exchange relationships transaction partners can be either 'trustworthy/reliable' (characterized by $u_B(C_A, D_B) < u_B(C_A, C_B)$) or 'untrustworthy/non-reliable' (characterized by $u_B(C_A, D_B) > u_B(C_A, C_B)$). When called upon to play in the first-mover or A-Role a rational actor will decide on C_A iff $p_A u_A(C_A, C_B) > u_A(C_A, D_B)(1 - p_A)$;

where p_A is the subjective probability estimate of the player in the A-role that the partner in the B-role is trustworthy. If we assume (i) that p, $0 \leq p \leq 1$, is the share of trustworthy types in the population and (ii) that there is no way to tell the trustworthy ex ante from the non-trustwor-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

thy then under purely random matching we have

$$pu_A\,(C_A, C_B) > u_A\,(C_A, D_B)\,(1-p)\,;$$

as a sufficient condition of offering cooperation in the trustor role. Yet, if information concerning B can be acquired after random matching by actors in the A-role at uniform material cost C, $0 \leq C \leq 1$, then this information will influence the rational expectations of players.

In Figure 2 one finds for alternative levels of costs C* horizontal arrows that indicate how p will evolve at that cost level from any starting combination (p*, C*).

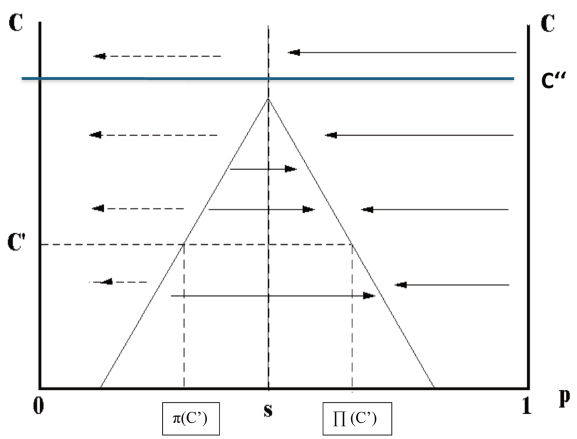


Figure 2 Evolution of population share p of committed actors for alternative cost levels C and given reliability of type detection

In Figure 2 values of p are indicated on the abscissa. On the ordinate we find alternative cost levels C at which a detection [signaling] [11] technology of a certain reliability is available to the interacting parties. Horizontal lines in the triangle indicate for alternative values of C the population

11 The detection and signaling technologies relevant here amount basically to the same thing. Either way a stochastic signal of the probability that the co-player is reliable respectively non-reliable is received by the trustor who can update her beliefs accordingly; (see for analytical details (Güth and Kliemt 2000)).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

compositions for which it is worthwhile to employ a detection technology of given reliability; e.g. for C=C' the relevant interval reaches from $\pi(C')$ to $\sqcap(C')$.

The arrows indicate the direction in which p will evolve whenever evolutionary success monotonically depends on direct substantive success (proximate fitness) on each round of interaction. For all initial (p*, C*) combinations it can be characterized where the process will ultimately be heading for C*.

So, for C′ in Figure 2

if $p \in (\pi(C'), \sqcap(C'))$ the population share of the trustworthy will increase up to $\sqcap(C')$;
if $p < \pi(C')$, the population share of the trustworthy will decrease down to 0;
if $p > \sqcap(C')$, the population share of the trustworthy will decrease down to $\sqcap(C')$.

Generally, the higher the costs of the detection technology, the smaller the interval of p-values for which it is rational to invest in the costly technology. Only for generic intervals the given technology will render a positive share of committed actors sustainable.[12] Assuming an infinite population and random pairing the process will converge either, for $p > \pi(C)$ towards $p = \sqcap(C)$ and otherwise towards $p = 0$.

Due to the dual nature of the conceptual framework underlying the indirect evolutionary approach (i) the choices are made according to subjective expected utility u_A, u_B, yet (ii) the population share p_{t+1} is determined solely according to the substantive success ξ_A, ξ_B of these choices.

The parameter s denotes the population share p=s at which, without access to type information, the expected values of showing trust or not would be equal. Knowing only the population share p, yet not the type of the specific partner, *all* A-role actors will show trust if $p > s$ and will show no trust if $p < s$. Without specific type information untrustworthy types in the second mover role will do better than trustworthy types. The untrustworthy will crowd out the trustworthy if beyond knowledge of the population share p of trustworthy types, no type-information is available

12 Figure 2 shows the relevant triangle for a technology of given reliability. This technology makes two signals available, one of trustworthiness and one of non-trustworthiness. The signals may differ in reliability yet since nothing substantial hinges on this for the present purposes we leave it out of account.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

at sufficiently low costs: As long as $p>s$, untrustworthy B-role actors will exploit the rational bets placed by A-role actors. If $p<s$, untrustworthy B-role actors will be offered a chance to exploit first-mover trust only if A-role actors erroneously show trust. Since the chances to exploit "noise trust", $p<s$, are less frequent than under conditions in which trust is the better choice for all A-role actors, the noise-dynamic is slower – indicated by dashed arrows – than the choice-dynamic – indicated by bold arrows. Obviously if A-role actors can make use only of their knowledge of the population composition p then ultimately the share p will converge towards $p=0$.

In sum, from our modeling approach we know that for any $C<C''$ not only $p=0$ but also $p>0$ may be an evolutionarily stable rest point. (For $C>C''$ the process converges to p=0 irrespective of the (initial) type composition.) A precise RCM makes it plausible that there can be a "niche" for intrinsically motivated moral types even in large scale competitive interactions.

III. Intermediate conclusion

According to the preceding line of argument the a priori thesis that market competition of necessity will drive out CSR investments is untenable.[13] Our example of pursuing an indirect evolutionary approach shows that it is at least conceivable that individuals who due to their commitments are merely boundedly opportunistic will not necessarily be crowded out by uncommitted competitors. Provided that individuals command the ability to detect committed actor-types with sufficient reliability at sufficiently low costs, commitments that express themselves in proximately disadvantageous CSR choices may indirectly contribute to the ultimate success of committed types.[14]

We are in principle content to let it rest with the insight that commitments to CSR can be sustainable under certain conditions.[15] Commit-

13 A parallel argument is made in (Güth and Kirchkamp 2021).

14 Dawkins observation that in case of personal actors long term survival is the greatest ultimate success carries over to the case of corporate actors as well; see (Dawkins 1995)

15 Whether the argument will work beyond mutually advantageous exchange and contract is open. But even if the viability of "doing well by doing good" were restricted to

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ments to execute (incomplete) contracts not only according to the letter but in spirit – i.e. genuine trustworthiness rather than reliability – will be sought after in corporate actors as potential transaction partners as it is sought in personal partners. Yet, since "hunger is not bread"[16], the fact that commitments would be desirable does not prove their feasibility. We believe that contrary to common sense corporate actors may be seen as commanding a comparative advantage over ordinary personal actors when it comes to having a conscience or commitments functionally equivalent to it.

C. Corporate constitutions as "corporate conscience"

The dual function model underlying the indirect evolutionary approach remains externalist in focusing on overt behavior rather than on internal processes that generate behavior.[17] We now illustrate by concrete examples how empirically the RCM framework may be specified.

I. Creating and signaling general corporate constitutional commitments

Trust in an agent's (trustee's) intrinsic motivation to pursue the aims ends or values of the trustor (principal) can be substituted by explicit contract. But if complete contracts specifying substantive outcomes are either impossible or undesirable, potential interaction partners may place value on internal *procedures* of a corporate actor that provide a functional equivalent of a personal actor's trustworthiness (e.g. in using her discretionary powers according to Rule of Law principles). A simple example can illustrate how this may be accomplished in principle (though real practices will be more complex).[18]

---

exchanges and contracts that are directly advantageous to all involved it would still be a highly relevant result.

16 See ((Bentham 1843) fallacy II).

17 In the standard approach there is no room for intrinsic motivation per se. Since only the choice behavior resulting from motivational processes matters is suffices to represent the choices made – or so the standard argument runs.

18 The existence of procedural rules in practice, of course, requires fundamental rule following behavior of individuals. On experiments concerning the empirical side of rule following see (Kimbrough and Vostroknutov 2016; 2018; Gächter u. a. 2025).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

II. Simple majority rule as commitment device[19]

Imagine a group of five individuals operating as a corporate actor under simple majority rule in some realm of choice. One individual is of age up to 1, one up to 2, one up to 3, one up to 4, and one up to 5. Once an actor becomes member he will be able to go for at least five periods. At the beginning of each period the oldest actor presently on the team, upon turning 5, exits from the community and is substituted by an individual of age 0. There is no exit other than the regular one after five periods and no new member is admitted unless a regular exit has taken place.

In our stylized example of procedural corporate governance, the age composition and the simple majority rule were fixed. The median voter is always aged less than three and has to go on for at least three periods of corporate choice making (including the one in which the decision is made). These aspects of the internal structure of the corporate actor may be fixed in externally enforceable contracts. But there is and must be residual discretionary power in collective decision making of the corporate actor. All individuals who have a voice in the decisions retain their powers of judgment and the freedom to vote for or against proposals, respectively, as seems fit to them. *Decision procedures rather than decisions are fixed.* The procedure sees to it that when exercising their discretionary power, the median and all younger voters must always consider at least two future periods of interaction. Therefore, the collective decisions of the corporate actor are not suffering from the same "end-game" effects as decisions of individual actors. Incentives to show opportunity taking behavior on the corporate level are in this sense bounded even if the individuals act fully opportunistically.

Individuals who join such a corporate actor know that the group as a whole will be committed at every point in time to adopt a perspective that goes beyond the short term. Personal or corporate actors who face such a corporate actor in external relations also understand that the actor is substantively unrestrained but will use its discretionary power in a way that reaches beyond the direct consequences for each single round of play taken separately even if the personal actors "within" the corporate actor are acting as opportunistically as in the outside world.

---

19 For the simple modeling idea originally due to Brennan see (Brennan and Kliemt 1994), for more general and sophisticated treatments of the folk theorem logic involved, see e.g. (Kandori 1992), (Salant 1991).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

In sum, we have both, discretionary powers and a restriction on the nature of the opportunity set that will induce less short-term oriented and in that operational sense more "responsible" behavior.

Artificial and simple as the example may be, it nevertheless illustrates how procedures of collective decision-making "within" a corporate actor can serve as a commitment device. Of course, the workings of the mechanism depend on being committed to playing by the rules. To the extent that people have reasons to believe in rule of law and can rely on the legal order they can also have good reason to rely on the external enforcement of procedural rules internal to the corporate actor.[20] They can *rely* on the fact that the collective actor will have the internal age structure described before, decides by simple majority, and has a potentially infinite life if it can sustain itself. Thereby the externally monitored internal corporate governance structure can induce external interaction partners of the corporation to expect corporate social responsibility in the sense of its bounded rather than unrestricted opportunism. This is a clear sense in which a functional equivalent of trustworthiness may be created by an appropriate constitutional decision architecture.

Internal commitments to certain procedures of corporate governance are not only relevant for external relations of the corporate actor. They are also relevant for its relations to its own members. Again, an extremely stylized simple example may serve as an illustration of the potential advantages of such forms of *procedurally embodied* corporate social responsibility.

III. Unanimous co-determination of decisions

Other than shareholder, so-called stakeholder approaches argue that managers and the companies they run should be treating all potential interaction partners of the company respectfully and fairly in an ethically non-discriminatory manner.[21] Whether this can plausibly hold for external

20 We admit that institutionalized constitutional commitments to "the political ideal of the rule of law" (Hayek 1955) ultimately depend on *informal practices* of individual rule following. See on the nature and significance of the political ideal of the rule of law (Kliemt 2024).

21 On Freeman's influential variant of such an approach see (Freeman 1984), (Freeman u. a. 2010).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

relations with suppliers and customers as stakeholder conceptions often insinuate seems doubtful.[22] With respect to internal relations with employees or shareholders such an ideal of non-discriminatory treatment may seem more plausible and can indeed be spelled out in terms of unanimous decision-making.[23]

The unanimity rule must in any event seem very attractive to adherents of stakeholder conceptions of corporate social responsibility since it implies that shareholder interests do not command "ethical" priority vis a vis stakeholder-interests. It expresses the stakeholder conceptions' core ethical value of equal moral concern (operationalized as equal veto power) for all stakeholder interests in concrete institutional rule terms.[24]

Though it is natural for an ethicist to subscribe to norms of equal ethical respect this subscription rests uneasily with the privileges of private property and ownership in WEIRDS (Western, Educated, Industrialized, Rich, Democratic Societies) which in all likelihood are part of the secret of the success of such societies (Henrich 2020; 2016). It also disregards the fact that contracts with other individuals on behalf of the corporation in the last resort derive their authority from being authorized by residual claimants or owners. If we are not willing to brush these rather obvious *normative* objections aside then following the precepts of a stakeholder approach, to the extent that it can be justified at all, must be justified indirectly: to adopt a stakeholder perspective in 'human resource management' must pay off over the long haul for shareholders of the corporation. Unless it is a variant of 'doing well by doing good' it is hard to see how it could be compatible with respecting property rights of shareholders under the rule of law.[25]

Despite this, as far as relations to employees are concerned a corporate culture in which the values of fairness and equal respect for each and

22 See on some of such doubts in a management context (Kliemt 2022)

23 Of course, the complementarity of formal and informal "rules" in corporate culture is acknowledged; see for instance succinctly on this the comments in (Engel and Güth 2006)

24 See on the ways and means of spelling this idea out (Brennan and Kliemt 2018; 2022; Kliemt 2021).

25 The exit option plays a crucial role here: if management uses the discretionary powers vested in it in ways that keep the share value up, those shareholders who disagree with its decisions get the best start when selling and using their own resources for their own aims outside the corporation. Likewise, employees can exit the employment nexus of a corporation; see for related issues (Buenstorf u. a. 2016).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

everybody are guiding principles of interaction may seem to have great promise. No wonder that senior management expresses support for such a corporate culture in speeches. However, potential employees may demand something more reliable than speeches in which principles of corporate social responsibility in dealing with stakeholder employees are endorsed merely as lofty ideals. To win (potential) employees over, the implementation of corporate governance procedures that *embody* equal respect may be necessary.[26] But as always, the devil is in the details, and we have a hunch that the devil might ultimately have the upper hand here.[27]

D. Final observations

To amount to anything real, "stakeholder-talk" must be translated into "stakeholder rules"[28]. Majority rule though giving an equal vote to everyone does not express equal respect sufficiently in procedural terms. As has been observed through the ages majorities may overrule minorities the same way as bosses in a hierarchy overrule their underlings. Taking stakeholder conceptions at their word only the strict unanimity rule could express guarantees of equal concern procedurally and in institutional terms.[29]

Nobody would suggest in earnest that all decisions in real world affairs should be made unanimously by all affected individuals. Only for certain fundamental decisions veto power may be procedurally granted to individuals of a certain group of stakeholders (or to representatives of homogenous stakeholder groups). Though in some circles of contractarian political philosophers and contractarian political economists, *conceivable*

26 Acquiring firm specific human capital can be conceived as analogous to showing trust in the interaction of Figure 1. Putting the employer in a role analogous to the B-role actor of Figure 1 the A-role employee has to be aware that his exit option may decrease in value and the employer's bargaining position may be strengthened (in particular the threat point of negotiations may alter in the course of time in ways unfavorable to the specializing employee).

27 See for an assault on the problem (Güth 2014).

28 Implementing deferred bonuses would be another obvious way to credibly signal a commitment of the corporation and its management to the longer-term

29 Though protection by substantive contractual claims and rights may be useful they can only regulate what is anticipated.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

unanimity is regarded as a useful tool of conceptual analysis a merely conceivable is not a real-world procedure (hunger is not bread).[30]

Due to the elegance of the revealed preference approach economists became almost oblivious of the fact that stenographical representations of the results of decision-making will work well only for understanding contexts in which the internal springs of action do not matter much.[31] This will be the case if observed behavior is – as in case of some markets – mostly determined by the *structural* conditions of *inter*action and the substantive or material payoffs that characterize them (typically in monetary terms).[32] In such cases of market competition actors who decide against what substantive payoffs require will be swiftly eliminated and economists can feel safe in ignoring behavior that is not substantively advantageous (one of the reasons why we modeled the evolutionary niche for CSR in so much detail).

It should be noted well, though, that the unitary actor model will suffice only under the special conditions described before. Only then can the psychology be taken out of economics without losing much. But if we consider more complex forms of behavior that at least allow for trade-offs between the short and the long term we cannot avoid shedding some light onto how behavior is generated.

Looking at internal processes of corporate actors we can treat their internal decision processes as equivalent to mental processes of personal actors who are cognitively and emotionally bounded (committed) to some

30 Even the most ardent adherents of unanimity try to cope with this by making its operation more indirect, for the seminal analysis see (Buchanan and Tullock 1962) and the still devastating criticism of classical contractarianism in "of the original contract" (Hume 1985).

31 Some of the most interesting work in making economic models of individual behavior more adeaquate make use of such techniques. To name but two examples one might think of (Bolton and Ockenfels 2000) as well as (Fehr and Schmidt 1999) which both build distributional and other regarding concerns into "individual utility function". We have no objection against devoting attention to motivational processes internal to the actor. We resent, however, insisting that human behavior must by all means be described as if maximizing some function. This can be justified only if one has – as in a context of evolutionary competition – good reason to assume that inconsistent choice making that cannot be so represented will be selected against strongly.

32 It is not by chance that the models in which economists tend to take the greatest pride "predict" certain outcomes of interaction mostly independently of specific assumptions on the behavioral level; e.g. evolutionary models as again in (Schumpeter 1959), (Alchian 1950), experimental markets (Smith 1962), or chess board models like in (Schelling 1971), (Hegselmann 2012).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

extent. Both, corporate and personal actors need not be utilitarian consequentialists. They might as well start from premises of interpersonal respect of a broadly speaking Kantian type. Whether such views are ethically reasonable or not is discussed in normative ethics. But it seems beyond reasonable debate that Corporate Social Responsibility can be a serious corporate governance strategy only if it can be institutionalized in operational terms.

First, it must be shown how CSR can be embodied in institutional rules of corporate governance.

Second, institutionalized CSR commitments must be shown to be sustainable.

To the extent that we can by means of corporate governance prepare at all for future decision situations that cannot yet be anticipated it should not be in terms of prescriptions of substantive choices.[33] Neither personal nor corporate actors will want to be operating according to a fully specified explicit strategy leaving no discretionary power to future manifestations of the actor. Commitments therefore must concern the procedures rather than the substance of choices. Why CSR should be an exception to this is as hard to see as why it should be impervious to the strictures of sustainability in competition. On the other hand, RCM can be used to make it entirely clear what it takes that CSR can be sustainable in market competition and how the implicit managerial promise to hold up share values can be compatible with CSR. In defense of shareholder value conceptions we should not forget that shareholders are also stakeholders. Strengthening the shareholders' exit option by keeping the share value up is expressive of *ethical* respect for the plural values of this particular class of stakeholders who can always invest their capital in pursuit of their particular aims, ends and values with some competitor.

33 The enumeration of several mutually exclusive sufficient action triggers can always be made collectively exhaustive by an "else-clause" specifying the corresponding action. But no person in her right mind would believe in her ability to anticipate "everything else" sufficiently to warrant a substantively fixed response, *no matter what*.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

References

Alchian, Armen A. 1950. "Uncertainty, Evolution, and Economic Theory". *Journal of Political Economy* Vol. 58: 211–21.

Bentham, Jeremy. 1843. *Anarchical Fallacies*. Herausgegeben von John Bowring. Bd. 2. The Works of Jeremy Bentham. William Tait. https://oll.libertyfund.org/titles/bentham-works-of-jeremy-bentham-11-vols.

Böhm, Franz. 1966. "Privatrechtsgesellschaft und Marktwirtschaft". *ORDO* 17: 75–151.

Böhm, Franz. 1989. "Rule of Law in a Market Economy". In *Germany's Social Market Economy: Origins and Evolution*, herausgegeben von Alan Peacock and Hans Willgerodt. Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20145-7_4.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition". *American Economic Review* 90: 166–93.

Brennan, Geoffrey, and Hartmut Kliemt. 1994. "Finite Lives and Social Institutions". *Kyklos* 47 (4): 551–71.

Brennan, Geoffrey, and Hartmut Kliemt. 2018. "Kantianism and Political Institutions". In *The Oxford Handbook of Public Choice, Volume 1 (Oxford Handbooks) (English Edition): Roger D. Congleton, Bernard N. Grofman, Stefan Voigt (eds.)*. Oxford Handbooks.

Brennan, Geoffrey, and Hartmut Kliemt. 2022. "Politics as Exchange?" *Homo Oeconomicus*, Online-Vorab-Publikation, August 17. https://doi.org/10.1007/s41412-022-00128-5.

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent*. University of Michigan Press.

Buenstorf, Guido, Christoph Engel, Sven Fischer, and Werner Gueth. 2016. "Non-compete clauses, employee effort and spin-off entrepreneurship: A laboratory experiment". *Research Policy* 45 (10): 2113–24. https://doi.org/10.1016/j.respol.2016.08.005.

Dawkins, Richard. 1995. *River out of Eden. A Darwinian View of Life*. Science Masters. Weidenfeld & Nicolson.

Engel, Christoph. 2010. "The behaviour of corporate actors: How much can we learn from the experimental literature?" *Journal of Institutional Economics* 6 (4): 445–75. Cambridge Core. https://doi.org/10.1017/S1744137410000135.

Engel, Christoph. 2018. "The Proper Scope of Behavioral Law and Economics (January 1, 2018)." MPI Collective Goods Discussion Paper No. 2018/2,. Bonn. http://dx.doi.org/10.2139/ssrn.3104475.

Engel, Christoph, and Werner Güth. 2006. "Corporate Culture as a Resource for Management: Comment". *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 162 (1): 97–100. JSTOR.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation". *Quarterly Journal of Economics* 114: 817–68.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Frank, Robert. 1987. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *The American Economic Review* 77/4: 593–604.

Freeman, R. Edward. 1984. *Strategic Management: A Stakeholder Approach*. Pitman.

Freeman, R. Edward, Jeffrey S. Harrison, Andrew C. Wicks, Bidhan L. Parmar, and Simone de Colle. 2010. *Stakeholder Theory: The State of the Art*. Cambridge University Press.

Gächter, Simon, Lucas Molleman, and Daniele Nosenzo. 2025. "Why people follow rules". *Nature Human Behaviour*, Online-Vorab-Publikation, Mai 26. https://doi.org/10.1038/s41562-025-02196-4.

Güth, Werner. 2014. "Endogenous community formation and collective provision – A procedurally fair mechanism". *Journal of Economic Behavior & Organization* 108 (Dezember): 389–95. https://doi.org/10.1016/j.jebo.2014.01.005.

Güth, Werner, and Oliver Kirchkamp. 2021. "Believing in Corporate Social Responsibility: An Indirect Evolutionary Analysis". *Journal of Institutional and Theoretical Economics*.

Güth, Werner, and Hartmut Kliemt. 1994. "Competition or Co-Operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes". *Metroeconomica* 45 (2): 155–87. https://doi.org/10.1111/j.1467-999X.1994.tb00018.x.

Güth, Werner, and Hartmut Kliemt. 2000. "Evolutionarily Stable Co-operative Commitments". *Theory and Decision* 49 (3): 197–222.

Güth, Werner, and Hartmut Kliemt. 2007. "The Rationality of Rational Fools.The Role of Commitments, Persons and Agents in Rational Choice Modeling". In *Rationality and Commitment*, herausgegeben von Fabienne Peter and Hans Bernhard Schmid. Oxford University Press.

Hayek, F.A. 1955. *The Political Ideal of the Rule of Law*. Fiftieth anniversary commemoration lectures. National bank of Egypt. https://books.google.de/books?id=5REeAAAAMAAJ.

Hazlitt, Henry. 1988. *Economics in One Lesson: The Shortest and Surest Way to Understand Basic Economics*. Crown Business.

Hegselmann, Rainer. 2012. "Thomas C. Schelling and the Computer: Some Notes on Schelling's Essay 'On Letting a Computer Help with the Work'". *Journal of Artificial Societies and Social Simulation* 15 (4): 9.

Henrich, Joseph. 2016. *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Henrich, Joseph. 2020. *The Weirdest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. 01 Edition. Allen Lane.

Hume, David. 1985. *Essays. Moral, Political and Literary*. Liberty Fund.

Kandori, Michihiro. 1992. "Repeated games played by overlapping generations of players". *The Review of Economic Studies* 59 (1): 81–92.

Kay, John. 2011. *Obliquity. Why our Goals are best achieved indirectly*. Profile Books.

Kimbrough, Erik O., and Alexander Vostroknutov. 2016. "Norms Make Preferences Social". *European Journal of Political Economy* 14 (3): 608–38.

Kimbrough, Erik O., and Alexander Vostroknutov. 2018. "A Portable Method of Eliciting Respect for Social Norms". *Economics Letters* 168: 147–50.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Kliemt, Hartmut. 2021. "The logical foundations of constitutional democracy between legal positivism and natural law theory". *Public Choice* 195 (3): 269–81. https://doi.org/10.1007/s11127-021-00888-9.

Kliemt, Hartmut. 2022. "What should moral managers do? The moral virtues of shareholder value conceptions of management". *The'Betrieb'as Corporate Actor*, 108–21.

Kliemt, Hartmut. 2024. "The nature and sicgnificance of the Rule of Law". *Kyklos* 77 (4): 1084–102. https://doi-org.ezproxy.uni-giessen.de/10.1111/kykl.12407.

Lahno, Bernd. 1995a. "Trust, Reputation, and Exit in Exchange Relationships". *Journal of Conflict Resolution* 39 (3): 495–510.

Lahno, Bernd. 1995b. *Versprechen: Überlegungen zu einer künstlichen Tugend*. Oldenbourg Wissenschaftsverlag.

Lahno, Bernd. 2002. *Der Begriff des Vertrauens*. Mentis.

Mayer, Colin. 2013. *Firm Commitment: Why the corporation is failing us and how to restore trust in it*. Oxford University Press.

Salant, David J. 1991. "A repeated game with finitely lived overlapping generations of players". *Games and Economic Behavior* 3 (2): 244–59. https://doi.org/10.1016/0899-8256(91)90025-A.

Schelling, Thomas C. 1971. "Dynamic models of segregation". *The Journal of Mathematical Sociology* 1 (2): 143–86. https://doi.org/10.1080/0022250X.1971.9989794.

Schumpeter, Joseph Alois. 1959. *The Theory of Economic Development*. Harvard University Press.

Smith, Vernon L. 1962. "An Experimental Study of Competitive Market Behavior". *Journal of Political Economy* 70:2 (April): 111–37.

Toulmin, St. 1961. *Foresight and understanding. An enquiry into the aims of science*. Hutchison.

Zahavi, Amotz, and Avishag Zahavi. 1997. *The Handicap Principle. A Missing Piece of Darwin's Puzzle*. Oxford University Press.0.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Bridging Minds and Practices
Integrating Cognitive Psychology and Practice Theory in Legal Reasoning

Anne van Aaken

Prelude For Engel-Angel

Not winged, nor robed in light,
but in the quiet brilliance
of listening without hurry.

He walks the corridors of the Max Planck House
where thought waits in shadow,
drawing it into the open
with the gentlest of questions.

He walks the Socratic path –
not to display what he knows,
but to draw forth what others can see.

He does not bind thought
with the iron of finality;
he lets it breathe,
turning the law over
like a stone in the river
to see what lies beneath.

Empirical evidence is his compass,
jurisprudence his map –
together they lead him
through the fog of opinion
toward the clearer shore.

He listens to data whispered in columns, charts and LLMs,
testing each claim
against the weight of evidence,
the light of reason.

Fairness is his north star,
and in the constellations of academia

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

he charts a course
that keeps even the smallest vessel
from drifting into darkness.

He glows in the patience
with which he hears another's thought
all the way to its quiet end,
and in the careful questions
that make it rise again,
stronger, surer, more itself.

And so, the name fits,
though not for flights above us,
but for walking beside us,
reminding us that
knowledge,
like justice,
is best pursued
with open hands, hearts, and minds.

A. Introduction

The work of Christoph Engel has spanned many issue-areas and working with Christoph Engel is always following a path of intellectual inquiry. A project we have collaborated upon led us from experiments with front-line humanitarian negotiators to practice theory in international law and international relations. I would like to explore this path further in this chapter, albeit in a more conceptual manner and without self-conducted empirical research – which is so dear to Christoph Engel.

Faute-de-mieux, the chapter remains exploratory concerning the intersection of cognitive psychology and practice theory to provide a more holistic understanding of legal interpretation, particularly in international law – the issue area of law which was the starting point of Christoph Engel in his intellectual journey.[1] Legal decision-making has long been one of

1 Christoph Engel, *Völkerrecht als Tatbestandsmerkmal deutscher Normen* (Duncker & Humblot 1989).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the interests of Christoph Engel.[2] While practice theory emphasizes the social and embodied nature of interpretive routines, cognitive psychology can offer insight into the mental processes that underlie those routines. The tools for legal reasoning used – interpretative methods – often determine the outcome of interpretation.[3] Research has concentrated on showing experimentally that adjudicators fall prey to biases in *substantive* decisions, focusing on experimental variations in substantive law or facts. They have not looked at the *second order rules of interpretation* and how their use may be influenced by cognitive biases. Yet those rules of interpretation have been the subject of practice theory as applied in adjudication. Interpretative methods are especially crucial if a treaty is incomplete or if there is no settled or coherent jurisprudence. These methods have been deemed to be an instrument of power:[4] they are not neutral, but they may be decisive for the outcome of a case. To paraphrase Emanuel Kant: "Substantive law without interpretational methods is blind; interpretational methods without substantive law are empty."

Commonly, the choice of interpretative methods is attributed *inter alia* to the legal culture.[5] Although this can be captured by practice theory, the cognitive dimension is missing – given that the choice of different tools may be provoked by different cognitive biases and heuristics which in turn are embedded in practice. The question is thus how can cognitive processes (like perception and heuristics) be understood not in isolation but as imbedded in practice and being shaped by and shaping legal interpretive practices. This chapter attempts to weave those strands together.

2 Christoph Engel and Keren Weinshall, 'Manna from Heaven for Judges: Judges' Reaction to a Quasi-Random Reduction in Caseload' (2020) 17 Journal of Empirical Legal Studies 722; Christoph Engel and Fritz Strack (eds), *The Impact of Court Procedure on the Psychology of Judicial Decision Making* (1 edn, Nomos 2007); Christoph Engel and Werner Gueth, 'Modeling a satisficing judge' (2018) 30 Rationality and Society 220; Engel Christoph, *Judicial decision-making a survey of the experimental evidence* (Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2022/6)) and more.

3 The interest of Christoph Engel on interpretative methods (textual interpretation) and Large Language Models is pursued in Richard McAdams and Christoph Engel, 'Asking GPT for the ordinary meaning of statutory terms' (2024) 2 Journal of Law, Technology & Policy 235.

4 Dieter Grimm, 'Methode als Machtfaktor' in Dieter Grimm (ed), *Recht und Staat der bürgerlichen Gesellschaft* (Suhrkamp 1987), at 347.

5 Pierre Legrand, 'European Legal Systems Are Not Converging' (1996) 45 International and Comparative Law Quarterly 52.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

How do legal actors' cognitive biases interact with field-specific interpretive tools? The chapter aims to develop a dual-layered account where the use of interpretational tools is not only a product of situated social practices (practice theory) but also constrained by mental architecture (cognitive psychology). Thus, I try to tease out the cognitive dimension of practice when adjudicators choose the tools of interpretation – which in turn influence decision-making. Combining these frameworks allows for a situated account of legal reasoning as embedded cognition that recognizes both embodied social dispositions and bounded rationality. It encourages empirical research combining experimental, ethnographic, doctrinal, and psychological approaches.

After cursorily depicting cognitive psychology and practice theory (II.), the chapter explores to apply a combination to legal interpretive methods as used in international law, necessarily remaining speculative but generating tentative hypotheses and using examples from international law (III.). The last part concludes (IV.).

B. Theoretical Foundations

Practice theory and cognitive psychology have many common points of integration yet have been methodologically apart and not much exploration has been done concerning their common insights.[6] Partially, this is a problem of translation and partially a problem of siloed thinking. In the following, I try to do some translation, in spite of that in every "translation" something gets lost.

I. Practice Theory

Practice theory is diverse, consciously and deliberately so.[7] Practice theory broadly refers to approaches in the social sciences that focus on the

6 Although practice theory and cognitive science are not so far away: Omar Lizardo, 'Pierre Bourdieu as Cognitive Sociologist' in Wayne H. Brekhus and Gabe Ignatow (eds), *The Oxford Handbook of Cognitive Sociology* (Oxford University Press 2019).

7 Theodore R. Schatzki, 'Introduction: Practice Theory' in Theodore R. Schatzki, Karin Knorr Cetina and Eike von Savigny (eds), *The Practice Turn in Contemporary Theory* (Routledge 2001); Emanuel Adler and Vincent Pouliot, 'International practices: introduction and framework' in Emanuel Adler and Vincent Pouliot (eds), *Interna-*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

routine, embodied actions and practices through which social life is produced and reproduced. Practice theory, as developed by scholars such as Pierre Bourdieu,[8] Andreas Reckwitz,[9] and Theodore Schatzki,[10] views social life as composed of routinized practices. Legal interpretation is seen not as the mechanical application of rules, but as a practice informed by shared norms, institutional settings, and professional socialization: "while an agent is reflecting upon what action will yield the most benefits or correspond to her normative commitments, she is doing so against a taken-for-granted background of habit that has already constrained her imaginable outcomes."[11] Bourdieu called these background beliefs, which are closely linked to schemes of perceptions and dispositions, the *habitus*. These dispositions are so deeply ingrained that they operate largely below the conscious surface.[12] Bourdieu's concept of *habitus* is central here: the embodied dispositions that guide legal actors in interpreting texts based on their legal training and experience.[13] Similarly, cognitive sociology explains behavior, roughly speaking, through cultural norms.[14]

Practice theory is called eclectic:[15] practices are both "individual (agential) and structural since, when 'disaggregated', practices are ultimately performed by individual social beings and thus they clearly are

tional Practices (Cambridge Studies in International Relations, Cambridge University Press 2011); Christian Bueger and Frank Gadinger, *International Practice Theory: New Perspectives* (2nd edn, Palgrave Macmillan 2018).

8 Pierre Bourdieu, *Practical Reason: On the Theory of Action* (Polity Press 1998).

9 Andreas Reckwitz, 'Toward a Theory of Social Practices:A Development in Culturalist Theorizing' (2002) 5 European Journal of Social Theory 243.

10 Schatzki, 'Introduction: Practice Theory', n. 7.

11 Ted Hopf, 'The logic of habit in International Relations' (2010) 16 European Journal of International Relations 539, 547.

12 Anne van Aaken and Moshe Hirsch, 'Introduction: International Legal Theory and the Cognitive Turn' in Anne van Aaken and Moshe Hirsch (eds), *International Legal Theory and the Cognitive Turn* (Oxford University Press 2025).

13 Mikael Rask Madsen and Salvatore Caserta, 'International Judicial Habitus: Pierre Bourdieu and the Cognitive Turn' in Anne van Aaken and Moshe Hirsch (eds), *International Legal Theory and the Cognitive Turn* (Oxford University Press 2025).

14 Seminal, Zerubavel 1999., especially 3–52, closing the gap between the unique solitary thinker and the cognitive-psychological view, searching for the universal foundations of human cognition by introducing the social mind.

15 Jérémie Cornut, 'Analytic Eclecticism in Practice: A Method for Combining International Relations Theories1' (2015) 16 International Studies Perspectives 50; Hopf, 'The logic of habit in International Relations', n. 11.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

what human agency is about."[16] The practice literature emphasizes the doings of actors and often refers to these doings as practices. Using practice theory moves beyond viewing legal interpretation as simply a cognitive or linguistic exercise. Instead, it's a socially situated, historically conditioned practice, but still agency matters for the choice of tools and this agency is also subject to cognitive bounds.

II. Cognitive Psychology

Cognitive psychology investigates how individuals perceive, process, and recall information and has been widely applied in behavioral law and economics,[17] including judicial decision-making. This research relies on the dual-process theory whereby System 1 (fast, intuitive thinking) is useful for familiar legal rules or routine decisions, but vulnerable to bias and System 2 (slow, deliberative thinking) is activated in complex, novel, or high-stakes legal interpretations. It is more accurate but resource-intensive. Legal interpretation can thus be influenced by cognitive phenomena such as heuristics and biases (such as anchoring, availability, confirmation, hindsight), but those can also be overridden. Numerous experiments with adjudicators[18] show that they also fall prey to biases and heuristics, in spite of that there are also counterexamples where professionalism can shield against those biases and heuristics within the domain of expertise of the adjudicator.[19] These cognitive patterns shape how legal texts are understood, especially under conditions of uncertainty and complexity. The choice of interpretative method may also be influenced by deliberate

16 Emanuel Adler and Vincent Pouliot, 'International practices' (2011) 3 International Theory 1, at 3.

17 Instead of many Eyal Zamir and Doron Teichman, *Behavioral Law and Economics* (Oxford University Press 2018).

18 Chris Guthrie, Jeffrey J. Rachlinski and Andrew J. Wistrich, 'Inside the Judicial Mind' (2001) 86 Cornell Law Review 777Chris Guthrie, Jeffrey J. Rachlinski and Andrew J. Wistrich, 'Blinking on the Bench: How Judges Decide Cases' (2007) 93 Cornell Law Review 1; Susan Franck and others, 'Inside the Arbitrator's Mind' (2017) 66 Emory Law Journal 1115.

19 Anne van Aaken and Roee Sarel, Framing Effects in Proportionality Analysis, Journal of Law & Empirical Analysis (2025) 2 Journal of Law & Empirical Analysis, 174. Yahli Shereshevsky and Tom Noah, 'Does Exposure to Preparatory Work Affect Treaty Interpretation? An Experimental Study on International Law Students and Experts' (2018) 28 European Journal of International Law 1287.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ignorance, that is, the "conscious choice not to seek or use knowledge (or information)".[20] Depending on which tool is chosen, information may be avoided, e.g. more so if textual analysis remains the main tool. Not using further interpretative tools avoids certain outcomes, e.g. if subsequent state practice is ignored.[21]

Cognitive psychology has been much less discussed with a view to how it matters for hermeneutics and legal reasoning, including the choice of interpretative tools.[22] Hermeneutics, understood as the theory of interpretation, investigates some questions that are also asked in the cognitive sciences. The nature of human understanding, the way that we gain and organize knowledge, the role played by language, and the relations between conscious and unconscious knowledge are all examples from the intersection of hermeneutics and the cognitive sciences.[23]

III. Integrating the Frameworks

The concept of *habitus* aligns well with schema theory in cognitive psychology.[24] Habitus can be defined as "the habitual, patterned ways of understanding, judging, and acting which arise from our particular position as members of one or several social fields, and from our particular trajectory in the social structure."[25] Schema theory is a branch of cognitive science concerned with how the brain structures knowledge. It represents a cognitive framework that helps individuals organize and interpret infor-

20 Ralph Hertwig and Christoph Engel, 'Homo Ignorans: Deliberately Choosing Not to Know' in Ralph Hertwig and Christoph Engel (eds), *Deliberate Ignorance: Choosing Not to Know* (MIT Press 2021), at 3.

21 See for details on investment law on (not) chosen interpretative tools, Aaken Anne van, 'Interpretational Methods as an Instrument of Control in International Investment Law' (2014) 108 Proceedings of the Annual Meeting (American Society of International Law) 196.

22 Anne van Aaken, 'The Cognitive Psychology of Rules of Interpretation in International Law' (2021) 115 AJIL Unbound 258.

23 Shaun Gallagher, 'Hermeneutics and the Cognitive Sciences' (2004) 11 Journal of Consciousness Studies 162.

24 Peter. J. Hampson and Peter. E. Morris, *Understanding Cognition* (Blackwell Publishers Inc. 1996).

25 Pierre Bourdieu, 'The Force of Law: Toward a Sociology of the Juridical Field' (1987) 38 Hastings Law Review 805, 811.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

mation based on their past experiences and knowledge.[26] Habitus relies on schemata. Individuals access schema to guide current understanding and action.[27] Legal actors rely on mental schemas – structured clusters of preexisting knowledge – to interpret laws.[28] For example, an arbitrator might have a schema for what constitutes "fair and equitable treatment" in international investment law, influencing how they read treaties or facts and also what kind of interpretative tools they use for interpretating "fair and equitable treatment".

Schemata in turn, are connected to framing since each acquired schema is a frame for a decision-problem. Framing is thus not to be viewed purely in isolation, as is mostly the case in behavioral economics, but is embedded. Framing as used in behavioral economics comes in two forms: A framing effect exists "when different ways of describing the same choice problem change the choices that people make, even though the underlying information and choice options remain essentially the same."[29] This phenomenon is called equivalence framing, and it is pervasive, e.g. when a choice is described as a loss rather than a gain, people decide differently.[30] Many experiments have explored these effects.[31] Framing plays a subtler role when a given problem description influences the decision maker's choice (mostly unconsciously) with regard to how to address a broader problem. This happens primarily by making certain aspects of an information set more salient, to the detriment of other aspects of the information set ('issue framing') and emerges from research in political

26 Robert Axelrod, 'Schema Theory: An Information Processing Model of Perception and Cognition' (1973) 67 American Political Science Review 1248.

27 Jeff Pankin, 'Schema theory and concept formation' (2013) Presentation at MIT, Fall; available at: https://webmitedu/pankin/www/Schema_Theory_and_Concept_Formationpdf .

28 Mikael Madsen and Salvatore Caserta, 'International Judicial Habitus: Pierre Bourdieu and the Cognitive Turn' in Anne van Aaken and Moshe Hirsch (eds), *International Legal Theory and the Cognitive Turn* (Oxford University Press 2024).

29 R. Cookson, 'Framing Effects in Public Goods Experiments' (2000) 3 Experimental Economics 55, 55. On different theories about framing, see Tore Ellingsen and others, 'Social framing effects: Preferences or beliefs?' (2012) 76 Games and Economic Behavior 117, 118. Framing effects similarly violate the rationalist axiom of 'descriptive invariance'.

30 Daniel Kahneman and Amos Tversky, 'Choices, Values, and Frames' (1984) 39 American Psychologist 341.

31 ibid.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

communication.[32] Judges and legal practitioners develop cognitive structures through legal education and professional experience that guide how they interpret legal texts. This applies also to the tools they use when interpreting texts; issue framing thus becomes important and is closely linked schemata, yet, also equivalence framing can play a role.

The encoding hypothesis of schema theory posits that schemas affect what information is the focus of our attention and what is not (e.g. how long we look at something), thus affecting the encoding of memory from short-term to long-term memory. Schemas guide how we interpret new information and may be quite powerful in their influence. Here, cognitively psychology would distinguish several underlying biases and heuristics which come to the fore also at the moment of adjudication: the availability bias,[33] anchoring[34] as well as the well-known confirmation bias, relevant to numerous international legal situations.[35] The latter bias refers to the tendency to search for, interpret, favour, and recall information in a way that confirms or supports one's prior beliefs or values.[36] Thus, people display this bias when they select information that supports their previous views, underestimating or ignoring contrary information, or when they interpret ambiguous evidence as supporting their existing attitudes. This mirrors the psychological concept of procedural memory and automaticity, where frequent practice leads to less conscious effort. Interpretive methods such as textualism or teleology can become routinized, reducing cognitive load.

---

32 James N. Druckmann, 'Political Preference Formation: Competition, Deliberation and the (Ir)relevance of Framing Effects' (2004) 98 American Political Science Review 671.

33 Amos Tversky and Daniel Kahneman, 'Availability: A Heuristic for Judging Frequency and Probability' (1973) 5 Cognitive Psychology 207.

34 Adrian Furnham and Hua Chu Boo, 'A literature review of the anchoring effect' (2011) 40 The Journal of Socio-Economics 35; Piotr Bystranowski and others, 'Anchoring effect in legal decision-making: A meta-analysis' (2021) 45 Law and Human Behavior 1.

35 See, e.g., Shiri Krebs, 'The Invisible Frames Affecting Wartime Investigations: Legal Epistemology, Metaphors, and Cognitive Biases' in Andrea Bianchi and Moshe Hirsch (eds), *International Law's Invisible Frames: Social Cognition and Knowledge Production in International Legal Processes* (Oxford University Press 2021), 132 et seq.

36 See, e.g., Raymond S. Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2 Review of General Psychology 175. Similarly Axelrod, 'Schema Theory: An Information Processing Model of Perception and Cognition', n. 26.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. Application to Legal Interpretive Methods in International Law

From a cognitive psychological perspective, biases and heuristics are both associated with dual-process theories. Ideally, legal interpreters always reason via System 2.[37] From a practice theory perspective, all legal interpretation is embedded in habitus.[38] In the following, the insights from both are combined to understand the use of interpretative tools. Rules of interpretation are meant to help the interpreters in overriding intuition. But they may be prone to biases and heuristics themselves. International law is especially interesting given its open texture and decentralized nature, its codified interpretive tool pluralism in the Vienna Convention on the Law of Treaties (VCLT[39]), and the pluralism on the international bench, shaped by different legal cultures. Thus, cognitive schemas, narratives, and biases can play a more pronounced role in interpretation. VCLT Articles 31 and 32 mandate the rules of treaty interpretation to be used, with ordinary meaning, context, and object and purpose being the starting point[40] - and the focus here. The VCLT does not prescribe how exactly these tools are used.

I. Textualism

Textualism as an interpretative tool focuses on the plain, ordinary meaning of the words used in the legal text. It is highly routinized and symbolically powerful in some legal cultures, especially the common law.[41] It is also a means to remain faithful to the originators of the text, yielding legitimacy and shielding the interpreter from judicial activism accusation. Literal interpretation is appealing in international law due to the lack of a central authoritative body, the perceived clarity and the desire for ob-

37 Christoph Engel, 'Uncertain Judges' (2020) 176 Journal of Institutional and Theoretical Economics 44.

38 Bourdieu, 'The Force of Law: Toward a Sociology of the Juridical Field', n. 25.

39 Vienna Convention on the Law of Treaties, May 23, 1969, 1155 U.N.T.S. 331.

40 As stated early on in Sir Jennings and Arthur Watts (eds), *Oppenheim´s International Law Vol. I, Part 2 to 4 (Peace)* (Longman 1992/1905), 1267: "The finding whether a treaty is clear or not is not the starting point but the result of the process of interpretation."

41 The most ardent representative of textualism is arguably Antonin Scalia, *A Matter of Interpretation: Federal Courts and the Law* (Princeton University Press 1997).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

jectivity. Textualism, compared to other interpretative tools, also shows cognitive economy: interpreting text literally is cognitively efficient – it relies often on System 1 (fast, automatic thinking) and avoids the effort of contextual or teleological analysis. Semantic priming[42] would suggest that adjudicators activate mental dictionaries of word meanings; these in turn are shaped by culture, language exposure, and profession – in other words – by schemas. Furthermore, the actual words of the treaty text anchor reasoning. Even if a broader interpretation might make sense, adjudicators may tend to stay close to the literal wording due to this bias and more so if their background derives from a legal culture where the dominant schema used for interpretation is the literal tool of interpretations. Additionally, the status quo bias may kick in here, given that any deviation from the wording needs additional argumentation – the literal meaning is the default. Status Quo bias and anchoring bias may also explain the use of precedent,[43] even if not an official tool of interpretation in international law but widely found in international adjudication.

II. Systematic Interpretation in International Law

This interpretive tool emphasizes systemic coherence. It aligns with cognitive tendencies like narrative coherence and consistency bias, and is cultivated through transnational professional networks.[44] Context is closely connected to ordinary meaning. Often, the main reason for looking to the context is to confirm an ordinary meaning. Yet, context can include very different factors ranging from surrounding words, heading of articles, and punctuation, to more remote elements such as other provisions

42 Semantic priming is a psychological phenomenon where the processing of a stimulus (a *target*) is facilitated by prior exposure to a semantically related stimulus (a *prime*). It is taken as evidence that words with related meanings are stored closely in semantic memory, and activating one concept spreads activation to related ones. See Timothy P. McNamara, *Semantic Priming: Perspectives from Memory and Word Recognition* (Psychology Press 2005).

43 Gilbert Guillaume, 'The Use of Precedent by International Judges and Arbitrators' (2011) 2 Journal of International Dispute Settlement 5 who attributes the use of precedent to purposes of legal certainty and for fear that decisions might be challenged before higher instances.

44 Nora Stappert, 'Practice theory and change in international law: theorizing the development of legal meaning through the interpretive practices of international criminal courts' (2020) 12 International Theory 33.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

in the same text, similar provisions in other treaties, or even the object and purpose.[45] There is thus a wide range of context elements open to choose from. Although the context used can be influenced by schemata, two cognitive biases may also kick in here. First, the availability bias is a mental shortcut that relies on immediate examples or clues that come to a given person's mind when evaluating a specific topic, concept, method, or decision. A specific context may be made salient by counsel, by previous decisions, or in deliberations, leading to the neglect of other elements. Second, the cognitive bias of motivated reasoning is well known.[46] Motivated reasoning refers to people's tendencies to conform their assessments of information, including from logical arguments or their own sense impressions, to some end or goal extrinsic to judgment accuracy; in other words, motivated reasoning is the "tendency to find arguments in favor of conclusions we want to believe to be stronger than arguments for conclusions we do not want to believe."[47] This is not to say that motivated reasoning necessarily occurs when choosing the context, but it suggests that when using the wide range of possible context elements, the schemas of the adjudicator may play a role when choosing the context, accompanied by motivated reasoning. Ambiguous terms in international law, of which there are many (e.g., "reasonable," "due process") are then interpreted through context which in turn may be influenced by individual cognition and background schema or habitus. Context may also be invoked selectively and then live on against all better knowledge, as the argument that foreign investors need special protection given that they cannot vote, as in investment arbitration where the European Convention of Human Rights was invoked selectively.[48]

45 Richard K. Gardiner, *Treaty Interpretation* (Oxford University Press 2008), 178.

46 Dan M. Kahan, 'Ideology, motivated reasoning, and cognitive reflection' (2013) 8 Judgment and Decision Making 407.

47 Z. Kunda, 'The case for motivated reasoning' (1990) 108 Psychol Bull 480, 480 et seq.

48 Anne van Aaken and Jan-Philip Elm, 'Framing in and through International Law' in Andrea Bianchi and Moshe Hirsch (eds), *International law's invisible frames – Social cognition and knowledge production in international legal processes* (Oxford University Press 2022).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

III. Teleological Interpretation

Another VCLT requirement is to assess a treaty in the light of its object and purpose, linked to the principle of effectiveness. Finding the object and the purpose of a treaty is not an easy task. The object and purpose can be found in the preamble or sometimes even the title of the treaty.[49] But often it seems that it is found by the courts by intuition without transparent explanation.[50] Several questions therefore arise for an interpreter: what is the meaning of the object and purpose, how are they to be identified, what is the relationship between the object and the purpose, how can multiple purposes of a treaty be balanced. Since the object and purpose rule requires instrumental rationality (the respective interpretation should effectively achieve the object and purpose), the interpreter is asked to consider real world consequences. Yet, in spite of this, adjudicators often show narrative thinking – purposive interpretation taps into people's deep cognitive preference for coherent stories and goal-directed reasoning and draws on their intuition.

Teleological interpretation is more flexible but can be cognitively demanding, and is often tied to particular legal cultures or issue areas. Teleological reasoning is especially prevalent in specific fields with highly specified expertise, encouraged by shared field expectations (e.g., coherence of legal systems). For example, interpretation the European Court of Human Rights often involves purposive reasoning grounded in human rights values. The Court emphasizes that the Convention is a "living instrument" and must be interpreted in light of present-day conditions and human dignity – a teleological, evolving understanding of rights. Madsen documents how ECtHR judges' interpretive choices reflect both institutional habitus and cognitive framing effects, including moral reasoning and empathy-driven interpretation.[51] Similarly, in international investment law, oftentimes the object and purpose of a treaty is used rather blindly based on early precedent as being solely investor protection. Given

49 David S. Jonas and Thomas N. Saunders, 'The Object and Purpose of a Treaty: Three Interpretive Methods,' (2010) 43 Vanderbilt Journal of Transnational Law 565.

50 Isabelle Buffard and Karl Zemanek, 'The "Object and Purpose" of a Treaty: An Enigma?' (1998) 3 Austrian Review of International & European Law 311.

51 Mikael Madsen, 'The Protracted Institutionalization of the Strasbourg Court: From Legal Diplomacy to Integrationist Jurisprudence' in Jonas Christoffersen and Mikael Madsen (eds), *The European Court of Human Rights between Law and Politics* (Oxford University Press 2011).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

the close-nit network of investment arbitrators[52] with a similar background, it comes to no surprise that habitus can explain the repeated use of an object and purpose of investment treaty against all political science and economic insights.[53] The framing of investment law as a protective device for foreign investors influences the decisions of the tribunals sometimes crucially. The vaguer the investment treaty is, the more important a teleological interpretation becomes. In the view of some tribunals, those treaties are instruments for the maximization of investor protection;[54] accordingly, uncertainties concerning ambiguous treaty provisions should be resolved in favor of foreign investors. This reasoning is reproduced through formal education and professional training across transnational legal networks.[55] Given the flexibility in purpose finding, the confirmation bias increases the risk that interpreters see the purpose they *want* to find, based on their habitus. Yet, it also can be explained by the availability of the argument and the anchoring of the arbitrators.

C. Conclusion and Outlook

This chapter aims at moving beyond dichotomy and argues that in order to understand the use of interpretative tools, we need to move past the false opposition between internalist (cognitive) and externalist (sociological) accounts. Legal interpretation is both socially and cognitively situated. It pleads for a contextual account of cognition, recognizing that cognitive processes are context-sensitive and practice-conditioned. Bounded rationality is situated and legal reasoning is not purely logical nor purely social – it is shaped by embodied dispositions and cognitive constraints. Integrating practice theory and cognitive psychology enables

52 Sergio Puig, 'Social Capital in the Arbitration Market' (2014) 25 European Journal of International Law 387.

53 Anne van Aaken and Tobias Lehmann, 'Sustainable Development and International Investment Law: An Harmonious View from Economics' in Roberto Echandi and Pierre Sauvé (eds), *Prospects in International Investment Law and Policy* (Cambridge University Press 2013).

54 E.g., *Siemens A.G. v. Argentine Rep.*, ICSID Case No. ARB/02/8, Decision on Jurisdiction, Aug. 3, 2004, para. 81.

55 Yves Dezalay and Bryant G. Garth, *Dealing in Virtue. International Commercial Arbitration and the Construction of a Transnational Legal Order* (Chicago University Press 1998).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

a more nuanced understanding of how legal meaning is produced, reproduced, and sometimes contested. This interdisciplinary perspective can improve both, scholarly analysis and institutional design in international legal interpretation. Methodological implications are manifold: It would be desirable to combine qualitative methods (e.g., ethnography, interviews) with experimental or observational data. Furthermore, legal judgments can be studied for patterned interpretive biases and practical reasoning strategies. Yet, many open questions arise – can training that modifies interpretive practice also affect cognitive biases (e.g., reduce anchoring) and do judges from different legal cultures exhibit different interpretive heuristics based on their professional socialization in international law? Maybe Christoph Engel or some of his disciples inquire those questions further.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Towards an Integrative Legal Research

Eyal Zamir

A. Introduction

In the past decades, interdisciplinary legal studies have flourished in the United States and in several other systems, such as Israel. More recently, thanks in large part to the pioneering work and influence of Christoph Engel, this trend is also gradually taking root in Europe. Economic, behavioral, philosophical and empirical legal analyses, to name just a few examples, make key contributions to legal theory. However, a disturbing byproduct of this development in systems where this interdisciplinary revolution has succeeded, is the fragmentation of legal scholarship. Many legal scholars write exclusively or primarily for others within their own school of thought, such as legal economists or legal empiricists. In addition to the increasing fragmentation within legal academia, the interdisciplinary legal scholarship is often inaccessible to legal practitioners – legislators, judges, or attorneys.

This essay argues that legal scholars should not suffice with engaging in conventional genres of academic legal research – such as doctrinal, comparative, historical, or interdisciplinary studies of various kinds – but rather strive to integrate these different genres, as well. The essay characterizes integrative legal research, describes the need for it, examines its prospects and perils, and maintains that it has the potential to make a substantial contribution to the development of law and even to breakthroughs in legal scholarship.

Integrative legal research acknowledges the contributions of the various research methods and perspectives it incorporates. However, the claim is that, much like in the parable of the *blind men and the elephant*, each methodology captures at most only part of the whole. According to this ancient story, which originates in East Asia, a group of blind individuals attempts to understand what an elephant is by each touching a different part of its body. The one who feels the elephant's leg believes it resembles a tree, another who touches the tusk thinks it is like a spear, the one who explores the trunk assumes it is akin to a snake, and so on. Inte-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

grative legal research, as the term implies, strives to integrate the different perspectives. It aligns with intellectual curiosity about diverse viewpoints, openness to new ideas, and a commitment to normative and theoretical pluralism.

B. Interdisciplinary Research and the Fragmentation of Legal Academy

Legal scholars have always differed in their areas of interest, the extent to which they engaged with methods such as comparative law or legal history, and the level of theoretical analysis in their work. However, these differences did not, until a few decades ago, prevent scholars from understanding one another, appreciating each other's work, and engaging in intellectual dialogue. In recent decades, however, this situation has changed due to the fracturing of legal scholarship, at least in the United States. The primary driver of this shift has been the significant expansion of interdisciplinary legal research. Therefore, it is useful to clarify at the outset the meaning – or meanings – of the term *interdisciplinary research*.

Schematically, one can distinguish between three levels of interaction between fields of research and knowledge (in addition to the absence of any interaction). The lowest level of interdisciplinarity, sometimes referred to as *multidisciplinarity*, involves cases where a particular problem has multiple facets. To address it more effectively, researchers from different disciplines collaborate, with each analyzing the problem using the concepts and methods of their respective fields. The findings from these separate studies are then combined into a final product consisting of a series of chapters, each addressing the problem from a distinct perspective.[1]

The next level, which I will call *weak interdisciplinarity* (and which is sometimes referred to as *cross-disciplinarity*), involves interdisciplinary research that draws on insights, perspectives, and research methods from one field to enrich research in another field. A common metaphor used in this context is that of "crossing boundaries" between the territories of different disciplines.[2]

1 David R. Holmes, *Confronting Complex Public Health Problems: The Development of Interdisciplinary Research*, 2 J. Pub. Health Pol'y 361, 362–64 (1981); Julie Thompson Klein, Interdisciplinarity: History, Theory, and Practice 56–63 (1990).

2 Joseph J. Kockelmans, *Why Interdisciplinarity?*, in Interdisciplinarity and Higher Education 123, 127–28 (Joseph J. Kockelmans ed., 1979).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Finally, *strong interdisciplinarity* (or simply *interdisciplinarity* in the narrow sense) does not merely borrow ideas, concepts, and methods developed in one discipline to apply them in another. Instead, it synthesizes and integrates two or more disciplines to create a new field of knowledge. A fitting metaphor for this type of interdisciplinarity is two rivers merging into one.[3] Examples of such interdisciplinary fields include biochemistry, criminology, and gender studies.

The boundary between the last two levels of interdisciplinarity is not sharp, and in truth, there exists a spectrum – both analytically and, at times, developmentally – ranging from single-discipline research to multidisciplinarity, weak interdisciplinarity, and strong interdisciplinarity. Moreover, interactions between different fields of knowledge are only part of a more complex picture of the processes occurring in academic research and thought, such as the fragmentation of established fields into increasingly independent subfields.

In our context, legal scholars, influenced by the American legal realism movement, have found that legal phenomena can be better understood by employing research methods and analytical tools borrowed from other academic disciplines. Familiarity with how questions of interest to legal scholars are discussed and conceptualized in other fields generates fresh insights and enables a critical examination of prevailing assumptions in the legal domain.[4]

A very prominent interdisciplinary approach is law and economics. However, many other interdisciplinary fields have also flourished, including *law and literature*, *law and society*, *law and culture*, *legal philosophy*, *sociology of law, feminist legal analysis, law and psychology, and law and political science*. Additionally, in recent years, empirical (including experimental) studies in law have gained increasing prominence. These method-

3 Arabella Lyon, *Interdisciplinarity: Giving Up Territory*, 54 College English 681 (1992); Judith A. Allen & Sally L. Kitch, *Disciplined by Disciplines? The Need for an Interdisciplinary Research Mission in Women's Studies*, 24 Feminist Stud. 275, 276–78 (1998); Robert L. Scott, *Personal and Institutional Problems Encountered in Being Interdisciplinary*, in Interdisciplinarity and Higher Education, *supra* note 2, at 306, 307–08.

4 Douglas W. Vick, *Interdisciplinarity and the Discipline of Law,* 31 J. Law & Society 163, 181–82 (2004).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ologies are used to address distinctly legal questions, as well as to explore the complex interrelations between legal and non-legal phenomena.[5]

Although there are significant differences among these interdisciplinary approaches, they typically involve the use of terminology, methods, and research paradigms derived from another field of knowledge (such as economics, sociology, or philosophy). In other words, contemporary writing in the various *law and…* genres mostly reflects *weak interdisciplinarity* – applying insights, research methods, and arguments developed in other disciplines to enrich the understanding and evaluation of legal norms, institutions, and phenomena – rather than *strong interdisciplinarity*, which entails the integration of two disciplines to create a new one. That said, some approaches come closer to strong interdisciplinarity. For example, in parts of economic analysis of law, conventions have gradually developed that differ not only from traditional legal scholarship but also from standard economic analysis.

One way or another, because the perspectives, research methods, and concepts vary significantly from one interdisciplinary approach to another, legal academia has almost become a modern Tower of Babel, where researchers do not engage in dialogue simply because they speak different languages.[6] Frequently, engagement in one form of interdisciplinary legal research is accompanied by disregard of other forms of interdisciplinary work.[7] The only commonality among scholars from different interdisciplinary approaches sometimes seems to be their disregard for doctrinal legal writing – scholarship that engages with the law from within.

Under these circumstances, many legal scholars write exclusively or primarily for others within their own school of thought. This is reflected in the specialized jargon they use, in differing conventions regarding how arguments are presented, and in the venues in which they publish. Con-

5 Mathias M. Siems, *The Taxonomy of Interdisciplinary Legal Research: Finding the Way out of the Desert*, 7(1) J. Commonwealth L. & Legal Educ. 5 (2009).

6 Sanford Levinson & J.M. Balkin, *Law, Music, and other Performing Arts*, 139 U. Pa. L. Rev. 1597, 1653 (1991); Reinhard Zimmermann, *Law Reviews: A Foray through a Strange World*, 47 Emory L.J. 659, 688 (1998); Jerry Mashaw, *Deconstructing Debate, Reconstructing Law*, 87 Cornell L. Rev. 682, 682 (2002).

7 *See, e.g.,* Eyal Zamir & Yuval Farkash, *Standard Form Contracts: Empirical Studies, Normative Implications, and the Fragmentation of Legal Scholarship: Comments on Florencia Marotta-Wurgler's Studies*, 12 Jrslm Rev. Legal Stud. 137, 167–69 (2015) (criticizing pro-consumer scholars for disregarding rigorous empirical studies in law and economics, even when those studies support their arguments).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

comitantly, much of legal academia, at least in the United States, neither engages with the materials used by legal practitioners nor aspires to influence legal practice, whether in legislation, adjudication, or legal counselling.[8]

Sometimes, scholars who analyze legal phenomena from an external perspective show little interest in the complexity of legal texts. For example, economic analysis of law is often highly sophisticated in terms of its economic tools but overly simplistic in its understanding of the legal doctrines it analyzes, their practical significance, and their interconnections with other legal doctrines.[9]

What started as a commendable effort to expand the boundaries of legal analysis has, in some instances, resulted in a form of narrow-mindedness, as some scholars affiliated with certain schools of thought tend to overlook legal doctrine, doctrinal research, and legal studies employing other methodologies. What started as a provocative and stimulating challenge to old canons and truths has, in some cases, become rigid adherence to new canons and "truths." This phenomenon risks impoverishing academic inquiry rather than enriching it. Furthermore, even within distinct schools of thought, some scholars express a sense of diminishing returns or even intellectual stagnation.[10]

Is it possible to benefit from the immense contributions of interdisciplinary analysis to legal scholarship without paying the price of the fragmentation described above? In the next section, I argue that it is possible – although it requires effort and is not without risks.

---

8 Harry T. Edwards, *The Growing Disjunction Between Legal Education and the Legal Profession*, 91 Mich. L. Rev. 34 (1992). This article was followed by seventeen responses and a counter-response, published in 91 Mich. L. Rev. 1921–2219 (1993). *See also* Judith S. Kaye, *One Judges' View of Academic Law Review Writing*, 39 J. Legal Edu. 313 (1989); Michael D. McClintock, *The Declining Use of Legal Scholarship by Courts: An Empirical Study*, 51 Okla. L. Rev. 659, 682–95 (1998); Zimmermann, *supra* note 6, at 679–81.

9 Pierre Schlag, *"Le Hors De Texte, Ce Moi": The Politics of Form and the Deconstruction of Deconstruction*, 11 Cardozo L. Rev. 1631, 1651 (1990); Gareth Davies, *The Relationship between Empirical Legal Studies and Doctrinal Legal Research*, 13 Erasmus L. Rev. 3, 8–12 (2020).

10 *See, e.g.,* Eric Posner, *Economic Analysis of Contract Law After Three Decades: Success or Failure?*, 112 Yale L.J. 829 (2003)

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

C. The Promise of Integrative Legal Research

I. The Challenge

Each research method and perspective used to examine legal issues or the law itself has the potential to enrich legal scholarship. However, each method achieves, at most, a limited objective. For instance, standard economic analysis may indicate the most efficient legal rule, given certain assumptions about market structure and human behavior. However, to determine whether these assumptions about market structure are correct, empirical research is required. To better understand human motives and decisions, psychological studies are necessary. Economic research may point to the most efficient legal rule, but an ethical-normative discussion is needed to address the question of how important maximizing overall welfare is and how it should be balanced against other considerations, such as distributive justice. Insights from political philosophy are essential for discussing whether promoting efficiency or redistributing welfare are legitimate roles of the state, while considerations from political science help assess the feasibility of achieving these goals through different institutions. To draw conclusions from theoretical analysis regarding the formulation and implementation of legal norms, one must understand the interplay between the legislature and the courts, and how legal norms are interpreted and evolve over time. Understanding the law in practice requires examining not only the law on the books and the law in action, but also the history of legal institutions and norms, and their interactions with non-legal institutions and norms. Fully grasping the significance of judicial decisions necessitates using critical tools that reveal different layers of judicial reasoning and possible gaps between judicial reasoning and the court's actual ruling. Almost every legal issue can benefit from insights gained from the experiences of other legal systems. Moreover, every legal norm should be assessed in terms of its impact on various sectors in society.

Without diminishing the contributions of different research methods, integrating diverse bodies of knowledge, perspectives, and methodologies is crucial for advancing both theoretical and applied legal scholarship. Such integration is particularly vital if the law is perceived as containing inherent tensions between power, interests, and politics on the one hand, and values, justifications, and social objectives on the other;

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

between science and craft; between tradition and change; and between experience and logic. This is the challenge of integrative legal research.

II. Different Types of Integration

While many interdisciplinary legal studies have, over the years, developed well-defined research paradigms, the common denominator among studies that seek to integrate different fields of knowledge is less clear. One type of integration involves incorporating insights from one or more interdisciplinary perspectives into traditional doctrinal legal analysis. Sometimes, such references serve merely as embellishments. In more successful cases, serious engagement with these insights significantly enriches the discussion.[11] Integrated legal research rejects viewing law as an isolated discipline, but it does not dismiss internal legal analysis. Doctrinal tools like statutory interpretation, precedent analysis, and analogical reasoning remain essential.[12]

Integrating interdisciplinary insights into traditional (doctrinal) legal analysis enhances their practical usefulness for legal practitioners. In my view, legal academia has a social role that goes beyond producing knowledge for its own sake – it should strive to improve and advance society by influencing legislation, judicial rulings, and other legal activities, as well as by training future legal practitioners.[13] Unlike some academic fields (such as theoretical physics or metaethics), but similar to others (such as

11 On this type of integration, see, e.g., Matyas Bodig, *Legal Doctrinal Scholarship and Interdisciplinary Engagement*, 8 Erasmus L. Rev. 43 (2015). Two examples of such successful integrations are: Richard Craswell, *Against Fuller and Perdue*, 67 U. Chi. L. Rev. 99 (2000) (criticizing the conventional, tripartite classification of interests protected by remedies for breach of contract from the perspectives of economic analysis, contractarian philosophical theories, retributive theories, distributive justice, and corrective justice theories); Melvin A. Eisenberg, *Mistake in Contract Law*, 91 Cal. L. Rev. 1573 (2003) (combining economic efficiency, deontological morality and institutional considerations associated with the functioning of the courts to reassess the doctrine of mistake).

12 Richard A. Posner, *The Decline of Law as an Autonomous Discipline 1962–1987*, 100 Harv. L. Rev. 761, 777 (1987).

13 *See also* Edward L. Rubin, *The Practice and Discourse of Legal Scholarship*, 86 Mich. L. Rev. 1835, 1881 (1988); Edward L. Rubin, *What Does Prescriptive Legal Scholarship Say and Who is Listening to It: A Response to Professor Meir Dan-Cohen,* 63 U. Colo. L. Rev. 731 (1992).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

medicine and social work), law is also an applied field in which academia should play a role in addressing real-world challenges.

It follows that legal scholarship should not regard legal policymakers, judges or lawyers merely as research subjects, but also as the audience and interlocutors of legal academia.[14] This perspective also suggests that, in addition to being members of the academic community, legal scholars can and should engage actively with the broader legal profession.[15] Finally, it implies that at least some academic legal research should address issues that concern policymakers and practitioners, and should be written in a language that is accessible to the latter.

Moreover, even if legal research is viewed primarily as a pursuit of knowledge for its own sake, overlooking the nuances of legal doctrine undermines the credibility of interdisciplinary analyses. These nuances are best uncovered through traditional legal analysis. Preference should thus be given to integrative legal research that seriously engages with doctrinal complexity. Similarly, legal theory can meaningfully inform empirical legal studies.[16]

A second type of integration, which is closer to conventional interdisciplinary writing, involves moving from a *dual-discipline* approach to a *tri-discipline* or even *quad-discipline* model, and beyond. A prominent example from recent decades is the enrichment of economic analysis of law with insights from the behavioral sciences. What initially appeared to be an antithesis to the mainstream law and economics movement – quickly became an integral part of the broader economic analysis of law, now known as *Behavioral Law and Economics*. The thesis and antithesis merged into a synthesis, and behavioral law and economics has secured a central role in legal-economic analysis.[17] A comparable integration of economics, philosophy, and law can highlight the problematic normative

14 For the opposite view, see Meir Dan-Cohen, *Listeners and Eavesdroppers: Substantive Legal Theory and Its Audience*, 63 U. Colo. L. Rev. 569 (1992).

15 On the challenges involved in fulfilling these two roles at the same time, see Theunis Robert Roux, *The Incorporation Problem in Interdisciplinary Legal Research – Some Conceptual Issues and a Practical Illustration*, 8 Erasmus L. Rev. 55, 59–61 (2015).

16 Hanoch Dagan, Roy Kreitner & Tamar Kricheli-Katz, *Legal Theory for Legal Empiricists*, 43 Law & Soc. Inquiry 292 (2018).

17 Eyal Zamir & Doron Teichman, Behavioral Law and Economics 141–45 (2018); Doron Teichman, Avishalom Tor & Eyal Zamir, *If You Can't Beat Them, Join Them: Richard Posner and Behavioral Law and Economics*, 31(3) History Of Economic Ideas 67 (2023).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

assumptions underlying standard economic analysis and address them by developing economic models that impose deontological constraints on social welfare maximization.[18]

But even research that does not seek to transform standard interdisciplinary perspectives can still be considered integrative if it examines legal issues from multiple angles simultaneously. This constitutes a third type of integration. For example, a study analyzing a legal issue from sociological, ethical-normative, historical, and comparative perspectives qualifies as integrative legal research, even if it does not aim to alter or improve each of these perspectives individually. Examining a legal issue through multiple lenses influences each of those perspectives. This type of analysis can strengthen certain normative arguments (when different perspectives lead to similar conclusions),[19] while weakening others (when they produce conflicting conclusions).

III. Contribution to Other Disciplines

Beyond enriching legal research, integrative legal research can make a significant contribution to other, non-legal disciplines. The inherent connection between law and practical life, law's pervasive role in all aspects of society, the responsibility that comes with the coercive power of law, and the prominent place of interdisciplinary research in legal academia – all enable interdisciplinary legal scholars to identify both the methodological and substantive limitations of adjacent fields. For example, legal scholars – including those specializing in law and economics – may be particularly attuned to the tensions between economic efficiency and distributive justice, as well as between economic efficiency and fundamental human rights. Likewise, legal scholars engaged in economic analysis are often more aware than their economist counterparts of the central role that legal systems play in economic activity. Thus, integrative legal re-

18 Eyal Zamir & Barak Medina, Law, Economics, and Morality (2010). The same is true of a possible integration of law, psychology, and normative ethics. *See* Eyal Zamir, Law, Psychology, and Morality: The Role of Loss Aversion (2015).

19 *See, e.g.,* Eyal Zamir, *The Inverted Hierarchy of Contract Interpretation and Supplementation,* 97 Colum. L. Rev. 1710 (1997); Eyal Zamir, *The Missing Interest: Restoration of the Contractual Equivalence*, 93 Va. L. Rev. 59 (2007).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

search is not only about adopting insights and analytical methods from other disciplines but also about critically examining and refining them.[20]

Similarly, legal academia provides a space where scholars engaged in one interdisciplinary approach (e.g., law and economics) may engage in dialogue and critique the assumptions and methodologies of another interdisciplinary approach (e.g., law and philosophy).[21] It is therefore unsurprising, for example, that legal scholars were among the pioneers in integrating insights from cognitive psychology into economic analysis.[22]

IV. Methodology and Ideology

Integrative legal research is consistent with normative pluralism – the view that moral and legal questions benefit from multiple, even conflicting, normative theories, such as consequentialism, deontology, or differing conceptions of well-being. Yet even normative monists can adopt integrative approaches, since many legal perspectives are not tied to specific normative positions. For example, a scholar focused solely on maximizing aggregate preference satisfaction can still gain valuable insights from fields like cognitive psychology, comparative law, and legal history.

It is also worth noting the typical association between research methodologies and ideological worldviews. For example, many scholars engaged in law and economics tend to hold conservative views on economic and social policy. They oppose state intervention in private life, resist redistributive measures by the state, and are decidedly critical of paternalism. Integrative legal research can, among other things, separate economic methodology from the normative assumptions commonly associated with it – and the same applies, where relevant, to other interdisciplinary perspectives. Economic analytical tools can be used to assess the effects of different redistributive measures under varying assumptions about market structure, evaluate the efficiency of paternalism, or choose

20 *See, e.g.,* Guido Calabresi, The Future of Law and Economics (2016).

21 *See, e.g.,* Richard A. Posner, Law and Literature: A Misunderstood Relation (1988); Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 Mich. L. Rev. 489 (1989); Robin West, *Authority, Autonomy, and Choice: The Role of Consent in the Moral and Political Visions of Franz Kafka and Richard Posner*, 99 Harv. L. Rev. 384 (1985).

22 *See, e.g.*, Mark Kelman, *Consumption Theory, Production Theory, and Ideology in the Coase Theorem*, 52 S. Cal. L. Rev. 669 (1979).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

between different policy options while taking deontological moral constraints into account.[23]

V. Preventing Subsumption

A possible advantage of an integrative approach – at least from the perspective of legal scholars – is that in specific interdisciplinary research, one discipline may become dominant over the other, setting the parameters of legitimate research paradigms, or in extreme cases, "swallowing" the less dominant field.[24] This risk is relevant to certain *Law and …* schools of thought, particularly those that are gradually evolving into strong interdisciplinary research, as may be occurring in some branches of law and economics. The dominance of economic methodology in this interdisciplinary research may diminish the role of legal reasoning and ultimately narrow the contributions of law and economics to legal scholarship.[25] Integrative legal research is less susceptible to this risk because its core focus remains on law – its rules, institutions, and principles – and because incorporating multiple perspectives inherently reduces the dominance of any single one.

D. Challenges and Concerns

Having analyzed the meaning and potential benefits of integrative legal research, this part discusses the challenges it faces and the concerns it raises.

---

23 *See, e.g.,* Bruce Ackerman, *Regulating Slum Housing Markets on Behalf of the Poor: Of Housing Codes, Housing Subsidies and Income Redistribution Policy*, 80 Yale L.J. 1093 (1971); Eyal Zamir, *The Efficiency of Paternalism*, 84 Va. L. Rev. 229 (1998); Zamir & Medina, *supra* note 17.

24 *Compare* Sheldon Stryker, *Consequences of the Gap between the "Two Social Psychologies"*, in The Future of Social Psychology: Defining the Relationship between Sociology and Psychology 83, 93–97 (Cookie W. Stephan, Walter G. Stephan & Thomas F. Pettigrew eds., 1991).

25 For the claim that legal studies are unlikely to be "colonized" by other disciplines, because law is primarily a profession, and legal scholars have a deeply felt professional identity, see M. Balkin, *Interdisciplinarity as Colonization*, 53 Wash. & Lee L. Rev. 949, 964–70 (1995).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

I. Superficiality

No single scholar can master all research methodologies and relevant literatures across multiple disciplines. There is a concern that attempting to cover too many fields may lead to superficiality or even charlatanism. However, I believe this challenge can be overcome.[26] Integration between research methods and perspectives does not replace specialized expertise in individual interdisciplinary fields. Rather, integrative research builds on and contributes to specialized interdisciplinary research. The experience of recent decades suggests that understanding the key insights of interdisciplinary legal research does not typically require deep expertise in adjacent disciplines. In fact, there is sometimes an inverse relationship between the methodological complexity of interdisciplinary legal research and its contribution to legal scholarship. While technically sophisticated research often advances a given interdisciplinary field, interdisciplinary legal studies that significantly impact legal thought generally do not rely on highly complex methodologies. Even when they do, their insights can usually be presented in intuitive and accessible terms.

Just as narrow interdisciplinary studies lay the foundation for integrative research, integrative research can, in turn, enrich these narrower studies. Integrative studies may suffer from gaps in specific expertise, but they gain from a broader perspective – or more precisely, from the multiplicity of perspectives they offer. They can generate new insights and ideas, reveal non-trivial connections between distinct bodies of knowledge, and challenge established truths. Some of these ideas, connections, and challenges may ultimately not withstand critical scrutiny, yet they will stimulate legal thinking.

Integrative legal research does not imply that all methodologies and competing schools of thought in legal academia are equally productive, and certainly not that they are equally productive in every legal context. There is no simple measure for choosing between different possible combinations of perspectives on any given issue. The choice depends on both the skills and inclinations of the researcher and on the issue itself. For example, an integrative analysis of market-related fields (such as contract law, corporate law, and antitrust law) would find it difficult to ignore insights from the economic analysis of law, and an integrative analysis

26 *See also* Guido Calabresi, *An Introduction to Legal Thought: Four Approaches to Law and to the Allocation of Body Parts*, 55 Stan. L. Rev. 2113, 2151 (2003).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of issues concerning the relationship between the government and parliament would likely require insights from political science. At the same time, integrative research should not shy away from incorporating feminist analytical tools in the study of contract law or from applying economic analysis in the study of criminal law.

II. Cumbersomeness

A prominent feature of the formation of a research discipline is the development of a unique vocabulary shared by those engaged in that field. Interdisciplinary research that addresses people who are not experts in one of the fields must be articulated in less technical language and sometimes even explain basic concepts from the other field.[27] These requirements apply even more strongly to integrative studies, which by their nature address individuals who are not experts in each of the fields that feed the research. Meeting these requirements may lead to lengthening the work or even making it cumbersome. In more severe cases, the price of meeting these requirements may be an oversimplification of the insights and concepts imported from various disciplines and a loss of subtlety.

Again, these risks should not be ignored, but they do not tip the scale against integrative studies. It appears that the great majority of contributions of interdisciplinary studies to legal theory can be presented in terms that do not require technical skill and expertise in other disciplines. Indeed, significant breakthroughs – such as Calabresi and Melamed's classification of property rules, liability rules, and inalienability,[28] and Ayres and Gertner's introduction of the notion of penalty default rules[29] – often prompt a series of responses that progressively refine the original insight through increasing methodological sophistication. However, the importance of these responses is often inversely proportional to the degree of sophistication required to present them. Usually, the strongest arguments against new ideas can be presented in language that is not overly technical.

27 *Cf.* Richard H. Steckel, *Strategic Ideas in the Rise of the New Anthropometric History and Their Implications for Interdisciplinary Research*, 58 J. Econ. History 803, 816 (1998).

28 Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 Harv. L. Rev. 1089 (1972).

29 Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 Yale L.J. 87 (1989).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

It follows that integrative legal research does not necessarily require the use of jargon or a collection of types of technical jargon. The common use of footnotes and the possible use of appendices in legal writing can help in this regard, as well. Footnotes can be used to explain terms taken from other disciplines, and appendices can be used to present arguments in a technical or formalized manner.

III. Community and Audience

One of the risks of engaging in integrative research is that there is often no well-defined reference group for such studies. For example, a study aiming to demonstrate, through philosophical analysis, that there is no contradiction between economic efficiency and paternalism, that findings from cognitive psychology indicate the necessity of paternalism, and that economic analysis methods can be used to assess the efficiency of paternalistic legal rules – may fall between the cracks. Scholars in the mainstream of law and economics might show little interest in it because its conclusions contradict the prevailing ideological perspectives in their field. At the same time, those who advocate communitarian views, altruism, or even paternalism might disregard the study due to their aversion to economic analysis.

When legal academia is structured into distinct silos (law and economics, feminist analysis, legal philosophy, etc.), studies that do not fit neatly into one of these silos may have a harder time gaining recognition than those that do. However, in this respect, integrative legal research is not fundamentally different from any other academic genre, which may initially encounter scepticism or even hostility from more established and entrenched schools of thought – including interdisciplinary schools. Moreover, there is also the potential for such research to resonate with scholars from different schools of thought, forging connections across disciplinary boundaries.

IV. Institutional and Pedagogical Implications

Encouraging integrative legal research has institutional and pedagogical implications. At the institutional level, fostering this type of research requires legal scholars to engage with multiple disciplines, collaborate with

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

experts from different fields, and participate in intellectual exchanges that bridge methodological divides. However, as many interdisciplinary studies demonstrate, such collaboration is feasible.

The encouragement of integrative legal research should also be pursued at the institutional level. Beyond promoting diverse academic training among law faculty members, appointing faculty with formal training in other disciplines, and establishing joint appointments between law and other university departments, it is essential to establish and nurture integrative research centers that provide the necessary conditions for fostering such interdisciplinary dialogue.

Similarly, one of the defining features of the fragmentation of legal academia in recent decades has been the establishment of associations and research centers, the organization of conferences, and the publication of journals focused on specific interdisciplinary fields. While these institutions, activities, and platforms are undoubtedly valuable, there is also merit in encouraging associations, research centers, and initiatives centered around substantive areas of law, such as contract law, that integrate diverse perspectives on these core legal fields. A dialogue between scholars engaged, for example, in feminist analysis of contract law and those conducting economic or historical analyses of the same doctrines may be just as fruitful as discussions between scholars applying economic analysis to contract law and those doing the same in criminal or tax law. Such interdisciplinary exchanges can contribute not only to the development of contract law but also to the refinement and advancement of each of the respective interdisciplinary approaches.

The integrative approach to legal research also has significant implications for legal education. In addition to doctrinal, comparative, and historical courses, as well as courses specializing in specific interdisciplinary perspectives, it is essential to develop integrative courses. These courses should expose students simultaneously to multiple perspectives on the law and diverse analytical methodologies.

E. Conclusion (On Elephants and Cathedrals)

The legal academy in systems that have undergone the interdisciplinary revolution is fragmented – and similar fragmentation may occur in systems currently undergoing this transformation. Greater tolerance for competing perspectives and a genuine appreciation of the value of plu-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ralism are needed. But these are not enough. An additional ideal to strive for is that, alongside the continued existence of separate research efforts within each school and the dialogue between scholars from different traditions, an effort should be made to integrate insights from these various schools.

The integrative research envisioned in this Essay respects and engages seriously with legal doctrines, aiming to contribute to their development and improvement. It regards legal practitioners – lawyers and judges – not merely as objects of legal research but also as partners and interlocutors in the evolution of the law. At its best, integrative legal research takes a genuine interest in other legal systems, both past and present, seeking to learn from how they have confronted and continue to address legal questions within their unique legal traditions and socio-economic contexts. This research approach has the potential to discern the valuable elements in doctrinal, comparative, historical, and interdisciplinary studies and extract the best from each.

Integrative legal research aspires to examine legal phenomena and analyze legal questions from multiple perspectives – both internally and externally – simultaneously. At times, it challenges the foundational assumptions of other disciplines, while at other times, it contributes to their development. Such research is likely to be more complex than many of the interdisciplinary studies prevalent today. However, simplicity – or even simplification – is not an advantage when the subject of study is a complex, multifaceted normative, institutional, and social system like the law.

From what I have said so far, one might infer that I believe that the integration of different methodologies and perspectives could yield a complete understanding of legal phenomena and definitive answers to legal questions. Like the parable of the elephant and the blind men, examining the elephant from all angles – rather than merely touching each of its individual parts – should lead to a full comprehension of its shape and essence. Scholars advocating for the integration of diverse theoretical and methodological perspectives sometimes conclude that such integration can produce a sophisticated grand theory or, at the very least, facilitate compromise and balance among competing theoretical frameworks.[30] This is not my argument.

30 Harold J. Berman, *Toward an Integrative Jurisprudence: Politics, Morality, History*, 76 Cal. L. Rev. 779 (1988); Winfried Brugger, *Legal Interpretation, Schools of Jurispru-*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The parable of the elephant (like the equally famous imagery in recent legal literature of Claude Monet's paintings of the Rouen Cathedral)[31] can be interpreted in multiple ways. According to one interpretation, the elephant – like normative truth or legal reality – has an objective existence, independent of the partial or clumsy attempts to discern and understand what the elephant looks like, what the correct legal ruling is, or what legal reality truly is. In this view, the Rouen Cathedral also stands unchanged regardless of how it appears at different times of the day or under varying weather conditions.

Another interpretation – aligning with the parable's original meaning in the Buddhist tradition – suggests that no single perspective, nor any combination of perspectives, can ever fully reveal the elephant, determine the correct legal resolution, or capture legal reality as it truly is. This may be because no objective legal reality or single correct answer exists at all or because we are incapable of comprehending and describing them within our conceptual framework.[32]

I am not claiming, therefore, that a sophisticated combination of perspectives and methodologies in legal literature will reveal a preexisting legal truth. At the same time, the argument in this Essay does not align with a nihilistic view of law or epistemology.[33] Like the parable of the elephant (and the cathedral imagery), this defence of integrative legal research is compatible with different epistemological perspectives and diverse conceptions of law, legal practice, and legal scholarship.[34]

Regardless of the perspective one adopts, the proposal here is to go beyond examining just one part of the elephant, viewing only a single image of the cathedral, or relying solely on one perspective on law. Instead, we should strive to integrate multiple examinations and viewpoints.

---

dence, and Anthropology, 42 Am. J. Comp. L. 395 (1994); Matthias Klat, *Integrative Jurisprudence: Legal Scholarship and the Triadic Nature of Law*, 33 Ratio Juris 380, 392 (2020).

31 Calabresi & Melamed, *supra* note 28.

32 David M. Zlotnick, *The Buddha's Parable and Legal Rhetoric*, 58 Wash. & Lee L. Rev. 957, 961–65 (2001).

33 For critical analyses of such *arrogant*, *optimistic* or *nihilist* uses of the elephant parable in legal literature, see Zlotnick, *supra* note 32.

34 *Cf.* Vincent C. Kavaloski, *Interdisciplinary Education and Humanistic Aspiration: A Critical Reflection*, in Interdisciplinarity and Higher Education, *supra* note 2, at 225.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Democracy and Paternalism

Axel Ockenfels & Carl Christian von Weizsäcker

A. Introduction

In Wikipedia we read the following definition of paternalism:

> **"Paternalism** is action that limits a person's or group's liberty or autonomy against their will and is intended to promote their own good. It has been defended in a variety of contexts as a means of protecting individuals from significant harm, supporting long-term autonomy, or promoting moral or psychological well-being. Such justifications are commonly found in public health policy, legal theory, medical ethics, and behavioral economics, where limited intervention is viewed as compatible with or even supportive of personal agency."

This definition is rooted in John Stuart Mill´s book "On Liberty" 1859/1991. So, the Wikipedia entry continues, quoting Mill:

> "It is, perhaps, hardly necessary to say that this doctrine [i.e. that individual liberty should only be restricted to protect a person or to protect others] is meant to apply only to human beings in the maturity of their faculties. We are not speaking of children, or of young persons below the age which the law may fix as that of manhood or womanhood."

Paternalism then is an attitude which promotes and justifies action from "above" to control and guide actions of others considered to be "below" those who are their "guides". The idea of "Democracy" corresponds to the opposite: the people "below" choose their agents, who are supposed to solve any joint problems in the interest of the people. Democracy, well understood, is a form of government of people who are free: liberty is a precondition of democracy.

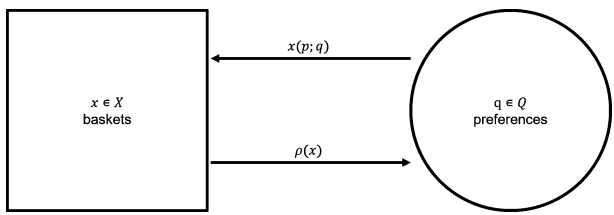
Based on earlier work by von Weizsäcker 2024 and Ambuehl, Bernheim, Ockenfels 2021 we try to come to grips about the relation between the two contrarian concepts of "paternalism" and "democracy".

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

B. Preferences and Freedom

Weizsäcker 2024 is about a consistent concept of welfare economics for the case that preferences of citizens are endogenously determined. The book generalizes traditional welfare economics, as pioneered by Samuelson 1938. In traditional welfare economics preferences are assumed to be exogenously given. In Weizsäcker 2024 it is shown that this welfare economic approach can be generalized, if and only if endogenously formed preferences are "adaptive". Further, the book also shows that the results of behavioral economics are consistent with the hypothesis of "adaptive preferences".

In the Weizsäcker 2024 approach endogenous preferences are modeled by means of the concept of a "preference system". We assume the existence of a space $\mathfrak{X}$ of potential consumption baskets, and we assume the existence of a space $\mathbb{Q}$ of potential preferences. The following figure indicates the idea of a preference system.



There is a mapping $x(p;q)$ from the space of preferences $\mathbb{Q}$ into the space of consumption baskets. It is the well-known demand function: the price vector p indicates the budget constraint; the preference vector q is an element of the space of preferences $\mathbb{Q}$. What distinguishes the preference system from the conventional demand function is the additional presence of a mapping $q = \rho\,(x)$ from the space of consumption baskets $\mathfrak{X}$ to the space of possible preferences $\mathbb{Q}$. This makes preferences an endogenous variable. Here we assume that it is the actual consumption vector x which, with some time delay, induces preferences. This is the concept of "induced preferences".

Within this general framework of a preference system, we then introduce a special case: it is the concept of "adaptive preferences". Intuitively speaking, adaptive preferences exhibit the feature of a certain preference

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

conservatism. To take as an example the well-known phenomenon of loss aversion. Within a conventional framework of the von-Neumann-Morgenstern utility function, loss aversion violates the assumption of fixed preferences. However, loss aversion exhibits a certain preference conservatism: people put an extra weight on a situation which avoids losses of their *actual* level of wealth.

The formal definition of adaptive preferences is this: consider two consumption baskets x and y such that with preferences $q = \rho(x)$ induced by x basket y is preferred over x. Preferences are defined to be *adaptive*, if it then is also the case that y is preferred over x, if preferences are induced by y.

By introducing the idea of "progress" or "improvement", we can provide an intuition, why adaptiveness of preferences is an important ingredient for welfare economics in a world of endogenously formed preferences. For simplicity of presentation, assume that inducement of preferences happens with a delay of one year. Consider the following sequence of baskets through time. In the first year we have basket x which induces preferences $q = \rho(x)$. These preferences prevail in the second year. Now, in the second year we have consumption basket y of which we assume that with prevailing preferences $q = \rho(x)$ the basket y is preferred over x. So, the move from x to y can be called an improvement. Let us then assume that in year 3 the basket x again prevails. Preferences in year 3 are induced by the basket y, which prevailed in year 2. Assume now that, contrary to the assumption of adaptive preferences, with preferences induced by y basket x is preferred over basket y. We then observe an improvement sequence from x to y and then back to x. We call this improvement sequence a pseudo-improvement. It is obvious that welfare economics basically becomes impossible, if such pseudo-improvements cannot be ruled out.

The assumption of adaptive preferences makes this particular kind of pseudo-improvements impossible. A large part of Weizsäcker 2024 is devoted to the proof that within the models developed in the book the assumption of adaptive preferences is sufficient to rule out pseudo-improvements and that, thereby, welfare economics of endogenously determined preferences is possible.

Theorems 1 and 2 provide such proof. Generally, we assume that the mappings between $\mathfrak{X}$ and $\mathbb{Q}$ are continuous for the topologies of these spaces. Furthermore, we have a non-satiation assumption and a certain single-crossing assumption. We operate in a discrete time model and in

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

addition in a continuous-time model. The latter we call the "real world model", because the real world exhibits continuous time.

Theorem 1 (in its different varieties, depending on the specifics of different models) says that, with adaptive preferences, any sequence of improving baskets of any finite length is a-cyclic, i.e., cannot come back to the origin.

Theorem 2 (in its different varieties) says this: if every sequence of improving baskets of any finite length is acyclic then preferences are adaptive.

There is then equivalence between the assumption of adaptive preferences and the assumption that welfare economics remains possible, when preferences are endogenously formed. As explained in detail in the book, for the case of endogenously formed preferences, the assumption of adaptive preferences can be seen as the generalization of the Samuelson-Houthakker axioms of revealed preference, which apply for the traditional model of fixed preferences. Note that fixed preferences are a special case of adaptive preferences.

As discussed in the book, not everybody in society acts in an autonomous mode. Thus, for example, small children do not behave in their own best interest. Every society accepts that parents interfere with their own small children so as to avoid behavior which is deleterious to the actor himself/herself. Freedom or liberty reasonably presupposes that people behave in an autonomous manner. However, autonomous behavior is not innate but must be acquired. This is the reason that every society requires a certain minimum age before people obtain all the rights a society of free people provides. Weizsäcker 2024 discusses this autonomy requirement for a society of free people (for example, Chapters 21, 22, 23, 24, 27).

The central idea for the path towards autonomy is "education". Successful schooling and education always requires that preferences are adaptive (Weizsäcker 2024, Chapter 20, Section 5). If we believe in education, we already implicitly believe in the prevalence of adaptive preferences. However, education is required for the goal of autonomous behavior. "Education" derives from Latin "educare" = educate, related to "educere" = "lead out" (Oxford Dictionary of English Etymology, Oxford 1966). By appropriate education people are brought (led out) from infantile immaturity towards autonomy.

As history has shown, such education towards civil liberty is not self-evidently successful. Karl Popper pointed to the great philosopher Plato as the intellectual fountain of 20th Century totalitarian ideology (Weizsäcker 2024, Chapter 27). In Plato´s "Republic" the "philosopher king" was the

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

only man in society with a successful "education" towards autonomy. And history, as well as our present world politics tell us of the many successes of such totalitarian ideology. There is the great temptation of solving the freedom problem by simply abolishing freedom.

Thus, in a sense, a democratic society of free people is not free in its choice of the details of the great educational enterprise toward citizens´ autonomy. It has to take account of the pitfalls of education towards mature citizenship. There are constraints to be taken in account in the design of educational procedures. In our present paper, we do not discuss the general characteristics of this educational enterprise. We only consider one aspect, which is personal intertemporal choice.

Individual intertemporal choice, for example in terms of consumption and labor supply, is very much linked to the subjectively perceived causation of future events. Animals have little understanding of causal chains. They simply follow their instincts, like hunger, fatigue, fear, curiosity, attachment to their mother, and, when grown up, sexual drives. Similarly, small children mainly follow their instincts. Paradise is the place, where their wants, generated by their instincts, are immediately satisfied. Paradise is the place with no need for intertemporal choice. Growing human understanding of the causal connection of events in the world at large, imposes on human action the need to anticipate the future. This leads to the need to think in terms of intertemporal trade-offs. It is the need to understand the opportunity cost of immediate satisfaction of wants. "Paradise lost". To learn about intertemporal trade-offs is an important part of the education towards autonomy.

The social process of education involves a majority of people as "educators", as "teachers". Parents are educators of their children; the social production machine is at the same time a complex educational machine with experienced workers as teachers and newcomers, apprentices as pupils. Modern society owns many schools, from kindergarten up to universities. We should understand that a large fraction of altruistic behavior within the family, among friends and colleagues takes the form of educating other people.

C. Experimental Evidence on Paternalism

Educating others is a quintessential form of altruism, and indeed of paternalism. The central dilemma is the tension between helping others

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

to make better decisions, and promoting learning while respecting the learner's autonomy.

Recent experimental evidence by Ambuehl, Bernheim, and Ockenfels (2021) shows how ordinary "choice architects" (CAs) navigate this trade-off. In their laboratory market for intertemporal payments, CAs design opportunity sets for "choosers". They regularly delete impatient options, even when non-binding advice could have been offered instead, because they believe the choosers will be better off.

Thus, in practice, education often takes the form of restricting choice rather than providing information.

However, the experiment also reveals the challenges associated with such interventions: CAs lack reliable information about another individual's preferences. How, then, do they decide which paths are helpful and which are harmful? Ambuehl et al. trace their judgements to the projective inferences first described by Adam Smith in his Theory of Moral Sentiments. These inferences follow from self-examination: "As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation." (Part I, Section I, Chapter I).

In this spirit, Ambuehl et al. identify two types of paternalism: those who reason about others based on their own mistakes and those who reason based on their own preferences.

A mistakes-projective paternalist assumes that others tend to share their susceptibility to error. They behave as if they are trying to help others avoid choices that they themselves would reject, but which they nevertheless are tempted to choose. This inclination generates a negative correlation between the choices she makes for herself and the restrictions she imposes on others in Ambuehl et al.'s setup.

In contrast, an ideals-projective paternalist behaves as if her own preferences are relevant to others, either because she believes they share her values or because she thinks her perspective is valid and theirs is not. Ideals-projective paternalism generates a positive correlation between the choices paternalists make for themselves and the restrictions they impose on others.

As ideals-projective reasoning very much dominates in the data, interventions are systematically misguided, even according to the CAs' own welfare criteria. They remove impatient options even when all payoffs are delayed, i.e. in settings that neutralize present bias arguments. In other words, CAs do not intervene by removing options that they wish they could re-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

sist when choosing for themselves (mistakes-projective paternalism); rather, they intervene as if they seek to align others' choices with their own aspirations (ideal-projective paternalism). False consensus bias compounds the error: patient CAs underestimate how many choosers they influence.

Furthermore, Ambuehl et al. demonstrate that the relationships between policy preferences and consumption outside the laboratory are consistent with ideals-projective paternalism. For example, lighter drinkers express significantly more support for increasing alcohol taxes in another jurisdiction.

More recent evidence from Isoni, Ockenfels, Sugden and Zheng (2025, in progress) complements these findings with data from the demand side. They begin with the observation that political approval for "helping" paternalistic interventions (i.e. the supply of paternalism) and the observation that some people choose to self-constrain (e.g. Sunstein and Thaler's New Year's resolution test) are sometimes presented as evidence of demand for paternalism. However, the supply of paternalism by CAs may be motivated by a desire to help or be helped (or both). In other words, a preference for self-constraint does not necessarily imply consent for externally imposed constraints.

Indeed, in a series of experiments, Isoni et al. found that subjects reported believing that the paternalistic decisions of an unknown other would be similar to their own, and that their own decisions were very similar to their inner demands. However, many subjects were happy to bind themselves, but objected to being bound by an unknown other. The demand for paternalism is much smaller than the supply.

Isoni et al.'s data exclude some hypotheses for this pattern, such as the existence of naïve individuals who do not realize they need help, or that subjects value decision autonomy in itself. An alternative approach could be based on bounded cognition: intuitive heuristics suffice for self-constraint and other-constraint, but make it difficult to evaluate whether constraints imposed by others are desirable.

To summarize, Weizsäcker (2024) provides a normative basis for combining welfare analysis and behavioral economics. His preference system recognizes that consumption today shapes preferences tomorrow. Welfare economics is only possible when preferences are adaptive, i.e. when improvements to welfare under current tastes are not undone once tastes adapt. In this context, adaptive preferences may justify educative or even paternalistic guidance as a learning aid on the path to autonomy.

Laboratory experiments, on the other hand, reveal the psychological

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

mechanisms of education and paternalism. While well-meaning, people often overstep the mark when intervening in others' choices, eventually limiting autonomy in a way that decreases welfare. CAs rely on their own (adapted) ideals as the best available proxy for the learner's preferences. However, the false consensus bias makes CAs overconfident in the relevance of their own ideals to learners. Moreover, while people seem happy to restrict the autonomy of others, they may not fully consider that they would reject being paternalized themselves.

Educational paternalism need not be a contradiction. When adaptive preferences make some paternal guidance unavoidable, the ideals of projective educators can provide the social energy that drives progress – but they must be guided by evidence of actual learner welfare. Blending Weizsäcker's macro theory with micro evidence from Ambuehl et al. and Isoni et al. provides an explanation of the circumstances in which paternalism promotes, rather than diminishes, autonomy.

Of course, there are forms of education that do not require direct behavior control. Providing people with advice and more options to empower them to make informed decisions that respect and extend one's authority is often possible. Ockenfels (2023), among others, provides examples in "behavioral market design". There does not always need to be a conflict between autonomy and education.

D. Institutions

There is a history of institutions spanning thousands of years. And, a long history of political philosophy reflecting the observed gradual or abrupt changes of prevailing institutions. What can we learn for our theme from this history? It is obvious that successful working of democracy requires an institutional framework. Historical attempts to abandon all existing institutions and to replace them with completely different ones have always failed. As examples we may point to the developments of the great French revolution, 1789 onwards, to the Russian revolution of 1917 onwards, or to Mao Zedong's "Cultural Revolution" in the sixties and seventies of the last century.

We then can ask: how much paternalism is contained in those institutions which are typical for flourishing modern democracies? Take the case of franchise or suffrage. Historically, suffrage was more restricted than it is today. In most countries suffrage for women was not introduced earlier

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

than in the 20th century. Obviously then the laws concerning franchise in earlier times contained a lot of paternalism.

Another, more subtle form of paternalism is the two-houses organization of legislatures. The Constitution of the United States of America prescribes that the legislature, the Congress, consist of two chambers: the Senate and the House of Representatives. Whereas there are general elections for the House every two years, senators are elected for a period of six years. The Constitution thereby expresses some distrust into the majority decisions of the House. Underlying is the well-known fact that the short-run effects of law-changes differ from their long-run effects. And the further idea that the voters do not fully understand the consequences of law changes. There is, of course, a substantial economics literature about the ensuing political short-termism.

The longer election period for senators then can be seen as a kind of constitutional paternalism – in this case against excessive short-termism of political decisions. We should add that the fathers of the American constitution were strongly influenced by Montesquieu, in particular by his great work on the "Spirit of Law". There he expounds on the "separation of powers" into legislative, executive, and judiciary.

The problem of potential short-termism is also at the core of the institution of an independent central bank. After the disastrous experience of two German hyperinflations after World War I and after World War II, the Federal Republic of Germany decided in the early fifties of the 20th century that the newly established central bank should obtain a high degree of independence. The Bundesbank then was able to make the Deutsche Mark a rather stable currency. In his work across several decades Alex Cukierman has empirically well established that currencies under control of an independent central bank had a substantially smaller inflationary bias than currencies under a more direct control of democratically elected governments: cf. for example Cukierman 1992. There is a short-termism of democratically elected governments. It is due to the fact that the electorate tends to evaluate the economic performance of the parties in power according to the actual state of the economy at the election time. Due basically to the division of labor, actual overall economic performance and the current rate of inflation are closely linked. So, the parties in power pursue an economic policy with an inflationary bias. By pursuing the primary goal of price stability, an independent central bank takes account of the welfare time lags involved in monetary policy. Only thereby can it provide the overall benefits which derive from price stability. It is then again a case

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

of paternalism so as to provide greater long run benefits at the expense of present short run benefits.

E. Conclusion

Through both theoretical and empirical lenses, we examined some aspects of the relationship between democracy and paternalism. Weizsäcker (2024) establishes a normative foundation for merging welfare analysis and behavioral economics by presenting a preference system that acknowledges the influence of current consumption on future preferences. Welfare economics is only possible when preferences are adaptive, meaning improvements to welfare under current tastes do not undo themselves once tastes adapt. In this context, adaptive preferences may justify educative or even paternalistic guidance as a learning aid on the path to autonomy.

Laboratory experiments reveal the psychological mechanisms underlying such interventions. Although well-intentioned, people often overstep when interfering with others' choices, thereby limiting autonomy and decreasing welfare. Choice architects rely on their own (adapted) ideals as the best available proxy for learners' preferences. However, false consensus bias makes them overconfident in the relevance of their own ideals. Furthermore, people readily restrict the autonomy of others but typically reject being paternalized themselves.

Educational paternalism need not be a contradiction. When adaptive preferences make some guidance unavoidable, the ideals of projective educators can provide the social energy that drives progress. However, they must be guided by evidence of actual learner welfare. Blending Weizsäcker's macro theory with micro evidence from Ambuehl et al. and Isoni et al. helps understanding when paternalism promotes rather than diminishes autonomy.

Of course, there are forms of education that do not require direct behavioral control. Providing advice and expanding options to empower informed decisions that respect and extend one's autonomy is often possible. Ockenfels (2023), among others, provides examples of "behavioral market design" that guides behavior by extending rather than narrowing down choice menus.

Ultimately, successful democratic institutions must navigate this tension by incorporating limited, evidence-based paternalistic elements, such

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

as age restrictions on voting, bicameral legislatures, and independent central banks, that protect long-term collective welfare while preserving the fundamental principle of self-governance. Some paternalism is necessary to create and maintain conditions conducive to genuine democratic autonomy. The challenge is to identify the optimal boundaries of such interventions and developing mechanisms to ensure they remain accountable to democratic oversight.

References

Ambuehl 2021, Sandro, B. Douglas Bernheim, Axel Ockenfels, What Motivates Paternalism? An Experimental Study. American Economic Review 111 (3): 787–830.

Cukierman, Alex 1992, Central Bank Strategy, Credibility, and Independence: Theory and Evidence, MIT Press, 496 pages.

Isoni 2025, Andrea, Axel Ockenfels, Robert Sugden, and Jiwei Zheng, The Demand and Supply of Paternalism, Work in Progress.

Mill 1859/1991, John Steward, On Liberty, published in Gray, John (ed), John Stuart Mill, On Liberty and other Essays, Chapter 1, Oxford, Oxford University Press.

Montesquieu 1748, Charles de Secondat, Baron de. *The Spirit of the Laws* (*De l'esprit des lois*). Geneva: Barillot & Fils.

Ockenfels 2023, Axel, Behavioral Market Design, Behavioral and Brain Sciences, 46, e171.

Samuelson 1938, Paul, A Note on the Pure Theory of Consumer´s Behavior, Economica, 5, 61–71.

Smith 1759, Adam. The Theory of Moral Sentiments. London: A. Millar, 1759.

Thaler 2008, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Weizsäcker 2024, Carl Christian von, Freedom and Adaptive Preferences, London, Routledge, XIV + 228 pages.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

The Engel Method
Psychological Insights and Practical Guidelines for Interdisciplinary Research Collaboration

Susann Fiedler & Andreas Glöckner

A. Introduction

Societies today face challenges that defy simple solutions: growing disparities in wealth and opportunity, demographic shifts that strain existing institutions, and the disruptive impact of new technologies on work, privacy, and civic life (Lyall, Bruce, Tait, & Meagher, 2011). Addressing these problems demands more than disciplinary expertise – it requires research teams capable of integrating different ways of thinking about systems, incentives, rules, and human behavior (Jacobs & Frickel, 2009; Klein, 1990). Yet while interdisciplinary collaboration is widely seen as essential to tackle these "wicked problems" and billions in research funding are invested in interdisciplinary science each year, many teams discover that combining diverse perspectives is harder in practice than in principle (Cummings & Kiesler, 2005). Many teams falter – not for lack of expertise, but because they underestimate the human dynamics of collaboration (Stokols, Hall, Taylor, & Moser, 2008; Fiore, 2008).

Every interdisciplinary team holds hidden knowledge, but often these insights remain locked away unless the right conditions make it safe and worthwhile to share them (Stasser & Titus, 1985; Edmondson, 1999). Interdisciplinary science promises creativity, but too often delivers confusion, status conflicts, or premature fragmentation (Khan, Dyaram, & Dayaram, 2022; Cauwelier, Ribiere, & Bennet, 2019). Understanding why some teams harness their diversity to generate breakthroughs while others descend into misunderstanding and inertia is therefore a central challenge for contemporary science (Wax, Asencio, Bentley, & Warren, 2024). Part of the answer lies in recognizing that interdisciplinarity is not merely an organizational or logistical challenge, but fundamentally a psychological one (Edmondson, 1999; Cannon & Edmondson, 2005). Of course, structural and institutional barriers play a role: misaligned incentives such

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

as tenure criteria privileging disciplinary outputs, or funding and recognition systems unsuited to collaborative work, as well as coordination complexity and difficulties in harmonizing data and tools (Cummings & Kiesler, 2005) make it difficult to sustain interdisciplinary projects. These dimensions are essential to address at policy and institutional levels. Yet even if structures and logistics are perfectly aligned, collaborations can still falter when psychological foundations – trust, openness, shared purpose, and a willingness to revise one's own assumptions – are not in place. This paper therefore focuses on the psychological and relational factors that determine whether diverse expertise becomes a source of progress, and offers evidence-based recommendations for building teams capable of transforming complex societal problems into actionable insights. In our examples we focus on interdisciplinary work in science, but most of the conclusions should be general enough to generalize to other areas as well.

B. Defining Interdisciplinarity

"Interdisciplinarity" at its core describes different ways of connecting knowledge across fields. Multidisciplinarity places disciplines side by side without true integration, interdisciplinarity combines perspectives into shared frameworks, and transdisciplinarity reaches beyond academia to co-produce knowledge with practitioners and stakeholders (Klein, 1990; Jacobs & Frickel, 2009; Lyall, Bruce, Tait, & Meagher, 2011). These forms sit on a continuum of integration: from loose cooperation to deeply joint problem-solving (Stokols, Hall, Taylor, & Moser, 2008). The higher the level of integration, the greater the psychological demands on researchers – coordination, trust, and openness to other ways of knowing.

From a psychological perspective, the challenge is not only that disciplines offer different answers, but that they ask different questions also concerning the same topic. Concerning the topic of how to generate knowledge in the scientific process, for example, economists may model incentives, psychologists study cognition and motivation, and legal scholars interpret institutions and rules. Such divergent perspectives make collaboration difficult, but also create the potential for insights that no single field could reach alone (Fiore, 2008; Salazar, Lant, Fiore, & Salas, 2012). Multiple perspectives might allow generating a multi-dimensional understanding of the topic, while keeping with one disciplinary perspective only can reduce the complexity of the topic onto a single dimension. In

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

this paper, interdisciplinary collaboration is understood as the intentional integration of disciplinary perspectives into shared insights and multi-dimensional mental models.

C. Barriers and Psychological Foundations

Interdisciplinary collaboration is widely expected to generate breakthroughs, yet many initiatives struggle to deliver on this promise (Cummings & Kiesler, 2005). While institutional incentives and logistical arrangements shape outcomes, they rarely determine them. This can be due to the fact that the mental model of group behavior fails to take into account important dimensions. Even well-resourced projects can lose momentum if, for example, the interpersonal and psychological dynamics within the team are not addressed. The following section therefore examines the underlying psychological mechanisms that help explain why some teams thrive while others stall.

I. Barriers Concerning Language, Methodology, and Epistemological Approach

One basic challenge in interdisciplinary work lies in the fact that sciences often use different subject-specific language both for matters of content as well as methodology. During their academic studies, researchers develop explicit and implicit knowledge concerning theories, concept definitions, methods, and paradigms. As a result, there is shared knowledge in the field concerning the meaning of technical labels (i.e., words), which allows communicating more efficiently. In psychology, for example, there is a shared knowledge concerning what dissonance, attitude, and reactance mean since these concepts have been developed in their respective theoretical frameworks. In economics, there is a shared understanding concerning the meaning of (orthodox) rationality, or Nash-equilibrium etc. Often these terms are related to certain research paradigms that are used in the respective field to conceptualize situations and to test theories such as social dilemmas, the gambling paradigm, or game theory.

A shared understanding of the meaning of terms allows discussion on a higher level of abstraction. Because humans can only hold a relatively small number of information units (i.e., 3 to 7 "chunks") active in working mem-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ory, prior knowledge is required to use larger information units (i.e. more abstract chunks) to capture the complexity of real-world problems that typically contain much more pieces of information. Knowledge of the concepts from neighboring disciplines could help achieve a more comprehensive understanding of the problem, without overcharging human memory.

In interdisciplinary work, challenges concerning concept labels can arise on three different levels: (a) individuals might not understand the technical labels of the other discipline at all (lack of knowledge), (b) the disciplines might use the same label to refer to different concepts (an interdisciplinary "jingle fallacy"), and (c) different labels are used to refer to the same concept or method (an interdisciplinary "jangle fallacy").

For an interdisciplinary research team, it is crucial to take sufficient time to develop a shared understanding of the most relevant concepts, theories, and paradigms for the problem at hand. To cover basics and more comprehensive thoughts, this can involve (a) joint workshops reading and discussing textbook chapters (e.g., on game theory), (b) paper clubs reading current relevant papers from all fields, and (c) mutually presenting and discussing own scientific work in early and later stages.

In the optimal case, these discussions lead to a deeper and broader understanding of the concepts, to the identification of weaknesses and blindspots of the own disciplines' theories, and acknowledgement of similar work and related concepts from other fields.

Additionally, disciplines apply different epistemological approaches. That is, disciplines differ in their commonly accepted methods for generating and updating knowledge and standards used to evaluate evidence. The quantitative empirical approach, often applied in psychology, uses empirical data to update beliefs about relations in the world that are summarized in theories. Results are evaluated concerning whether proper (experimental) research designs were used (i.e. assuring internal and external validity) and proper statistics were applied to analyze the data. A theoretical approach, applied for example in theoretical economics or mathematics, often starts from minimal basic assumptions or axioms and derives predictions using rules of logic and other mathematical operations to explain and predict phenomena in the world. These derivations have to be reproducible and free of errors, and generate new insights by deriving many predictions from minimal assumptions. In law, hermeneutic approaches dominate, focusing on interpreting written text and the construction of consistent arguments, with validity evaluated by the conclusiveness of the argument or measures of consensus.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

These epistemological differences are reinforced by divergent disciplinary languages (as introduced above), standards for presenting work, and assumptions about legitimate knowledge. They interact with group dynamics such as ingroup favoritism, collusion within disciplines, and differences in status or reputation. In such contexts trust in sources and appreciation of others' expertise are indispensable. Successful integration depends on remaining critical while staying open to belief revision, even when this challenges deeply held assumptions (e.g., debates about deception in experiments).

II. Relational Barriers to Integration

Establishing a shared language is an important foundation, but it does not remove all challenges. Interdisciplinary collaboration frequently confronts teams with high levels of epistemic uncertainty, diverging methodological assumptions, and the need to integrate heterogeneous standards of evidence (Jacobs & Frickel, 2009; Stokols, Hall, Taylor, & Moser, 2008).

Three relational mechanisms are particularly important. First, psychological safety and error culture shape whether researchers are willing to take interpersonal risks and treat mistakes as learning opportunities rather than threats (Edmondson, 1999; Cannon & Edmondson, 2005; Wax, Asencio, Bentley, & Warren, 2024). Second, the hidden profile problem demonstrates how easily unique knowledge is overlooked when groups focus on shared information rather than eliciting unshared expertise (Stasser & Titus, 1985; Dayeh & Morrison, 2020). Third, perceived hierarchies and power asymmetries across disciplines can marginalize certain perspectives, reducing the diversity of voices and limiting integration (Cauwelier, Ribiere, & Bennet, 2019; Khan, Dyaram, & Dayaram, 2022).

Taken together, these relational barriers underline that interdisciplinarity is not only about reconciling language, theories or methods; it is equally about cultivating trust, openness, and mutual respect (Fiore, 2008; Salazar, Lant, Fiore, & Salas, 2012). Without such conditions, epistemological differences become sources of division rather than opportunities for synthesis. The following section therefore turns to relational barriers to interdisciplinary collaboration.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

1. Psychological Safety and Error Culture in Interdisciplinary Research Teams

Interdisciplinary work confronts researchers with ambiguity and the constant need to engage across boundaries of method and evidence. Under such conditions, psychological safety, the shared belief that the context is safe for interpersonal risk-taking, becomes a decisive factor in determining whether collaboration succeeds (Edmondson, 1999; Wax, Asencio, Bentley, & Warren, 2024). When researchers feel safe, they are more likely to voice doubts, challenge assumptions, and propose unconventional ideas; when safety is absent, silence, self-censorship, and disciplinary retreat are more likely (Cauwelier, Ribiere, & Bennet, 2019; Khan, Dyaram, & Dayaram, 2022).

A critical component of psychological safety is the treatment of errors. Disciplines differ markedly in their error cultures: in some fields, mistakes are framed as personal failings, while in others they are considered natural by-products of innovation. Where teams cultivate an open error management culture, mistakes become opportunities for learning and reflection; where blame dominates, members censor themselves and innovation stalls (Cannon & Edmondson, 2005; Oliveira, Santos, & Ratten, 2022). Leaders play a key role by modeling vulnerability, acknowledging their own limitations, and rewarding openness rather than punishing missteps (Agarwal & Farndale, 2017; Ahmad & Umrani, 2019).

Psychological safety also helps to mitigate status asymmetries that are common in interdisciplinary teams. Perceived prestige differences, such as between quantitative and qualitative methods, can silence scholars from less dominant traditions (Cummings & Kiesler, 2005; Khan et al., 2022). Safety provides the conditions in which all voices can be heard, allowing unique disciplinary perspectives to be expressed and integrated (Fiore, 2008; Salazar, Lant, Fiore, & Salas, 2012).

In practice, fostering psychological safety in research collaborations requires deliberate strategies. Teams benefit from norm setting and role models that emphasize respect, inclusion, and the value of dissent (Edmondson, 1999; Wax, Asencio, Bentley, & Warren, 2024). Equally important are transparent decision-making and error-handling procedures, which reduce ambiguity and strengthen trust (Stokols, Hall, Taylor, & Moser, 2008). In collaborative science, this transparency extends to research practices themselves. Open science tools such as preregistration, registered reports, and shared data and material repositories institution-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

alize an error-friendly culture by making it legitimate to report null results, deviations from hypotheses, and unforeseen challenges (Nosek et al., 2018; Chambers, 2019). Such practices protect researchers from reputational risks that often undermine psychological safety, while simultaneously improving the robustness of scientific outputs.

Policy and institutional frameworks can further reinforce these dynamics. Funding agencies and journals are beginning to recognize that the success of research collaborations depends not only on scientific merit but also on transparent processes and a culture of learning. By incentivizing preregistration, replication studies, and open sharing of protocols, institutions reduce the stigma of error and create structural conditions that align with psychological safety (Nosek et al., 2015). Moreover, initiatives that explicitly reward collaboration quality – such as team-based evaluations, cross-disciplinary grants, and recognition for open practices – signal that innovation depends on openness as much as on disciplinary expertise. Ultimately, psychological safety and error culture are not peripheral concerns but foundational to whether interdisciplinary research teams can move beyond disciplinary defensiveness and harness their diversity for cumulative, transparent, and impactful science.

2. Hidden Profiles and Information Sharing

Even when interdisciplinary teams establish a climate of psychological safety, they face a persistent cognitive barrier: the hidden profile problem. Classic research shows that groups tend to spend the majority of their time discussing information already shared by all members, while unique expertise and unshared insights remain overlooked (Stasser & Titus, 1985; Wittenbaum, Hollingshead, & Botero, 2004). This bias toward common information undermines the very rationale of interdisciplinarity, which depends on surfacing and integrating distinct disciplinary perspectives (Mesmer-Magnus & DeChurch, 2009).

The mechanisms behind hidden profiles are well documented. Teams prefer shared information because it feels more fluent, socially validating, and easier to process (Stasser & Stewart, 1992; Lu, Yuan, & McLeod, 2012). For interdisciplinary collaborations, this challenge is amplified. Scholars not only hold different pieces of information, but also rely on divergent epistemological assumptions, methods, and standards of evidence. These differences increase the likelihood that unique disciplinary knowledge is not voiced (Cummings & Kiesler, 2005; Salazar, Lant, Fiore, & Salas,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

2012). Research further shows that perceived competence shapes information sharing: members who see themselves as less competent relative to others are paradoxically more likely to share and seek information, leading to better decisions, while those who feel more competent often withhold their unique knowledge (Dayeh & Morrison, 2020). In interdisciplinary settings, this dynamic risks silencing precisely those disciplinary contributions that are most needed for integration.

Addressing hidden profiles requires structured processes that counteract natural tendencies. Explicit protocols such as assigning roles, conducting "rounds" for unshared insights, or using structured decision-making techniques (e.g., Delphi procedures) have been shown to increase the disclosure of unique information (Stasser & Stewart, 1992; Sohrab, Waller, & Kaplan, 2015; Wittenbaum et al., 2004). Ultimately, the hidden profile literature highlights a paradox at the heart of interdisciplinarity: the knowledge that most needs to be voiced is least likely to be shared spontaneously. Overcoming this paradox requires deliberate structuring of collaboration processes, recognition of disciplinary asymmetries, and reinforcement of norms that value the surfacing of unique information. A simple solution in interdisciplinary teams is that (e.g.) the leader of the group ensures that each discipline provides their perspective on the topic at hand by directly asking "what does [discipline] say to this topic?", and appreciating the elicited answers.

3. Perceived Hierarchies and Power Dynamics

Interdisciplinary teams not only bring together different domains of expertise but also researchers with varying levels of mastery in methods, theories, and certain knowledge areas. These asymmetries create implicit hierarchies between perceived "experts" and "novices" that can strongly shape collaboration processes. Expertise often affords status, making it more likely that an individual's contributions will be taken seriously, while those who are newer to a field or less confident in their methodological skills may hesitate to speak up (Boix Mansilla, Lamont, & Sato, 2016). These dynamics can have both costs and benefits. On the one hand, experts anchor discussions and provide legitimacy to methodological choices. On the other hand, they may inadvertently dominate conversations, dismiss alternative viewpoints, or steer decisions toward familiar solutions rather than innovative integrations. Novices, by contrast, may bring fresh perspectives precisely because they are less constrained by dis-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ciplinary conventions, but only if they feel encouraged and empowered to voice them (Edmondson, 1999; Khan, Dyaram, & Dayaram, 2022).

Power dynamics also shape how errors and uncertainties are handled. Research shows that members with lower status are less likely to admit mistakes or raise doubts, fearing reputational costs, whereas high-status experts may downplay their own errors to maintain authority (Cannon & Edmondson, 2005). These patterns undermine the open exchange of knowledge, especially in collaborations where methodological complexity means that no single member can oversee all aspects of the project (Hall, Vogel, Huang, & Crowston, 2018). Overcoming these barriers requires structuring participation so that both experts and novices contribute meaningfully. Explicit invitations for input from less experienced members can redistribute influence and create a more balanced environment (Salazar, Lant, Fiore, & Salas, 2012; Boix Mansilla et al., 2016). Training that emphasizes humility for experts and confidence-building for novices can further help teams avoid reproducing rigid hierarchies and instead use differences in expertise as a resource for creativity and integration. A pragmatic measure is asking less experienced persons (e.g., PhD students) to provide their opinions first, and the leader of the group to speak last. Generally, it should be assured that critical thoughts concerning the dominant opinion (potentially established by high-power persons) are voiced and encouraged (Janis, 1971).

In this way, perceived hierarchies are not inherently detrimental. When managed well, they can provide structure and clarity while still leaving space for learning, innovation, and the constructive questioning of established assumptions. When unmanaged, however, they silence valuable contributions and reinforce defensiveness – precisely the opposite of what interdisciplinary collaboration requires.

III. Cognitive Diversity and Belief Revision

Cognitive diversity – the presence of different perspectives, heuristics, and problem-solving strategies within a team – lies at the heart of interdisciplinarity. By combining varied lenses, teams increase their potential to generate creative insights, uncover blind spots, and develop more sustainable solutions (Page, 2007; Hong & Page, 2004). In interdisciplinary collaborations, cognitive diversity is not incidental but constitutive: each field brings distinctive assumptions about causality, methods, and evi-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dence, which together expand the space of possible solutions (Salazar, Lant, Fiore, & Salas, 2012; Hall, Vogel, Huang, & Crowston, 2018).

Yet cognitive diversity is a double-edged sword. While it increases the potential for innovation, it also raises coordination costs and the risk of conflict (Mello & Rentsch, 2015; Mannix & Neale, 2005). Disagreements over fundamental assumptions can stall progress if team members are unwilling or unable to revise their beliefs. Interdisciplinary collaborations thus depend critically on the ability of researchers to not only present their views but also to update them in light of unfamiliar evidence and reasoning.

Belief revision is a demanding psychological process. Studies show that individuals exhibit confirmation bias and motivated reasoning, preferring information that aligns with their prior views (Nickerson, 1998; Taber & Lodge, 2006). These tendencies are amplified in group settings where disciplinary identities are salient, making it easier to defend one's own paradigm than to integrate competing ones. At the same time, cognitive conflict, if managed constructively, can lead to deeper elaboration, more systematic processing of information, and ultimately better decisions (De Dreu & Weingart, 2003; van Knippenberg, De Dreu, & Homan, 2004).

Successful teams develop the ability to surface, evaluate, and combine diverse perspectives while maintaining a climate of mutual respect (Salazar et al., 2012). Leadership and team processes are key: transformational leadership, structured deliberation, and explicit emphasis on critical reflection all increase the likelihood that cognitive diversity will translate into innovation rather than gridlock (Wang, Kim, & Lee, 2016; Boix Mansilla, Lamont, & Sato, 2016). Importantly, belief revision is facilitated by trust: members are more willing to adjust their assumptions when they regard colleagues as competent and benevolent (Dirks & Ferrin, 2001; Edmondson, 1999).

For interdisciplinary science, the challenge is not to eliminate cognitive diversity but to harness it. Teams must learn to tolerate prolonged periods of uncertainty, to hold competing representations in play, and to adopt procedures that encourage genuine updating rather than defensive entrenchment. When achieved, this capacity transforms disciplinary plurality from a source of friction into a wellspring of creativity and resilience. Without it, collaborations risk dissolving into parallel monologues, with each field defending its own territory rather than building something new.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

D. Recommendations for Practice

The mechanisms reviewed in this chapter: barriers concerning language and methodology, psychological safety, error culture, information sharing, expertise dynamics, and belief revision highlight not only why collaborations falter, but also what is needed to create effective conditions for interdisciplinarity. The goal is not simply to remove obstacles, but to design environments where diverse expertise can flourish. There are many examples for successful interdisciplinary scientific work, often relying on different approaches. In this section we use the organization of an interdisciplinary research group at the Max Planck Institute for Research on Collective Goods by Christoph Engel as a case-study to illustrate a solution that worked extremely well in practice and had substantial long-term consequences. We refer to the distilled approach as the *Engel-Method*.

I. Develop a Shared Language and Epistemological Awareness

An ideal interdisciplinary environment requires a common language that allows researchers to communicate across disciplinary boundaries and make their contributions mutually intelligible. Without it, collaborations risk misinterpretations and the persistence of disciplinary silos.

Christoph Engel created multiple structures to foster this shared language. He established paper seminars and a common corpus of readings so that everyone could engage with the same foundational texts. Joint methods seminars and summer schools gave researchers the opportunity to compare their approaches and identify common ground. Presentations of the same topic from different disciplinary perspectives encouraged participants to recognize both divergences and overlaps in their conceptual frameworks. Practical tools such as a joint glossary and shared "work horse" projects (i.e. projects with a joint research paradigm, cf. Betsch & Bröder, 2025) further anchored collaboration in a common vocabulary.

Engel also modeled the translational work himself by moving across fields – law, game theory, behavioral economics, psychology, meta-science, and computer science – becoming a novice every few years. By experiencing the disorientation of learning a new field, he demonstrated both the necessity and the value of building a shared language for integration. Combined, these strategies provide the cognitive infrastructure that

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

allows teams to operate at a higher level of abstraction and to integrate diverse insights into coherent solutions.

II. Create a Culture of Psychological Safety and Constructive Error Management

At the foundation of an ideal collaborative environment lies a culture in which researchers feel safe to voice doubts, challenge assumptions, and admit a lack of knowledge. Psychological safety is cultivated when leaders model vulnerability, curiosity, and frame errors as learning opportunities (Edmondson, 1999; Cannon & Edmondson, 2005). Christoph Engel consistently embodied this principle by openly discussing how his own papers and results had come about – not only the successes, but also the manipulations that failed and the effects that could not be shown. Some of his publications, as he frequently explained, had emerged directly from such failures and the need to rethink assumptions. When teaching new methods, first in Stata and later in Python, he would deliberately demonstrate where analyses could go wrong, which assumptions had to be revised, and how his own thinking evolved over time. He also invited external experts, such as David E. Jaeger for statistics, to tackle open questions alongside students, making clear that even senior researchers continue to learn. By sharing these experiences candidly, Engel normalized error as part of scientific progress, signaling that not knowing is never a problem – only the refusal to engage and learn is.

III. Ensure Systematic Information Sharing

Interdisciplinary teams achieve their potential best when all members' unique expertise is surfaced. Environments that institutionalize structured information sharing – whether through rotating facilitators, formal "information rounds," or the facilitation of informal discussions – make it less likely that valuable insights remain hidden (Stasser & Titus, 1985; Mesmer-Magnus & DeChurch, 2009). Beyond individual projects, infrastructures such as shared databases, joint seminars, and transparent communication platforms can embed these practices into the daily routines of collaborative research.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Christoph Engel actively created such conditions at the Max Planck Institute. He introduced a mentorship system that paired senior and junior scientists from different research areas, encouraging informal exchanges outside of large meetings and lowering the threshold for asking questions. The physical environment reinforced these opportunities: the table tennis court and the institute's garden became central spaces for spontaneous conversations, allowing researchers to discuss viewpoints and problems in smaller, more relaxed rounds. The environment encouraged joint lunch and coffee breaks across disciplinary borders. Engel also shaped seminar formats so that a junior researcher would present a paper or method, supported by a senior colleague. This practice both empowered early-career scientists and ensured that disciplinary boundaries were bridged in constructive ways.

Engel himself modeled and encouraged information sharing by raising respective questions – famously asking, "what does [*psychology*] say to this?" – and by consistently bringing new methods, ideas, and results to the group, making them available for everyone to use. These practices demonstrated that information sharing was not only encouraged but expected, and that knowledge was a collective resource rather than an individual possession. In this way, Engel embedded both structures and informal mechanisms for exchange into the daily fabric of the institute, ensuring that valuable insights surfaced and interdisciplinary collaboration became the norm rather than the exception.

IV. Balance Expertise with Opportunities for Growth

Ideal interdisciplinary environments balance the authority of established experts with opportunities for novices to learn, contribute, and develop. Without such balance, collaborations risk either being dominated by senior voices or leaving less experienced researchers without pathways to growth. Environments that deliberately recognize the expertise of all members, regardless of career stage, create the conditions for both integration and innovation (Boix Mansilla, Lamont, & Sato, 2016).

Christoph Engel consistently worked to achieve such balance within his institute. He ensured that senior and junior researchers were paired in complementary ways, and when the distribution of staff was uneven, he invited external scholars to fill gaps, maintaining diversity of perspectives and methods. Importantly, Engel emphasized that every junior researcher

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

was an expert in their own right: they brought specialized knowledge of a content area, and regularly highlighted these contributions in group discussions. This recognition elevated junior members from passive learners to active collaborators, reinforcing the idea that interdisciplinary work depends on drawing from everyone's expertise.

Knowledge growth was further accelerated through extensive summer schools, where researchers, regardless of seniority, moved quickly through in-depth courses and their own first experimental projects. These intensive programs compressed learning and experimentation into a short time, allowing juniors to gain rapid confidence in their methods and insights. Notably, Engel was not only the teacher in these settings but often also a curious student himself – participating in the same curriculum, engaging in discussions, and modeling the humility required for continuous learning. By creating structures in which senior and junior voices were intentionally balanced, and by positioning himself as a co-learner, Engel ensured that expertise flowed in multiple directions. This approach both empowered younger scientists and reminded established scholars of the value of openness and curiosity.

V. Embed Reflexivity and Belief Revision

Interdisciplinary teams thrive when belief revision is built into their routines. This requires more than ad hoc discussions: it calls for institutionalized reflexivity – regular workshops where assumptions are questioned, methods are debated, and integration is made explicit. In ideal environments, disagreement is not seen as a threat but as a signal of productive diversity, provided it is handled in a climate of trust.

At the Max Planck Institute, Christoph Engel established precisely such an environment. Reflexivity was not confined to formal workshops but woven into the everyday routines of research. Opportunities to meet and openness to discuss were abundant: many projects began over lunch or during informal conversations in the garden or hallway. This somewhat grass-roots, even "anarchistic" approach ensured that new ideas could surface quickly and without the burden of rigid hierarchies. The architecture and meeting culture reinforced this openness. Offices had open-door policies, feedback was given quickly and constructively, and researchers could rely on colleagues who had no other commitments than to focus on science and the facilitation of new ideas. This created conditions in which

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

"flow states" of collaboration were supported and reflexivity became a natural part of daily interaction.

Engel also institutionalized specific practices that encouraged belief revision. He allowed for fast appointments of seminar presentations – sometimes even opening the 2 p. m. lunch slot – so that ideas could be tested and refined without delay. The internal discussion paper series provided friendly but rigorous feedback, offering a safe forum for critical exchange before external submission. Across these formats, there was a strong perception that colleagues were eager to help and discuss – not only in scheduled meetings, but also in the spontaneous encounters that made reflexivity a lived, collective norm. Through this combination of openness, flexibility, and institutionalized feedback mechanisms, Engel created an environment where assumptions were continuously surfaced and revised, not as an exception but as the everyday practice of interdisciplinary science.

E. Conclusion

Interdisciplinary research succeeds when structural support and institutional incentives are matched by environments that foster openness, curiosity, and respect. Psychological safety, information sharing, balanced expertise, and reflexivity help ensure that disciplinary diversity becomes a resource rather than a barrier.

Christoph Engel showed us what such an environment can look like. By treating errors as opportunities, by honoring the expertise of juniors alongside seniors, by creating spaces for spontaneous exchange, and by stepping into new fields as a learner himself, he built a culture where curiosity, openness, and mutual respect were the norm. All these experiences teach us that interdisciplinarity is not only about combining knowledge – it is about building communities of trust where learning never stops. Christoph Engel applied a very unique, both deeply humanistic and scientific approach, which justifies referring to it as the Engel Method.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

References

Agarwal, P., & Farndale, E. (2017). High-performance work systems and creativity implementation: The role of psychological capital and psychological safety. *Human Resource Management Journal*, *27*(3), 440–458. https://doi.org/10.1111/1748-8583.12148

Ahmad, I., & Umrani, W. A. (2019). The impact of ethical leadership style on job satisfaction. *Leadership & Organization Development Journal*, *40*(5), 534–547. https://doi.org/10.1108/lodj-12-2018-0461

Betsch, T., & Bröder, A. (2025). The wheels of scientific progress: Integrative theory building in decision science. *Zeitschrift für Psychologie / Journal of Psychology, 233*(4). https://doi.org/10.1027/2151-2604/a000605

Boix Mansilla, V. B., Lamont, M., & Sato, K. (2015). Shared cognitive-emotional-interactional platforms. *Science Technology & Human Values*, *41*(4), 571–612. https://doi.org/10.1177/0162243915614103

Cannon, M. D., & Edmondson, A. C. (2005). Failing to learn and learning to fail (intelligently): How great organizations put failure to work to innovate and improve. *Long Range Planning, 38*(3), 299–319. https://doi.org/10.1016/j.lrp.2005.04.005

Cauwelier, P., Ribiere, V. M., & Bennet, A. (2019). The influence of team psychological safety on team knowledge creation: A study with French and American engineering teams. *Journal of Knowledge Management, 23*(6), 1157–1175. https://doi.org/10.1108/jkm-07-2018-0420

Chambers, C. (2019). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice.* Princeton University Press.

Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science, 35*(5), 703–722. https://doi.org/10.1177/0306312705055535

Dayeh, V., & Morrison, B. W. (2020). The effect of perceived competence and competitive environment on team decision-making in the Hidden-Profile paradigm. *Group Decision and Negotiation*, *29*(6), 1181–1205. https://doi.org/10.1007/s10726-020-09701-2

Dirks, K. T., & Ferrin, D. L. (2001) The role of trust in organizational settings. *Organization Science, 12*(4), 450–467. https://doi.org/10.1287/orsc.12.4.450.1064

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Fiore, S. M. (2008). Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research, 39*(3), 251–277. https://doi.org/10.1177/1046496408317797

Hall, K. L., Vogel, A. L., & Crowston, K. (2019). Comprehensive collaboration plans: practical considerations spanning across individual collaborators to institutional supports. In *Strategies for team science success: Handbook of evidence-based principles for cross-disciplinary science and practical lessons learned from health researchers* (pp. 587–612). Cham: Springer International Publishing.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Hong, L., & Page, S. E. (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl. Acad. Sci. U.S.A., 101*(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101

Jacobs, J. A., & Frickel, S. (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology, 35*, 43–65. https://doi.org/10.1146/annurev-soc-070308-115954

Janis, I. L. (1971). Groupthink: The desperate drive for consensus at any cost. *Psychology Today, 5*(6), 43–76.

Khan, N., Dyaram, L., & Dayaram, K. (2022). Team faultlines and upward voice in India: The effects of communication and psychological safety. *Journal of Business Research, 142*, 540–550. https://doi.org/10.1016/j.jbusres.2022.01.009

Klein, J. T. (1990). *Interdisciplinarity: History, theory, and practice*. Wayne State University Press.

Lu, L., Yuan, Y. C., & McLeod, P. L. (2011). Twenty-five years of hidden profiles in group decision making. *Personality and Social Psychology Review*, *16*(1), 54–75. https://doi.org/10.1177/1088868311417243

Lyall, C., Bruce, A., Tait, J., & Meagher, L. (2011). *Interdisciplinary research journeys: Practical strategies for capturing creativity*. Bloomsbury Academic.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, *6*(2), 31–55. https://doi.org/10.1111/j.1529-1006.2005.00022.x

Mello, A. L., & Rentsch, J. R. (2015). Cognitive diversity in teams. *Small Group Research*, *46*(6), 623–658. https://doi.org/10.1177/1046496415602558

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, *94*(2), 535–546. https://doi.org/10.1037/a0013773

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … & Yarkoni, T. S. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.https://doi.org/10.1126/science.aab2374

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Oliveira, M. F., Santos, E., & Ratten, V. (2022). Strategic perspective of error management, the role of leadership, and an error management culture: A mediation model. *Journal of Economics Finance and Administrative Science*, *28*(55), 160–175. https://doi.org/10.1108/jefas-01-2022-0028

Page, S. E. (2007). Making the difference: Applying a logic of diversity. *Academy of Management Perspectives*, *21*(4), 6–20. https://doi.org/10.5465/AMP.2007.27895335

Salazar, M., & Lant, T. (2018). Facilitating innovation in interdisciplinary teams: The role of leaders and integrative communication. *Informing Science: the Inter-*

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

national Journal of an Emerging Transdiscipline, 21, 157–178. https://doi.org/10.28945/4011

Salazar, M. R., Lant, T. K., Fiore, S. M., & Salas, E. (2012). Facilitating innovation in diverse science teams through integrative capacity. *Small Group Research*, *43*(5), 527–558. https://doi.org/10.1177/1046496412453622

Sohrab, S. G., Waller, M. J., & Kaplan, S. (2015). Exploring the hidden-profile paradigm. *Small Group*

Research, *46*(5), 489–535. https://doi.org/10.1177/1046496415599068

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology*, *63*(3), 426–434. https://doi.org/10.1037/0022-3514.63.3.426

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467–1478. https://doi.org/10.1037/0022-3514.48.6.1467

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine, 35*(2), S77–S89. https://doi.org/10.1016/j.amepre.2008.05.002

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Wang, X., Kim, T., & Lee, D. (2016). Cognitive diversity and team creativity: Effects of team intrinsic motivation and transformational leadership. *Journal of Business Research, 69*(9), 3231–3239. https://doi.org/10.1016/j.jbusres.2016.02.026

Wax, A., Asencio, R., Bentley, J. R., & Warren, C. (2024). Safety first! Psychological safety as a driver of learning in functionally diverse, self-assembled teams. *Team Performance Management, 30*(7/8), 173–194. https://doi.org/10.1108/tpm-02-2024-0017

Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs*, *71*(3), 286–310. https://doi.org/10.1080/0363452042000299894

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

VII.
Appendices

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Christoph Engel: List of Publications

A. Publications in Peer-Reviewed Journals

106 Is the German Constitutional Court Partisan?, in: Journal of Legal Studies ***.
105 Out of Sight is Out of Mind? Experimentally Testing a Gradually Materializing Public Bad, in: Journal of Economic Psychology *** (with Alexander Egberts and Joshua Fairfield).
104 Behaviorally Efficient Remedies. An Experiment, in: International Review of Law and Economics *** (with Lars Freund).
103 Crime for a Living, in: Journal of Institutional and Theoretical Economics 181 (2025) 133–135.
102 Who is Afraid of the Pink Elephant? Evidence on (Not) Ignoring Inadmissible Evidence and Debiasing Interventions, in: Judgement and Decision Making *** (with Jasmin Golder and Rima Rahal).
101 Integrating Machine Behavior into Human Subject Experiments: A User-friendly Toolkit and an Application to Framed Prisoner's Dilemmas, in: Experimental Economics *** (with Max Großmann and Axel Ockenfels).
100 LLM as a Law Professor. Having a Large Language Model Write a Commentary on Freedom of Assembly, in: Artificial Intelligence and Law *** (with Johannes Kruse).
99 When Do Guardians Deter Offending? An Experimental Test of Informal Social Control Mechanisms, in: Journal of Quantitative Criminology *** (with Charles Lanfear and Daniel Nagin).
98 Sabotaging Competitors, Both Real and Illusory, in: Judgement and Decision Making *** (with Dan Simon).
97 The Valuation of Moral Rights. A Field Experiment, in: American Law and Economics Review *** (with Stefan Bechtold).
96 Calling in at Dr. GPT, in: Journal of Institutional and Theoretical Economics 180 (2024) 365–367.
95 Deliberately Ignoring Inequality to Avoid Rejecting Unfair Offers, in: Communications Psychology 2 (2024) 48 (with Dorothee Mischkowski, Konstantin Offer, Zoe Rahwan).
94 Is Transparency a Blessing or a Curse? An Experimental Horse Race Between Accountability and Extortionary Corruption, in: International Review of Law and Economics 78 (2024) 106189 (with Eyal Zamir).
93 What Makes Cooperation Precarious? in: Journal of Economic Psychology 101 (2024) 102712 (with Bettina Rockenbach).
92 Bounded Cognition and the Taking of Personal Property: An Experiment, in: American Law and Economics Review *** (with Oren Bar-Gill).
91 Code is Law. How COMPAS Affects the Way the Judiciary Handles the Risk of Re-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

cidivism, in: Artificial Intelligence and Law 2024, https://doi.org/10.1007/s10506-024-09389-8 (with Lorenz Linhardt and Marcel Schubert).

90 Outcomes or Participation? Experimentally Testing Competing Sources of Legitimacy for Taxation, in: Economic Inquiry 62 (2024) 563–583 (with Luigi Mittone and Azzurra Morreale).

89 Does the Fundamental Transformation Deter Trade? An Experiment, in: American Law and Economics Review 25 (2023) 338–380 (with Eric Helland).

88 Hide and Seek in the Judiciary, in: Journal of Institutional and Theoretical Economics 179 (2023) 152–155.

87 Do Not Draw a Black Ball from the Justice Lottery, in: Journal of Institutional and Theoretical Economics 179 (2023) 118–120.

86 How Little Does It Take to Trigger a Peer Effect? An Experiment on Crime as Conditional Rule Violation, in: Journal of Research in Crime and Delinquency 60 (2023) 455–492.

85 Diffusion of Legal Innovations, in: Annual Review of Law and Social Science 18 (2022) 139–152 (with Keren Weinshall).

84 Lucky You: Your Case is Heard by a Seasoned Panel: Panel Effects in the German Constitutional Court, in: Journal of Empirical Legal Studies 19 (2022) 1179–1221.

83 Intensified Support for Juvenile Offenders on Probation. Evidence from Germany, in: Journal of Empirical Legal Studies 19 (2022) 447–490 (with Sebastian Goerg and Christian Traxler).

82 Challenges in the Interdisciplinary Use of Comparative Law, in: American Journal of Comparative Law 69 (2021) 777–797.

81 Managing Expectations. How Selective Information Affects Cooperation and Punishment in Social Dilemma Games, in: Journal of Economic Behavior and Organization 187 (2021) 111–136 (with Sebastian Kube and Michael Kurschilgen).

80 People Are Conditional Rule Followers, in: Journal of Economic Psychology 85 (2021) 102384 (with Pieter Desmet).

79 Social Mindfulness is Normative When Costs are Low, but Rapidly Declines with Increases in Costs, in: Judgment and Decision Making 16 (2021) 290–322 (with Paul van Lange).

78 Machine Advice with a Warning about Machine Limitations. Experimentally Testing the Solution Mandated by the Wisconsin Supreme Court, Journal of Legal Analysis 13 (2021) 284–340 (with Nina Grgic-Hlaca).

77 Tax Morale and Fairness in Conflict. An Experiment, in: Journal of Economic Psychology 81 (2020) 102314 (with Luigi Mittone and Azzurra Morreale).

76 Manna from Heaven for Judges. Judges' Reaction to a Quasi-Random Reduction in Caseload, in: Journal of Empirical Legal Studies 17 (2020) 722–751 (with Keren Weinshall).

75 Estimating Heterogeneous Reactions to Experimental Treatments, in: Journal of Economic Behavior and Organization 178 (2020) 124–147.

74 The Fragility of a Nudge. The Power of Self-Set Norms to Contain a Social Dilemma, in: Journal of Economic Psychology 81 (2020) 102293 (with Michael Kurschilgen).

73 Coherence-based Reasoning and Order Effects in Legal Judgments, in: Psychol-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

ogy, Public Policy and the Law 26 (2020) 333–352 (with Andreas Glöckner and Sinika Timme).

72 Uncertain Judges, in: Journal of Institutional and Theoretical Economics 176 (2020) 44–47

71 Deciding on Behalf of Others Does Not Mitigate Selfishness: An Experiment, in: Economics Letters 183 (2019) 108616 (with Claudia Cerrone).

70 When Does Transparency Backfire? Putting Jeremy Bentham's Theory of General Prevention to the Experimental Test, in: Journal of Empirical Legal Studies 16 (2019) 881–908.

69 How to Deal with Inconsistent Choices on Multiple Price Lists, in: Journal of Economic Behavior & Organization 160 (2019) 138–157 (with Oliver Kirchkamp).

68 The Accuracy–Discrimination Trade-Off, in: Journal of Institutional and Theoretical Economics 175 (2019) 123–125.

67 How to Protect Entitlements: An Experiment, in: Journal of Law and Economics 61 (2018) 525–553 (with Oren Bar-Gill).

66 Diffusion of Legal Innovations. The Case of Israeli Class Actions, in: Journal of Empirical Legal Studies 15 (2018) 708–731 (with Alon Klement and Keren Weinshall).

65 If the Worst Comes to the Worst. Dictator Giving When Recipients' Endowments are Risky, in: European Economic Review 105 (2018) 51–70 (with Sebastian Goerg).

64 Modeling a Satisficing Judge, in: Rationality and Society 30 (2018) 220–246 (with Werner Güth).

63 Empirical Methods for the Law, in: Journal of Institutional and Theoretical Economics 174 (2018) 5–23.

62 The Dark Side of Price Cap Regulation. A Lab Experiment, in: Public Choice 173 (2017) 217–240 (with Klaus Heine).

61 Insuring your Donation. An Experiment, in: Journal of Empirical Legal Studies 14 (2017) 858–885 (with Renate Buijze and Sigrid Hemels).

60 You Are In Charge. Experimentally Testing the Motivating Power of Holding a Judicial Office, in: Journal of Legal Studies 46 (2017) 1–50 (with Lilia Zhurakhovska).

59 The Hidden Cost of Compensation, in: Journal of Institutional and Theoretical Economics 173 (2017) 106–109.

58 Non-Compete Clauses, Employee Effort and Spin-off Entrepreneurship: A Laboratory Experiment, in: Research Policy 45 (2016) 2113–2124 (with Guido Bünstorf, Sven Fischer and Werner Güth).

57 At the Mercy of a Prisoner. Three Dictator Experiments, in: Applied Economics Letters 24 (2017) 774–778 (with Thorsten Chmura and Markus Englerth).

56 The Hog Cycle of Law Professors. An Econometric Time Series Analysis of the Entry-level Job Market in Legal Academia, in: PLOS One 11 (2016[7]) 1–22 (with Hanjo Hamann).

55 A Random Shock is not Random Assignment, in: Economics Letters 145 (2016) 45–47.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

54 Bargaining in the Absence of Property Rights. An Experiment, in: Journal of Law and Economics 59 (2016) 477–495 (with Oren Bar-Gill).
53 Symmetric vs. Asymmetric Punishment Regimes for Collusive Bribery, in: American Law and Economics Review 18 (2016) 506–556 (with Sebastian Goerg and Gaoneng Yu).
52 Does Class Action Have a Deterrent Effect?, in: Journal of Institutional and Theoretical Economics 172 (2016) 104–107.
51 When is the Risk of Cooperation Worth Taking? The Prisoner's Dilemma as a Game of Multiple Motives, in: Applied Economics Letters 23 (2016) 1157–1161 (with Lilia Zhurakhovska).
50 Homo Ignorans. Deliberately Choosing Not to Know, in: Perspectives on Psychological Science 11 (2016) 359–372 (with Ralph Hertwig).
49 Unpacking Negligence Liability. Experimentally Testing the Governance Effect, in: Journal of Empirical Legal Studies 13 (2016) 116–152 (with Theodore Eisenberg).
48 The People's Hired Guns? Experimentally Testing the Motivating Force of a Legal Frame, in: International Review of Law and Economics 43 (2015) 67–82 (with Alicja Reuben).
47 Tacit Collusion. The Neglected Experimental Evidence, in: Journal of Empirical Legal Studies 12 (2015) 537–577.
46 Who is Afraid of the Stick? Experimentally Testing the Deterrent Effect of Sanction Certainty, in: Review of Behavioral Economics 2 (2015) 405–434 (with Daniel Nagin).
45 Scientific Disintegrity as a Public Bad, in: Perspectives on Psychological Science 10 (2015) 361–379.
44 Randomized Information about the Law as an Instrument, in: Journal of Institutional and Theoretical Economics 171 (2015) 171–175.
43 Conditional Cooperation with Negative Externalities – An Experiment, in: Journal of Economic Behavior and Organization 108 (2014) 252–260 (with Lilia Zhurakhovska).
42 On Probation. An Experimental Analysis, in: Journal of Empirical Legal Studies 12 (2015) 252–288 (with Heike Hennig-Schmidt, Bernd Irlenbusch, Sebastian Kube).
41 Who is Afraid of Pirates? An Experiment on the Deterrence of Innovation by Imitation, in: Research Policy 44 (2015) 20–33 (with Marco Kleine).
40 Has The World Changed? My Neighbor Might Know: Effects of Social Context on Routine Deviation, in: Journal of Behavioral Decision Making 28 (2015) 50–66 (with Tilman Betsch, Stefanie Lindow, Corina Ulshöfer, Janet Kleber).
39 Assuring Civil Damages Adequately Deter: A Public Good Experiment, in: Journal of Empirical Legal Studies 11 (2014) 301–349 (with Theodore Eisenberg).
38 Social Preferences Can Make Imperfect Sanctions Work: Evidence from a Public Good Experiment, in: Journal of Economic Behavior and Organization 108 (2014) 343–353.
37 A Dynamic View on Justification, in: Journal of Institutional and Theoretical Economics 170 (2014) 189–192.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

36 What Does "Clean" Really Mean? The Implicit Framing of Decontextualized Experiments, in: Economics Letters 122 (2014) 386–389 (with David Rand).
35 dhreg, xtdhreg, bootdhreg: Programs to Implement Double Hurdle Regression, in: Stata Journal 14 (2014) 778–797 (with Peter Moffatt).
34 First Impressions are More Important than Early Intervention. Qualifying Broken Windows Theory in the Lab, in: International Review of Law and Economics 37 (2014) 126–136 (with Martin Beckenkamp, Andreas Glöckner, Bernd Irlenbusch, Heike Hennig-Schmidt, Sebastian Kube, Michael Kurschilgen, Alexander Morell, Andreas Nicklisch, Hans-Theo Normann, Emmanuel Towfigh).
33 Can We Trust Intuitive Jurors? Standards of Proof and the Probative Value of Evidence in Coherence Based Reasoning, in: Journal of Empirical Legal Studies 10 (2013) 230–252 (with Andreas Glöckner).
32 The Coevolution of Behavior and Normative Expectations. An Experiment, in: American Law and Economics Review 15 (2013) 578–609 (with Michael Kurschilgen).
31 Nudged to Be Consistent, in: Journal of Institutional and Theoretical Economics 169 (2013) 45–48.
30 Role Induced Bias in Court. An Experimental Analysis, in: Journal of Behavioral Decision Making 26 (2013) 272–284 (with Andreas Glöckner).
29 Fair Exclusion, in: Journal of Institutional and Theoretical Economics 168 (2012) 171–175.
28 The Emergence of a New Rule of Customary Law. An Experimental Contribution, in: Review of Law and Economics 7 (2011) 767–789.
27 Fairness Ex Ante and Ex Post. Experimentally Testing Ex Post Judicial Intervention into Blockbuster Deals, in: Journal of Empirical Legal Studies 8 (2011) 682–708 (with Michael Kurschilgen).
26 Dictator Games. A Meta-Study, in: Experimental Economics 14 (2011) 583–610.
25 When is Intellectual Property Needed as a Carrot for Innovators?, in: Journal of Competition Law and Economics 7 (2011) 277–299.
24 Contract as Exposure to Attack, in: Journal of Institutional and Theoretical Economics 167 (2011) 72–76.
23 The Behaviour of Corporate Actors. How Much can we Learn from the Experimental Literature?, in: Journal of Institutional Economics 6 (2010) 445–475.
22 The Multiple Uses of Experimental Evidence in Legal Scholarship, in: Journal of Institutional and Theoretical Economics 166 (2010) 199–202.
21 Erga Omnes. Why does Public International Law Ignore Privity of Contract, in: Journal of Institutional and Theoretical Economics 165 (2009) 24–28.
20 Learning the Law, in: Journal of Institutional Economics 4 (2008) 275–297.
19 Poor Judicial Performance: When Should the Parties Care?, in: Journal of Institutional and Theoretical Economics 164 (2008) 95–98.
18 How Much Collusion? A Meta-Analysis of Oligopoly Experiments, in: Journal of Competition Law and Economics 3 (2007) 491–549.
17 The Impact of Institutions on the Decision How to Decide, in: Journal of Institutional Economics 3 (2007) 323–349 (with Elke U. Weber).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

16 Competition in a Pure World of Internet Telephony, in: Telecommunications Policy 31 (2007) 530–540.
15 Exclusionary Bundling and the Effects of a Competitive Fringe, in: Journal of Institutional and Theoretical Economics 163 (2007) 133–137.
14 The Cognitive Effect of a Minimum Wage, in: Journal of Institutional and Theoretical Economics 163 (2007) 52–55.
13 Clumsy Solutions for a Complex World. The Case of Climate Change, in: Public Administration 84 (2006) 817–843 (with Marco Verweij, Mary Douglas, Richard Ellis, Frank Hendriks, Susanne Lohmann, Steven Ney, Steve Rayner, Michael Thompson). reprinted, with a postscript, in: Richard Fardon (ed.): Mary Douglas. Cultures and Crises, London 2013, 146–174 and 191–197.
12 The Role of Law in the Governance of the Internet, in: International Review of Law, Computers & Technology 20 (2006) 201–216.
11 Corporate Design for Regulability. A Principal-Agent-Supervisor Model, in: Journal of Institutional and Theoretical Economics 162 (2006) 104–124.
10 Corporate Culture as a Resource for Management, in: Journal of Institutional and Theoretical Economics 162 (2006) 97–100 (with Werner Güth).
9 Governing the Egalitarian Core of the Internet, in: International Journal of Communications Law and Policy 10 (2005) 1–25.
8 General and Specific Rules, in: Journal of Institutional and Theoretical Economics 161 (2005) 350–354.
7 Give the Journals Back to the Scientists, in: Journal of Institutional and Theoretical Economics 160 (2004) 35–38.
6 Modelling Fundamentalism, in: Journal of Institutional and Theoretical Economics 159 (2003) 163–170.
5 Causes and Management of Conflicts, in: Journal of Institutional and Theoretical Economics 159 (2003) 1–15.
4 Windows as an Institution Organizing the Markets for Applications Software, in: Journal of Institutional and Theoretical Economics 158 (2002) 155–162.
3 The European Charter of Fundamental Rights. A Changed Political Opportunity Structure and its Normative Consequences, in: European Law Journal 7 (2001) 151–170.
2 Delineating the Proper Scope of Government: A Proper Task for a Constitutional Court ?, in: Journal of Institutional and Theoretical Economics 157 (2001) 187–219.
1 Legal Responses to Bounded Rationality in German Administration, in: Journal of Institutional and Theoretical Economics 1994, 145–162.

B. Peer-reviewed Conference Proceedings

3 LEXam. Benchmarking Legal Reasoning on 340 Law Exams, in: International Conference on Learning Representations (ICLR), 2026, Rio de Janeiro (with Yu Fan, Jingwei Ni, Jakob Merane, Etienne Salimbeni, Yang Tian, Yoan Hermstrüwer, Yinya Huang, Mubashara Akhtar, Florian Geering, Oliver Dreyer,

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Daniel Brunner, Markus Leippold, Mrinmaya Sachan, Alexander Stremitzer, Christoph Engel, Elliott Ash, Joel Niklaus).

2 Lawma: The Power of Specialization for Legal Tasks, in: International Conference on Learning Representations (ICLR), 2025, Singapore (with Ricardo Dominguez, Vedant Nanda, Rediet Abebe, Stefan Bechtold, Jens Frankenreiter, Krishna Gummadi, Moritz Hardt, Michael Livermore).

1 Human Decision Making with Machine Assistance. An Experiment on Bailing and Jailing, in: 22nd Association for Computing Machinery (ACM) Conference on Computer-Supported Cooperative Work and Social Computing, 2019, Austin Texas (with Nina Grgić-Hlača and Krishna Gummadi).

C. Monographs

16 Die verfassungsrechtliche Zulässigkeit eines Entflechtungstatbestands im Kartellgesetz (Common Goods: Law, Politics and Economics 19) Baden-Baden (Nomos) 2008, 92 S.

15 Paketvermittelte Telefonie (Law and Economics of International Telecommunications 55) Baden-Baden (Nomos) 2006, 133 S.

14 Generating Predictability. Institutional Analysis and Institutional Design. Cambridge (Cambridge University Press) 2005, 315 S.

13 Verhandelter Netzzugang (Common Goods: Law, Politics and Economics 7) Baden-Baden (Nomos) 2002, 101 S.

12 Abfallrecht und Abfallpolitik (Common Goods: Law, Politics and Economics 5) Baden-Baden (Nomos) 2002, 358 S.

11 Global Networks and Local Values. A Comparative Look at Germany and the United States, Washington (National Academy Press) 2002, 241 S. (with the other members of the Steering Committee). German version under the titel "Globale Netze und lokale Werte. Eine vergleichende Studie zu Deutschland und der Vereinigten Staaten (Law and Economics of International Telecommunications 49) Baden-Baden (Nomos) 2002, 273 S.

10 Die Vorschriften des Telekommunikationsgesetzes über den Zugang zu wesentlichen Leistungen. Eine juristisch-ökonomische Untersuchung (Law and Economics of International Telecommunications 37) Baden-Baden (Nomos) 1998, 98 S. (with Günter Knieps).

9 Europarechtliche Grenzen für öffentlich rechtliche Spartenprogramme. Beihilfeaufsicht, Wettbewerbsregeln, Grundfreiheiten. Berlin (Vistas) 1996, 103 S.

8 Kabelfernsehen (Law and Economics of International Telecommunications 29) Baden-Baden (Nomos) 1996, 165 S.

7 Medienordnungsrecht (Law and Economics of International Telecommunications 28) Baden-Baden (Nomos) 1996, 144 S.

6 Gemischtwirtschaftliche Abfallentsorgung. Ein Lehrstück zur Verschränkung des Abfallrechts mit Kartellrecht, Kommunalrecht, dem Recht der öffentlichen Aufträge und Wirtschaftsverfassungsrecht (Schriften zum deutschen und europäischen Umweltrecht 6) Köln (Heymanns) 1995, 91 S.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

5 Privater Rundfunk vor der Europäischen Menschenrechtskonvention (Law and Economics of International Telecommunications 19) Baden-Baden (Nomos)1993, 487 S.
4 Planungssicherheit für Unternehmen durch Verwaltungsakt, Tübingen (Mohr Siebeck) 1992, 195 S.
3 Das Embargo gegen Irak und Kuwait. Entschädigungsansprüche gegen die Europäische Gemeinschaft und gegen die Bundesrepublik Deutschland – ein Rechtsgutachten (Wirtschaftsrecht und Wirtschaftspolitik 117) Baden-Baden (Nomos) 1991, 89 S. (with Ernst-Joachim Mestmäcker).
2 Der Einfluß des europäischen Gemeinschaftsrechts auf die deutsche Rundfunkordnung (Law and Economics of International Telecommunications 15) Baden-Baden (Nomos) 1990 (with Ernst-Joachim Mestmäcker, Karin Gabriel-Bräutigam und Martin Hoffmann; von Engel 1. Teil VII: Grundrechtsschutz in der Gemeinschaft = S. 75–101).
1 Völkerrecht als Tatbestandsmerkmal deutscher Normen (Tübinger Schriften zum internationalen und europäischen Recht 19) Berlin (Duncker & Humblot) 1989, 359 S.

C. Articles, not Peer-Reviewed

151 Interacting with the Disabled. Experimental Evidence on Helping, and on Deciding on Behalf of Others, in: Helen Alford / Fabio Ferrucci (eds.): Disability and the Human Condition, Vatican 2024, 300–314.
150 Gewichtsformel – wörtlich genommen. Ein empirischer Test mit der Hilfe eines Sprachmodells, in: FS Alexy, Tübingen 2025, 495–512.
149 Eye-Tracking as a Method for Legal Research, in: Kevin Tobia (ed.): The Cambridge Handbook of Experimental Jurisprudence, Cambridge 2025, 329–359 (with Rima Rahal).
148 Experimental Comparative Law 2.0? Large Language Models as a Novel Empirical Tool, in: Niels Philipsen / Sarah Schoenmaekers / Louis Visscher / Franziska Weber (eds.): By the end of the day. Liber amicorum Michael Faure, Deventer 2025, 417–428.
147 Allgemeines Verwaltungsrecht, in: Christoph Engel / Lawrence O'Hara / Stefanie Egidy / Yoan Hermstrüwer / Leonard Hoeft / Pascal Langenbach (eds.): Öffentliches Recht als Verhaltensordnung, Tübingen 2025, 135–166.
146 Verhalten: Ein Paradigma für das öffentliche Recht, in: Christoph Engel / Lawrence O'Hara / Stefanie Egidy / Yoan Hermstrüwer / Leonard Hoeft / Pascal Langenbach (eds.): Öffentliches Recht als Verhaltensordnung, Tübingen 2025, 1–10.
145 Kommentar ohne Autor. Können Sprachmodelle das Kommentieren übernehmen?, in: Juristenzeitung 79 (2024) 997–1007 (with Johannes Kruse).
144 Asking GPT for the Ordinary Meaning of Statutory Terms, in: Journal of Law, Technology & Policy 2024, 235–296 (with Richard McAdams).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

143 Treu und Glauben: Frag GPT, in Susanne Auer-Maier, Michael Holoubek et al. (Hrsg): Treu und Glauben im Wirtschaftsrecht, Wien 2024, 17–36.
142 Experiments, in: Lee Epstein /Gunnar Grendstad /Urska Sadl /Keren Weinshall (eds.): The Oxford Handbook of Comparative Judicial Behaviour (2024) 265–284.
141 Wettbewerb und Gemeinwohl, in: Uwe Kischel /Hanno Kube (eds.): Handbuch des Staatsrechts I (2023), 755–789.
140 United, but Not Uniform. Experimental Evidence About Risk Taking in the Family, in: Pierpaolo Donati (Ed.): The Family as a Relational Good (Pontifical Academy of Social Sciences), Vatican 2023, 329–352.
139 Grundrechtskonjunkturen, in: Juristenzeitung 2022, 593–599.
138 Sunlight is the Best Disinfectant – Or is it? Anonymity as a Means to Enhance Impartiality, in: Arizona Law Review 63 (2021) 1063–1087 (with Eyal Zamir).
137 The Proper Scope of Behavioural Law and Economics, in: Stefan Grundmann /Philipp Hacker (eds.): Theories of Choice. The Social Science and the Law of Decision Making, Oxford 2021, 19–36.
136 Rechtswissenschaft als Sozialwissenschaft – Spurensuche im Jahrgang 2019 der ZGR, in: FS Windbichler, Berlin 2020, 33–50.
135 Maverick: Experimentally Testing a Conjecture of the Antitrust Authorities, in: Vincent Buskens, Rense Corten, Chris Snijders (eds.): Advances in the Sociology of Trust and Cooperation. Liber Amicorum Werner Raub, Berlin 2020, 357–390 (with Axel Ockenfels).
134 Does Efficiency Trump Legality? The Case of the German Constitutional Court, in: Yun-chien Chang (ed.): Selection and Decision in Judicial Process Around the World. Empirical Inquiries, Cambridge 2020, 261 – 286 (peer reviewed).
133 Organisationen als Akteure, in: Julian Krüper (Hrsg.): FS Morlok. Tübingen 2019, 295–302.
132 Experimental Criminal Law. A Survey of Contributions from Law, Economics, and Criminology, in: Willem H. van Boom /Pieter Desmet /Peter Mascini (eds.): Empirical Legal Research in Action. Cheltenham 2018, 57–108.
131 The Solidarity Motive, in: Stefano Zamagni /Marcelo Sanchez Sorondo (eds.): . Vatican 2018, 214–232, also in: Michael Faure /Wicher Schreuders /Louis Visscher (eds.): Don't Take it Seriously. Essays in Law and Economics in Honour of Roger Van den Bergh. Cambridge 2018, 459–478.
130 Das legitime Ziel in der Praxis des Bundesverfassungsgerichts. Eine quantitative Analyse der Entscheidungen des Jahres 2011, in: Matthias Jestaedt/Oliver Lepsius (Hrsg.): Verhältnismäßigkeit. Tübingen 2015, 97–128.
129 Privacy as a Public Good, in: Duke Law Journal 65 (2015) 385–457 (with Joshua Fairfield).
128 Behavioral Law and Economics: Empirical Methods, in: Eyal Zamir/Doron Teichman (eds.): The Oxford Handbook of Behavioral Economics and the Law. Oxford 2014, 125–142.
127 Aufgaben, in: Hanno Kube et al. (eds.): Leitgedanken des Rechts. FS Paul Kirchhof. Heidelberg 2013, 57–69.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

126 Legal Experiments: Mission Impossible?, Erasmus Law Lectures 28. The Hague (Eleven) 2013.
125 Öffentliches Wirtschaftsrecht aus der Sicht der ökonomischen Theorie, in: Dirk Ehlers / Michael Fehling / Hermann Pünder (eds.): Besonders Verwaltungsrecht I. Heidelberg 2012, 40–62.
124 Law as a Precondition for Religious Freedom, in: Fides et Libertas (2011) 18–35 and in: Mary Ann Glendon / Hans F. Zacher (eds.): Universal Rights in a World of Diversity. The Case of Religious Freedom. Vatican 2012, 393–412, as well as (in Russian translation) in право вечнце 323 (2015) 151–172.
123 Die Bedeutung der Verhaltensökonomie für das Kartellrecht, in: Holger Fleischer / Daniel Zimmer (eds.): Beitrag der Verhaltensökonomie (Behavioral Economics) zum Handels- und Wirtschaftsrecht (Beiheft der Zeitschrift für das gesamte Handelsrecht und Wirtschaftsrecht 75). Frankfurt 2011, 100–121.
122 An Experimental Contribution to the Revision of the Guidelines on Research and Development Agreements, in: FS Möschel. Baden-Baden 2011, 227–240.
121 Competition as a Socially Desirable Dilemma. Theory vs. Experimental Evidence, in: Josef Drexl / Wolfgang Kerber / Rupprecht Podszun (eds.): Competition Policy and the Economics Approach, Cheltenham 2011, 245–269.
120 Ernst-Joachim Mestmäcker, in: Stefan Grundmann / Karl Riesenhuber (eds.): Deutschsprachige Zivilrechtslehrer des 20. Jahrhunderts in Berichten ihrer Schüler, Berlin 2010, 53–69.
119 Das schwindende Vertrauen in die Marktwirtschaft und die Folgen für das Recht, in: FS Hopt. Berlin 2010, 2733–2752.
118 Behavioural Law and Economics im Urhebervertragsrecht – Ein Werkstattbericht, in: Karl Riesenhuber / Lars Klöhn (Hrsg.): Das Urhebervertragsrecht im Lichte der Verhaltensökonomik. Berlin 2010, 17–37; under the title "Recht im Labor – der Bestsellerparagraph im Urheberrecht" also in Michael Fehling / Klaus Grewlich (Hrsg.): Struktur und Wandel des Verwaltungsrechts. Symposium zum 80. Geburtstag von Martin Bullinger, Baden-Baden (Nomos) 2011, 9–30.
117 Preponderance of the Evidence Versus Intime Conviction. A Behavioral Perspective on a Conflict Between American and Continental European Law, in: Vermont Law Review 33 (2009) 435–467.
116 Die technische Seite der Verbreitung öffentlich-rechtlichen Rundfunks in digitalisierten Kabelnetzen, in: Zeitschrift für Urheber- und Medienrecht 52 (2008) 904–927 (with Jörn Lüdemann).
115 REITs ante portas – Die Anpassung des deutschen Rechts an institutionelle Investoren in den Grundstücks- und Mietmärken, in: Juristenzeitung 63 (2008) 1027–1031.
114 Geistiges Eigentum als Anreiz zur Innovation – Die Grenzen des Arguments, in: Martin Eifert / Wolfgang Hoffmann-Riem (Hrsg.): Geistiges Eigentum und Innovation, Berlin (Duncker) 2008, 43–72.
113 Institutional Design Capitalizing on the Intuitive Nature of Decision Making, in: Christoph Engel / Wolf Singer (eds.): Better than Conscious?, Cambridge (MIT Press) 2008, 413–432 (with Mark Lubell, Paul Glimcher, Reid Hastie, Jeffrey J.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Rachlinski, Bettina Rockenbach, Reinhard Selten, Tania Singer and Elke U. Weber).

112 Institutions for Intuitive Man, in: Christoph Engel /Wolf Singer (eds.): Better than Conscious?, Cambridge (MIT Press) 2008, 391–410.

111 Better than Conscious? The Brain, the Psyche, Behavior, and Institutions, in: Christoph Engel /Wolf Singer (eds.): Better than Conscious?, Cambridge (MIT Press) 2008, 1–19 (with Wolf Singer).

110 Governments in Dilemma. A Game Theoretic Model for the Conclusion of Bilateral Investment Treaties, in: University of Illinois Law Review 2008, 305–318.

109 Herrschaftsausübung bei offener Wirklichkeitsdefinition – Das Proprium des Rechts aus der Perspektive des öffentlichen Rechts, in: Christoph Engel/Wolfgang Schön (Hrsg.): Das Proprium der Rechtswissenschaft, Tübingen (Mohr) 2007, 205–240.

108 The Difficult Reception of Rigorous Descriptive Social Science in the Law, in: Nico Stehr /Bernd Weiler (eds.): Who Owns Knowledge? Knowledge and the Law, New Brunswick (Transaction Publishers) 2007, 169–213.

107 Innovationsanreize aus Wettbewerb und Kollusion, in: Ansgar Ohly /Diethelm Klippel (eds.): Geistiges Eigentum und Gemeinfreiheit, Tübingen (Mohr) 2007, 19–43.

106 The Psychological Case for Obliging Judges to Write Reasons, in: Christoph Engel/Fritz Strack (eds.): The Impact of Court Procedure on Judicial Decision Making, Baden-Baden (Nomos) 2007, 71–109.

105 [Behavioral Law and Economics or Law and Psychology?] Comment on Christine Jolls: Behavioral Law and Economics, in: Peter Diamond/Hannu Vartiainen (eds.): Behavioral Economics and Its Applications, Princeton (Princeton University Press) 2007, 148–155.

104 Verhaltenswissenschaftliche Analyse: Eine Gebrauchsanweisung für Juristen, in: Christoph Engel/Markus Englerth/Jörn Lüdemann/Indra Spiecker genannt Döhmann (Eds.): Recht und Verhalten. Beiträge zu Behavioural Law and Economics, Tübingen (Mohr) 2007, 363–405.

103 Freiheit und Autonomie, in: Detlef Merten/Hans-Jürgen Papier (eds.): Handbuch der Grundrechte in Deutschland und Europa II, Heidelberg 2006, 389–436.

102 Clumsy Conclusions: How to Do Policy and Research in a Complex World, in: Marco Verweij/Michael Thompson (eds.): Clumsy Solutions for a Complex World, Houndmills 2006, 1–27 (with Marco Verweij, Michael Thompson).

101 The Case for Clumsiness, in: Marco Verweij/Michael Thompson (eds.): Clumsy Solutions for a Complex World, Houndmills 2006, 1–27 (with Marco Verweij, Mary Douglas, Richard Ellis, Frank Hendriks, Susanne Lohmann, Steven Ney, Steve Rayner, Michael Thompson).

100 Öffentlichkeitsarbeit, in: Josef Isensee/Paul Kirchhof (eds.): Handbuch des Staatsrechts der Bundesrepublik Deutschland IV, Heidelberg 2006, 477–505. also published as: Ökonomische Analyse in der Rechtswissenschaft – Öffentlichkeitsarbeit der Regierung, in: Ulrich von Alemann/Martin Morlok/Julian Küper (eds.): (Partei-) Politik im Zeichen des Marketing, Baden-Baden 2010, 42–69.

99 Wettbewerb als sozial erwünschtes Dilemma, in: Christoph Engel/Wernhard

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Möschel (eds.): Recht und spontane Ordnung. Festschrift für Ernst-Joachim Mestmäcker zum achzigsten Geburtstag, Baden-Baden 2006, 155–198.

98 How Do Heuristics Mediate the Impact of Law on Behavior?, in: Gerd Gigerenzer/Christoph Engel (eds.): Heuristics and the Law, Cambridge, Mass. 2006, 439–465 (with Daniel G. Goldstein, Hal R. Arkes, Martin Beckenkamp, Robert Cooter, Robert C. Ellickson, Chris Guthrie, Ralph Hertwig, Stephanie Kurzenhäuser, Elke U. Weber).

97 Social Dilemmas, Revisited from a Heuristics Perspective, in: Gerd Gigerenzer/Christoph Engel (eds.): Heuristics and the Law, Cambridge, Mass. 2006, 61–85.

96 Law and Heuristics. An Interdisciplinary Venture, in: Gerd Gigerenzer/Christoph Engel (eds.): Heuristics and the Law, Cambridge, Mass. 2006, 1–16 (with Gerd Gigerenzer)

95 Inconsistency in the Law. In Search of a Balanced Norm, in: Christoph Engel/Lorraine Daston (eds.): Is There Value in Inconsistency?, Baden-Baden 2006, 221–281.

94 The Impossibility of Turning Arrow's Impossibility Theorem Epistemic, in: Christoph Engel/Lorraine Daston (eds.): Is There Value in Inconsistency?, Baden-Baden 2006, 129–140.

93 Can Inconsistency be a Value?, in: Christoph Engel/Lorraine Daston (eds.): Is There Value in Inconsistency?, Baden-Baden 2006, 9–16.

92 Zuschaueranteile in der publizistischen Konzentrationskontrolle. Grenzwert oder bloß ein Indiz unter vielen?, in: Zeitschrift für Urheber-und Medienrecht 49 (2005) 776–782.

91 Rationale Rechtspolitik und ihre Grenzen, in: Juristenzeitung 60 (2005) 581–590.

90 Marktabgrenzung als soziale Konstruktion, in: Andreas Fuchs /Hans-Peter Schwintowski /Daniel Zimmer (Hrsg.): Festschrift für Ulrich Immenga zum 70. Geburtstag, München 2004, 127–147.

89 A Constitutional Framework for Private Governance, in: German Law Journal 5 (2004) 197–236.

88 Meinungsvielfalt durch Ministererlaubnis ?, in: Zeitschrift für Wettbewerbsrecht 1 (2003) 448–480.

87 Der egalitäre Kern des Internet. Eine vernachlässigte Herausforderung für Steuerungstheorie und Steuerungspraxis, in: Karl-Heinz Ladeur (Hrsg.): Innovationsoffene Regulierung des Internet. Neues Recht für Kommunikationsnetzwerke (Schriften zur rechtswissenschaftlichen Innovationsforschung 7) Baden-Baden 2003, 25–38.

86 The Constitutional Court – Applying the Proportionality Principle – as a Subsidiary Authority for the Assessment of Political Outcomes, in: Christoph Engel / Adrienne Héritier (Hrsg.): Linking Politics and Law (Common Goods: Law, Politics and Economics 9) Baden-Baden 2003, 285–314.

85 Organizing Co-Existence in Cyberspace: Content Regulation and Privacy Compared, in: Christoph Engel /Adrienne Héritier (Hrsg.): Linking Politics and Law (Common Goods: Law, Politics and Economics 9) Baden-Baden 2003, 219–284.

84 Linking Political Science and Law, in: Christoph Engel /Adrienne Héritier

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

(Hrsg.): Linking Politics and Law (Common Goods: Law, Politics and Economics 9) Baden-Baden 2003, 7–22.

83 Imposed Liberty and its Limits. The EC Treaty as an Economic Constitution for the Member States, in: Talia Einhorn (ed.): Spontaneous Order, Organization and the Law. Roads to a European Civil Society. Liber Amicorum Ernst-Joachim Mestmäcker, The Hague (Asser Press) 2003, 429–437.

82 Die Internet-Service-Provider als Geiseln deutscher Ordnungsbehörden. Eine Kritik an den Verfügungen der Bezirksregierung Düsseldorf, in: Multimedia und Recht 6 (2003) Beilage 4/2003, 1–35.

81 Reihen aus Kinofilmen. Die Regeln des Rundfunkstaatsvertrags über die Unterbrechung zur Ausstrahlung von Werbung, in: Zeitschrift für Urheber- und Medienrecht 47 (2003) 85–94.

80 European Telecommunications Law: Unaffected by Globalization ?, in: Jürgen Basedow / Harald Baum / Klaus J. Hopt / Hideki Kanda / Toshiyuki Kono (Hrsg.): Economic Regulation and Competition. Regulation of Services in the EU, Germany and Japan (European Business Law & Practice Series 18) The Hague 2002, 219–251.

79 Rechtliche Entscheidungen unter Unsicherheit, in: Christoph Engel / Jost Halfmann / Martin Schulte (Hrsg.): Wissen – Nichtwissen – Unsicheres Wissen (Common Goods: Law, Politics and Economics 8) Baden-Baden 2002, 305–350.

78 Das legitime Ziel als Element des Übermaßverbots. Gemeinwohl als Frage der Verfassungsdogmatik, in: Winfried Brugger / Stephan Kirste / Michael Anderheiden (Hrsg.): Gemeinwohl in Deutschland, Europa und der Welt (Interdisziplinäre Studien zu Recht und Staat 24) Baden-Baden 2002, 103–172.

77 Globale Netze und lokale Werte, in: Archiv für Presserecht 33 (2002) 119–128.

76 Die soziale Funktion des Eigentums, in: Thomas von Danwitz / Otto Depenheuer / Christoph Engel: Bericht zur Lage des Eigentums, Heidelberg 2002, 9–107.

75 Offene Gemeinwohldefinitionen, in: Rechtstheorie 32 (2001) 23–52.

74 Institutionen für einen schwachen, den menschlichen Verstand, in: Max-Planck-Gesellschaft Jahrbuch 2001, 783–788. Also under the title "Krücken für den hinkenden Verstand" in Max Planck Forschung 2/2002, 15–18, English under the title "Crutches for our Hobbling Reason" in Max Planck Research 3/2002, 15–18.

73 Hybrid Governance Across National Jurisdictions as a Challenge to Constitutional Law, in: European Business Organization Law Review 2 (2001) 569–584 (Liber Amicorum Ernst Joachim Mestmäcker).

72 Die Grammatik des Rechts, in: Hans-Werner Rengeling / Hagen Hof (Hrsg.): Instrumente des Umweltschutzes im Wirkungsverbund, Baden-Baden 2001, 17–49.

71 Institutionen zwischen Staat und Markt, in: Die Verwaltung 34 (2001) 1–24.

70 Die Europäische Grundrechtscharta und die Presse, in: Zeitschrift für Urheber- und Medienrecht 44 (2000) 975–1006.

69 Die Kabelbetreiber als Mäzene wider Willen. Die Pflicht zur unentgeltlichen Einspeisung von Rundfunkprogrammen in die Kabelnetze nach § 19 LMG Baden-Württemberg 1999, in: RTkom 3 (2000) 189–210.

68 Arbeitsmarkt und staatliche Lenkung, in: Veröffentlichungen der Vereinigung der Deutschen Staatsrechtslehrer 59 (2000) 56–98.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

67 Sonderlasten für Fernsehveranstalter mit einem Zuschaueranteil von 10 %, in: Zeitschrift für Urheber- und Medienrecht 44 (2000) 345–356.
66 Das Internet und der Nationalstaat, in: Berichte der Deutschen Gesellschaft für Völkerrecht 39 (2000) 353–425. English version under the title: The Internet and the Nation State, in: Christoph Engel/Kenneth h. Keller (eds.): Understanding the Impact of Global Networks on Local Social, Political and Cultural Values (Law and Economics of International Telecommunications 42) Baden-Baden 2000, 201–260.
65 Abfallrecht und Abfallpolitik, in: Max-Planck-Gesellschaft Jahrbuch 1999, Göttingen 1999, 828–833. Slightly altered version: Der Müll, der Staat und sein Gesetz, in: Max-Planck-Forschung 4/1999, 50–55.
64 Selbstregulierung im Bereich der Produktverantwortung, in: Staatswissenschaften und Staatspraxis 9 (1998) [erschienen 1999] 535–591; auch in: Wolfgang Klett/Gerhard Schmitt-Gleser/Helmut Schnurer (Hrsg.): Deregulierung im Abfallrecht (Druckschrift zu den 7. Kölner Abfalltagen) Köln 1999, 227–300 [Titel dort von den Herausgebern verändert zu: Instrumente und deren inhaltliche Ausgestaltung (Selbstverpflichtungen, Zielfestlegungen, ökonomische Instrumente, Verordnungen)].
63 Europäisches Umweltrecht aus der Vogelperspektive. Das Handbuch zum europäischen und deutschen Umweltrecht, in: DVBl. 1999, 1069–1077.
62 Der Schutz der Verlagsrechte an Werken sowjetischer Komponisten, in: Erich Schulze/Dieter Meurer (Hrsg.): Urheberrechte, Urhebervertragsrecht und verwandte Schutzrechte. Gutachtensammlung 1958–1998, Marburg 1999, 215–249 (with Ernst-Joachim Mestmäcker).
61 Kommunikation und Medien, in: Manfred Dauses (Hrsg.): Handbuch des EU-Wirtschaftrechts, 7. Ergänzungslieferung 1999, E V, S. 1–81 (with Sebastian Seelmann-Eggebert).
60 Der Weg der deutschen Telekommunikation in den Wettbewerb, in: MMR 1999 Beilage 3, 7–14; englische Übersetzung unter dem Titel The Path to Competition for Telecommunications in Germany, in: J.Gregory Sidak/Christoph Engel/Günter Knieps (eds.): Competition and Regulation in Telecommunications. Examining Germany and America, Boston 2001, 15–39.
59 Internationaler Umweltschutz. Systematik der Probleme und der Lösungen, in: Festschrift Drobnig. Tübingen 1998, 247–276.
58 Die privatnützige Enteignung als Steuerungsinstrument, in: Die Verwaltung 1998, 543–558.
57 Rechtswissenschaft als angewandte Sozialwissenschaft. Die Aufgabe der Rechtswissenschaft nach der Öffnung der Rechtsordnung für sozialwissenschaftliche Theorie, in: ders. (Hrsg.): Methodische Zugänge zu einem Recht der Gemeinschaftsgüter (Common Goods: Law, Politics and Economics 1) Baden-Baden 1998, 11–40.
56 Die öffentliche Hand zwischen Innen- und Außensteuerung, in: Hans-Günther Henneke (Hrsg.): Organisation kommunaler Aufgabenerfüllung (Schriften zum deutschen und europäischen Kommunalrecht 6) Stuttgart 1998, 145–222.
55 Nebenwirkungen wirtschaftsrechtlicher Instrumente, in: Christoph Engel/

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Martin Morlok (Hrsg.): Öffentliches Recht als ein Gegenstand ökonomischer Forschung. Die Begegnung der deutschen Staatsrechtslehre mit der Konstitutionellen Politischen Ökonomie. Tübingen 1998, 173–206.

54 Das Recht der Gemeinschaftsgüter, in: Die Verwaltung 30 (1997) 429–479.

53 Die ansteckende Wirkung der ausländischen Liberalisierung von Telekommunikationsmärkten, in: JbNPÖ 16 (1997) 248–252.

52 Der Anspruch privater Kabelbetreiber auf ein Entgelt für die Durchleitung von Rundfunkprogrammen und das Medienrecht, in: Zeitschrift für Urheber- und Medienrecht 41 (1997) 497–520.

51 Verbreitung digitaler Pay-TV-Pakete in Fernsehkabelnetzen. Kartellrechtliche und medienrechtliche Überlegungen, in: Zeitschrift für Urheber- und Medienrecht 41 (1997) 309–330.

50 L'accès du publique et l'accès des opérateurs aux technologies avancées de communication, in: Fundamental Rights and New Information Technologies in the Audiovisual Sector (Schriftenreihe des Instituts für Europäisches Medienrecht 16) München 1996, 45–56.

49 Regulierung durch Organisation und Verfahren, in: Ulrich Immenga/Wernhard Möschel/Dieter Reuter (Hrsg.): FS Ernst-Joachim Mestmäcker. Baden-Baden 1996, 119–138.

48 Inhaltskontrolle im Internet, in: Archiv für Presserecht 1996, 220–227.

47 Europarechtliche Grenzen für die Industriepolitik, in: Hans-Werner Rengeling (Hrsg.): Europäisierung des Rechts (Osnabrücker rechtswissenschaftliche Abhandlungen 50) Köln 1996, 35–67.

46 Risikovorsorge im demokratischen Rechtsstaat, in: Die Verwaltung 1996, 265–271.

45 Legal Experiences in Competition among Institutions, in: Lüder Gerken (Ed.): Competition among Institutions (Hayek Symposium 1994) London 1995, 89–118.

44 Internationale öffentliche Unternehmen, in: Rabels Zeitschrift 1995, 495–544.

43 Die Messung der Fernsehnutzung als Voraussetzung eines Marktanteilsmodells zur Kontrolle der Medienkonzentration, in: Zeitschrift für Urheber- und Medienrecht 39 (1995), 1995, 653–666.

42 Anmerkung zu EuGH 09. 02. 1995 – Leclerc-Siplec, in: Juristenzeitung 1995, 940–942.

41 Multimedia und das deutsche Verfassungsrecht, in: Wolfgang Hoffmann-Riem/Thomas Vesting (Hrsg.): Perspektiven der Informationsgesellschaft (Symposien des Hans-Bredow-Instituts 16) Baden-Baden 1995, 155–171.

40 Wege zur Bewältigung der Konflikte in der globalen Informationsgesellschaft, in: Ernst-Joachim Mestmäcker (Hrsg.): Kommunikation ohne Monopole II. Ein Symposium über Ordnungsprinzipien im Wirtschaftsrecht der Telekommunikation und der elektronischen Medien (Law and Economics of International Telecommunications 23) Baden-Baden 1995, 179–210.

39 Zivilrecht als Fortsetzung des Wirtschaftsrechts mit anderen Mitteln. Rechtspolitische und verfassungsrechtliche Überlegungen am Beispiel des Haftungsrechts, in: Juristenzeitung 1995, 213–218.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

38 Werbung im öffentlich-rechtlichen Rundfunk nach 20 Uhr – verfassungsrechtliche und europarechtliche Einwände, in: Öffentlich-rechtlicher Rundfunk und Werbefinanzierung. Verfassungs-, medien- und ordnungsrechtliche Grenzen, Berlin 1995, 33–66.
37 Medienrechtliche Konzentrationsvorsorge, in: Die Sicherung der Meinungsvielfalt. Berichte, Gutachten und Vorschläge zur Fortentwicklung des Rechts der Medienkonzentrationskontrolle (Schriftenreihe der Landesmedienanstalten 4) Berlin 1995, 221–285.
36 Der Planungsvorbehalt für die Umwidmung von Frequenzen im Bestand öffentlich-rechtlicher Rundfunkanstalten, in: Zeitschrift für Urheber- und Medienrecht 1994, 536–543.
35 Rundfunk in Freiheit, in: Archiv für Presserecht 1994, 185–191.
34 "Effizienter Vertragsbruch" – eine juristische Antwort, in: Homo Oeconomicus 1994, 143–150.
33 Die Geltung des Brutto-Prinzips für die Unterbrechung von Spiel- und Fernsehfilmen durch Werbung, in: Zeitschrift für Urheber- und Medienrecht 38 (1994), 335–342.
32 Konzentrationsprobleme im Rundfunk und ihre rechtliche Bewältigung, in: Fusionskontrolle in der Sackgasse?/Pluralität im Wettbewerb (Rundfunkfreiheit contra Wettbewerbskontrolle) (Schriftenreihe des Instituts für Europäisches Medienrecht 8) München 1994, 86–96.
31 Die Bedeutung völkerrechtlicher Anerkennungen für das Internationale Privatrecht, in: Christoph Engel/Helmut Weber (Hrsg.): Festschrift für Dietrich Rothoeft zum 65. Geburtstag, München 1994, 87–96.
30 Einwirkungen des europäischen Menschenrechtsschutzes auf Meinungsäußerungsfreiheit und Pressefreiheit – insbesondere auf die Einführung von innerer Pressefreiheit, in: Archiv für Presserecht 1994, 1–9.
29 Vorsorge gegen die Konzentration im privaten Rundfunk mit den Mitteln des Rundfunkrechts – eine Analyse von § 21 Rundfunkstaatsvertrag 1991, in: Zeitschrift für Urheber- und Medienrecht 37 (1993), 557–585.
28 Versetzt die Europäische Menschenrechtskonvention der Sitztheorie des deutschen internationalen Gesellschaftsrechts den Todesstoß?, in: ZEuP 1993, 150–158.
27 Eigentumsschutz für Unternehmen, in: Archiv des öffentlichen Rechts 1993, 169–236.
26 Kommunikation und Medien, Beitrag zu Manfred Dauses (Hrsg.): Handbuch des EG-Wirtschaftsrechts 1993, E.V, S. 1–30.
25 Der Gesetzesvorbehalt für die Privatisierung öffentlich-rechtlichen Rundfunks, in: Zeitschrift für Urheber- und Medienrecht 37 (1993), 214–219.
24 Die Einwirkungen des Europäischen Gemeinschaftsrechts auf das deutsche Verwaltungsrecht, in: Die Verwaltung 1992, 437–476.
23 Telekommunikationsrecht. Grundstrukturen eines neuen Rechtsgebiets am Gegenstand des Europarechts, in: Zeitschrift für Urheber- und Medienrecht 36 (1992), 279–284.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

22 Is Trade in Services specific?, in: Thomas Oppermann / Josef Molsberger (Eds.): A new GATT for the Nineties and Europe '92, Baden-Baden 1991, 213–219.
21 Trade in Services between the European Communities and Third Countries – Its Regulation by Community Law, in: Daniel Friedmann / Ernst-Joachim Mestmäcker (Eds.): Rules for Free International Trade in Services. Symposium in Tel Aviv (Wirtschaftsrecht und Wirtschaftspolitik 107) Baden-Baden 1990, 107–175.
20 Der Ordnungsvorbehalt in den Schranken der Europäischen Menschenrechtskonvention, in: Schweizer Jahrbuch für internationales Recht 1989, 41–91.
19 Die Konkurrenz zwischen den deutschen öffentlich-rechtlichen Rundfunkanstalten und privaten Anbietern. Schlaglichter auf die Rechtsprechung des Bundesverfassungsgerichts aus dem Europäischen Gemeinschaftsrecht und aus der Europäischen Menschenrechtskonvention, in: Jörg Peter / Kay-Uwe Rhein (Hrsg.): Wirtschaft und Recht, Osnabrück 1989, 244–281.
18 Außenhandel mit Rundfunk: Rundfunkrichtlinie der Europäischen Gemeinschaft vs. Fernsehkonvention des Europarats, in: Rundfunk und Fernsehen 1989, 203–214.
17 Verfassungs-, Gesetzes- und Referendumsvorbehalt für Änderungen des Bundesgebiets und andere gebietsbezogene Akte, in: Archiv des öffentlichen Rechts 114, 1989, 46–112.
16 Ausstrahlungen der Europäischen Menschenrechtskonvention auf das Kollisionsrecht, in: Rabels Zeitschrift 1989, 3–51.
15 Offene Rundfunkordnung (Diskussionsbericht zum gleichnamigen Symposium), in: Ernst-Joachim Mestmäcker (Hrsg.): Offene Rundfunkordnung. Prinzipien für den Wettbewerb im grenzüberschreitenden Rundfunk. Gütersloh 1988, 505–556.
14 Der Zusammenhang von Sende- und Empfangsfreiheit unter der Europäischen Menschenrechtskonvention, in: Zeitschrift für Urheber- und Medienrecht 32 (1988), 511–526.
13 Die Bedeutung des Völkerrechts für die Anwendung in- und ausländischen Wirtschaftsrechts, in: Rabels Zeitschrift 1988, 271–302.
12 Europäische Konvention über grenzüberschreitendes Fernsehen. Ein Schritt in die falsche Richtung, in: Zeitschrift für Rechtspolitik 1988, 240–247.
11 Wirkung der völkerrechtlichen Normen über die Immunität im deutschen Rechtsraum, in: Zeitschrift für Vergleichende Rechtswissenschaft 1988, 33–45.
10 The position of public monopolies on broadcasting under the European Convention on Human Rights, in: Ernst-Joachim Mestmäcker (Ed.): The Law and Economics of Transborder Telecommunications. A Symposium (Law and Economics of International Telecommunications 1) Baden-Baden 1987, 55–70.
9 Die Schranken der Schranken der Europäischen Menschenrechtskonvention. Das Merkmal "notwendig in einer demokratischen Gesellschaft" in den Schrankenvorbehalten, das Diskriminierungsverbot, und die "margin of appreciation", in: Österreichische Zeitschrift für öffentliches Recht und Völkerrecht 1986, 261–287.
8 Die Folgen eines Verstoßes gegen § 114 Abs. 2 BSHG für das Verfahren vor dem Verwaltungsgericht, in: Zeitschrift für Sozialhilfe 1986, 317–321.
7 Abfindungsklauseln – Eine systematische Übersicht, in: Neue Juristische Wochenschrift 1986, 345–350.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

6 Das Völkerrecht des Telekommunikationsvorgangs, in: Rabels Zeitschrift 1985, 90–120.
5 Anmerkung zu Schleswig-Holsteinisches VG 30.10.1984, StAZ 1985, 79 (Namensänderung bei Stiefkindern), in: Standesamtszeitung 1985, 79–80.
4 Die französische Scheidung, Der praktische Fall aus dem Internationalen Privat- und Verfahrensrecht, in: Jura 1984, 381–390.
3 Vorläufige Maßnahmen gegen Täter von Verkehrsdelikten, die nicht im Besitz einer gültigen Fahrerlaubnis sind, in: Deutsches Autorecht 1984, 108–109.
2 Anmerkung zu LG Düsseldorf 10.3.1983, Juristenzeitung 1983, 625 (Fall Tabatabai), in: Juristenzeitung 1983, 627–629.
1 Der Film "Die Weiße Rose" und die Rechtslage bei Urteilen des Volksgerichtshofs, in: Juristenzeitung 1983, 192–195 (with Helmut Weber)

D. Special Issues and Edited Volumes

44 Crime: Analyzing Criminal Decision Making with the Tools of Economics and Criminology, in: Journal of Institutional and Theoretical Economics 181 (2025) 1–232 (with Dan Nagin).
43 Öffentliches Recht als Verhaltensordnung, Tübingen 2025 (with Lawrence O'Hara /Stefanie Egidy /Yoan Hermstrüwer /Leonard Hoeft /Pascal Langenbach).
42 Machine Learning and the Law, in: Journal of Institutional and Theoretical Economics 180 (2024) 211–399 (with Urs Schweizer).
41 Judicial Decision-Making, in: Journal of Institutional and Theoretical Economics 179 (2023) 1–249 (with Urs Schweizer).
40 Deliberate Ignorance. Choosing Not to Know, Cambridge (MIT Press) 2020 (with Ralph Hertwig).
39 Causality in the Law and in the Social Sciences, in: Journal of Institutional and Theoretical Economics 176 (2020) 1–213 (with Urs Schweizer).
38 Without Money, in: Journal of Institutional and Theoretical Economics 175 (2019) 1–205 (with Urs Schweizer).
37 Empirical Methods for the Law, in: Journal of Institutional and Theoretical Economics 174 (2018) 1–243 (with Urs Schweizer).
36 The Remedies Game, in: Journal of Institutional and Theoretical Economics 173 (2017) 1–207 (with Urs Schweizer).
35 Beyond Privity, in: Journal of Institutional and Theoretical Economics 172 (2016) 1–231 (with Urs Schweizer).
34 Does the Law Deliver?, in: Journal of Institutional and Theoretical Economics 171 (2015) 1–213 (with Urs Schweizer).
33 What Makes Intervention Legitimate?, in: Journal of Institutional and Theoretical Economics 170 (2014) 1–198 (with Urs Schweizer).
32 Behavioral Theory of Institutions, in: Journal of Institutioal and Theoretical Economics 169 (2013) 1–197 (with Urs Schweizer).

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

31 Testing Contracts, in: Journal of Institutional and Theoretical Economics 168 (2012) 1–208 (with Urs Schweizer).
30 Business-to-Consumer Transactions, in: Journal of Institutional and Theoretical Economics 167 (2011) 1–173 (with Urs Schweizer).
29 Gemeinwohl durch Wettbewerb?, Veröffentlichungen der Vereinigung der Deutschen Staatsrechtslehrer 69 (2010).
28 Jurimetrics, in: Journal of Institutional and Theoretical Economics 166 (2010) 1–203 (with Urs Schweizer).
27 Erosion von Verfassungsvoraussetzungen, Veröffentlichungen der Vereinigung der Deutschen Staatsrechtslehrer 68 (2009).
26 Coordination in the Absence of Sovereign Intervention, in: Journal of Institutional and Theoretical Economics 165 (2009) 1–192 (with Urs Schweizer).
25 Mechanism Design and the Law, in: Journal of Institutional and Theoretical Economics 164 (2008) 1–193 (with Urs Schweizer).
24 Better than Conscious?, Cambridge (MIT Press) 2008 (with Wolf Singer).
23 Public International Law and Economics, in: University of Illinois Law Review 2008, 1–436 (with Anne van Aaken and Tom Ginsburg).
22 Das Proprium der Rechtswissenschaft, Tübingen (Mohr) 2007 (with Wolfgang Schön).
21 The Impact of Court Procedure on Judicial Decision Making, Baden-Baden 2007 (with Fritz Strack).
20 Recht und Verhalten. Beiträge zu Behavioural Law and Economics, Tübingen 2007 (with Markus Englerth, Jörn Lüdemann and Indra Spiecker genannt Döhmann).
19 Experimental Law and Economics (24th International Seminar on the New Institutional Economics), in: Journal of Institutional and Theoretical Economics 163 (2007) 1–203 (with Urs Schweizer).
18 Recht und spontane Ordnung. Festschrift für Ernst-Joachim Mestmäcker zum achtzigsten Geburtstag, Baden-Baden 2006 (with Wernhard Möschel).
17 Heuristics and the Law (94th Dahlem Workshop), Cambridge, Mass. 2006 (with Gerd Gigerenzer).
16 Is There Value in Inconsistency? (Common Goods: Law, Politics and Economics 15) Baden-Baden 2006 (with Lorraine Daston).
15 Interacting with a Corporate Actor (23rd International Seminar on the New Institutional Economics), in: Journal of Institutional and Theoretical Economics 162 (2006) 1–225 (with Urs Schweizer).
14 Game Theory and the Law (22nd International Seminar on the New Institutional Economics), in: Journal of Institutional and Theoretical Economics 161 (2005) 191–355 (with Urs Schweizer).
13 The Generation and Distribution of Knowledge (21st International Seminar on the New Institutional Economics), in: Journal of Institutional and Theoretical Economics 160 (2004) 1–186 (with Urs Schweizer).
12 Linking Politics and Law (Common Goods: Law, Politics and Economics 9) Baden-Baden 2003 (with Adrienne Héritier).
11 Causes and Management of Conflicts (20th International Seminar on the New

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Institutional Economics), in: Journal of Institutional and Theoretical Economics 159 (2003) 1–258 (with Urs Schweizer).

10 Wissen – Nichtwissen – Unsicheres Wissen (Common Goods: Law, Politics and Economics 8) Baden-Baden 2002 (with Jost Halfmann und Martin Schulte).

9 Organizing and Designing Markets (19th International Seminar on the New Institutional Economics), in: Journal of Institutional and Theoretical Economics 158 (2002) 1–218 (with Urs Schweizer).

8 The Proper Scope of Government (18th International Seminar on the New Institutional Economics), in: Journal of Institutional and Theoretical Economics 157 (2001) 1–230 (with Urs Schweizer).

7 Competition and Regulation in Telecommunications. Examining Germany and America, Boston 2001 (with J. Gregory Sidak und Günter Knieps).

6 Governance of Global Networks in the Light of Differing Local Values (Law and Economics of International Telecommunications 43) Baden-Baden 2000 (with Kenneth H. Keller).

5 Understanding the Impact of Global Networks on Local Social, Political and Cultural Values (Law and Economics of International Telecommunications 42) Baden-Baden 2000 (with Kenneth H. Keller).

4 Methodische Zugänge zu einem Recht der Gemeinschaftsgüter (Common Goods: Law, Politics and Economics 1) Baden-Baden 1998.

3 Öffentliches Recht als ein Gegenstand ökonomischer Forschung. Die Begegnung der deutschen Staatsrechtslehre mit der Konstitutionellen Politischen Ökonomie. Tübingen 1998 (with Martin Morlok).

2 Festschrift für Dietrich Rothoeft zum 65. Geburtstag. München 1994. (with Helmut Weber).

1 Jahrbuch Junger Zivilrechtswissenschaftler 1 (1991) Stuttgart 1992. (with Harald Baum, Oliver Remien und Manfred Wenckstern).

E. Book Reviews

52 Mario Martini: Der Markt als Instrument hoheitlicher Verteilungslenkung. Möglichkeiten und Grenzen einer marktgesteuerten staatlichen Verteilung des Mangels, Tübingen 2008, in: Die Verwaltung 44 (2011) 131–133.

51 Jan Hecker: Marktoptimierende Wirtschaftsaufsicht. Öffentlich-rechtliche Probleme staatlicher Wirtschaftsinterventionen zur Steigerung der Funktionsfähigkeit des Marktes, Tübingen 2007, in: Archiv des öffentlichen Rechts 134 (2009) 151–153.

50 Jan Ole Püschel: Informationen des Staates als Wirtschaftsgut, Berlin 2006, in: Die Verwaltung 40 (2007) 298–300.

49 John Dewey: Human Nature and Conduct, New York 1922, in: Juristenzeitung 62 (2007) 84–85.

48 Otto Mayer: Deutsches Verwaltungsrecht, Berlin 1924, in: Juristenzeitung 61 (2006) 1168–1169.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

47 Julius Hermann von Kirchmann: Die Wertlosigkeit der Jurisprudenz als Wissenschaft, Berlin 1848 – Schlusswort, in: Juristenzeitung 61 (2006) 1118–1119.
46 Karl M. Meessen: Wirtschaftsrecht im Wettbewerb der Systeme, Tübingen 2005, in: Archiv des öffentlichen Rechts 131 (2006) 322–324.
45 Julius Hermann von Kirchmann: Die Wertlosigkeit der Jurisprudenz als Wissenschaft, Berlin 1848, in: Juristenzeitung 61 (2006) 614–614.
44 Todd Sandler: Global Collective Action, Cambridge 2004, International Studies Review 7 (2005) 328–330.
43 Paul W. Glimcher: Decisions, Uncertainty, and the Brain. The Science of Neuroeconomics, Cambridge, Mass. 2003, in: Rabels Zeitschrift 69 (2005) 193–198.
42 Gunnar Folke Schuppert: Staatswissenschaft, Baden-Baden 2003, in: Die Verwaltung 37 (2004) 567–569.
41 Adam Smith: The Theory of Moral Sentiments, Cambridge 1790, in: Juristenzeitung 2004, 667–668.
40 Ekkehart Schlicht: On Custom in the Economy, Oxford 1998, in: Rabels Zeitschrift 68 (2004) 419–423.
39 Mark Turner: Cognitive Dimensions of Social Science, Oxford 2001, in: Rabels Zeitschrift 68 (2004) 415–418.
38 Anne Peters: Elemente einer Theorie der Verfassung Europas, Berlin 2001, in: German Law Journal 4 (2003) 1375–1379.
37 Daniel Kahneman / Amos Tversky (eds.): Choices, Values, and Frames, Cambridge 2000, in: Rabels Zeitschrift 67 (2003) 781–786.
36 Johanna Hey: Steuerplanungssicherheit als Rechtsproblem, Köln 2002, in: Die Verwaltung 36 (2003) 409–412.
35 Jürgen G. Backhaus (ed.): The Elgar Companion to Law and Economics, Cheltenham 1999, in: Rabels Zeitschrift 67 (2003) 411–416.
34 Cass R. Sunstein (ed.): Behavioral Law and Economics, Cambridge 2000, in: Rabels Zeitschrift 67 (2003) 406–410.
33 Peter Axer: Normsetzung der Exekutive in der Sozialversicherung (Jus Publicum 49) Tübingen 2000, in: Verw. 34 (2001) 432–436 (unter dem Titel: Normsetzung der Exekutive in der Sozialversicherung).
32 Michael Kloepfer: Umweltrecht. München2 1998, in: Rabels Zeitschrift 65 (2001) 377–381.
31 Fritz Scharpf: Games Real Actors Play. Actor-centered institutionalism in Policy Research. Boulder, Colorado 1997, in: Rabels Zeitschrift 65 (2001) 192–197.
30 Gerd Gigerenzer / Peter Todd: Simple Heuristics that Make us Smart, New York 1999, in: Behavioral and Brain Sciences 23 (2000) 747–747 (unter dem Titel: Psychological Research on Heuristics Meets the Law).
29 Jack Knight: Institutionen und gesellschaftlicher Konflikt (Die Einheit der Gesellschaftswissenschaften 99) Tübingen 1997, in: Rabels Zeitschrift 64 (2000) 824–829.
28 Douglas G. Baird / Robert H. Gertner / Randal C. Picker: Game Theory and the Law. Cambridge (Mass.) 1995, in: Rabels Zeitschrift 64 (2000) 151–157.
27 Iris Bohnet: Kooperation und Kommunikation. Eine ökonomische Analyse in-

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

dividueller Entscheidungen (Die Einheit der Gesellschaftswissenschaften 98). Tübingen 1997, in: Rabels Zeitschrift 64 (2000) 146–151.

26 Monika Böhm: Der Normmensch. Materielle und prozedurale Aspekte des Schutzes des menschlichen Gesundheit vor Umweltschadstoffen (Jus Publicum 16). Tübingen 1996, in: Die Verwaltung 32 (1999) 121–123.

25 Gerhard Wegner: Wirtschaftspolitik zwischen Selbst- und Fremdsteuerung. Ein neuer Ansatz (Contributiones Jenenses 3). Baden-Baden 1995, in: Rabels Zeitschrift 63 (1999) 191–196.

24 Richard A. Epstein: Simple Rules for a Complex World. Cambridge 1995, in: Rabels Zeitschrift 62 (1998) 519–521.

23 Rudolf Richter / Eirik Furubotn: Neue Institutionenökonomik. Eine Einführung und kritische Würdigung. Tübingen 1996, in: Rabels Zeitschrift 62 (1998) 324–328.

22 Michael Kloepfer / Ekkehardt Mast: Das Umweltrecht des Auslandes (Schriften zum Umweltrecht 55) Berlin 1995, in: Rabels Zeitschrift 62 (1998) 187–190.

21 Wolfgang Fikentscher: Recht und wirtschaftliche Freiheit 1. Die Freiheit des Wettbewerbs, 2. Transnationales Marktrecht. Tübingen 1993, in: Rabels Zeitschrift 61(1997) 761–766.

20 Ulrich Büdenbender: Die Kartellaufsicht über die Energiewirtschaft. (Veröffentlichungen des Instituts für Energierecht an der Universität zu Köln 76) Baden-Baden 1995, in: AöR 122 (1997) 473–477.

19 Lothar Determann: Neue, gefahrverdächtige Technologien als Rechtsproblem. Beispiel: Mobilfunk-Sendeanlagen (Schriften zum Umweltrecht 68) Berlin 1996, in: DVBl. 1997, 1018.

18 Andreas Heinemann: Grenzen staatlicher Monopole im EG-Vertrag (Münchener Universitätsschriften. Reihe der Juristischen Fakultät 116) München 1996, in: ZHR 161 (1997) 561–564.

17 Franz Merli: Öffentliche Nutzungsrechte und Gemeingebrauch (Forschungen aus Staat und Recht 107) Wien 1995, in: Verw. 30 (1997) 421–424.

16 Geoffrey Brennan / James M. Buchanan: Die Begründung von Regeln. Konstitutionelle Politische Ökonomie (Die Einheit der Gesellschaftswissenschaften 83) Tübingen 1993, in: Rabels Z 1997, 363–368.

15 Martin Schulte: Schlichtes Verwaltungshandeln. verfassungs- und verwaltungsrechtsdogmatische Strukturüberlegungen am Beispiel des Umweltrechts (Jus Publicum 12) Tübingen 1995, in: JZ 1997, 401–402.

14 European Union – The Human Rights Challenge I–III. Baden-Baden 1991, in: Rabels Z 1996, 776–781.

13 Johannes Weisser: Corporate Opportunity. Zum Schutz der Geschäftschancen des Unternehmens im deutschen und US-amerikanischen Recht. Köln 1991, in: Rabels Z 1996, 578–583.

12 Gerrit Manssen: Privatrechtsgestaltung durch Hoheitsakt. Verfassungsrechtliche und verwaltungsrechtliche Grundfragen (Jus Publicum 9) Tübingen 1994, in: Die Verwaltung 1995, 411–415.

11 Rolf H. Weber: Vom Monopol zum Wettbewerb, Zürich 1995, in: AfP 1995, 446–447.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

10 Peter Schlechtriem (Hrsg.): Privatrecht und Wirtschaftsverfassung. Herstellung und Entwicklung des Privatrechts als Voraussetzung einer freiheitlichen Wirtschaftsverfassung in den ehemals sozialistischen Ländern (Schriften der Ernst von Caemmerer-Gedächtnis-Stiftung 2) Baden-Baden 1994, in: AöR 1995, 650–652.
9 Michael Bothe / Wolfgang Kilian: Rechtsfragen grenzüberschreitender Datenflüsse. Köln 1992, in: Rabels Z 1994, 605–609.
8 Ulrich Immenga / Ernst-Joachim Mestmäcker (Hrsg.): Gesetz gegen Wettbewerbsbeschränkungen. Kommentar. München 1992, in: Rabels Z 1994, 371–376.
7 Wolfgang Hoffmann-Riem: Rundfunkrecht neben Wirtschaftsrecht. Zur Anwendbarkeit des GWB und des EWG-V auf das Wettbewerbsverhalten öffentlichrechtlichen Rundfunks in der dualen Rundfunkordnung (Beiträge zum Rundfunkrecht 43) Baden-Baden 1991, in: Rabels Z 1994, 159–161.
6 Oliver E. Williamson: Die ökonomischen Institutionen des Kapitalismus. Unternehmen, Märkte, Kooperationen (Die Einheit der Gesellschaftswissenschaften 64) Tübingen 1990, in: Rabels Z 1993, 556–565.
5 Günther Bauer: Wettbewerbsbeschränkungen durch Staaten? Die Verantwortung der Mitgliedstaaten der Europäischen Gemeinschaft für den Wettbewerb im Gemeinsamen Markt (Europarecht, Völkerrecht 31) München 1990, in: Rabels Z 1992, 595–601.
4 Adelheid Puttler: Völkerrechtliche Grenzen von Export- und Reexportverboten. Eine Darstellung am Beispiel des Rechts der Vereinigten Staaten von Amerika und der Bundesrepublik Deutschland. Baden-Baden 1989, in: Rabels Z 1991, 398–400.
3 Andrew Clapham: Human Rights and the European Community. A Critical Overview – Antonio Cassese / Andrew Clapham / Joseph Weiler (eds.): Human Rights and the European Community. Methods of Protection – Antonio Cassese / Andrew Clapham / Joseph Weiler (eds.): The Substantive Law. Baden-Baden 1991, in: Rabels Z 1991, 178–181.
2 Alexander Georgieff: Kollisionen durch extraterritoriale staatliche Regelungen im internationalen Wirtschaftsrecht. Eine Untersuchung zu jüngsten Aspekten auf den Gebieten des Exportkontrollrechts und des Kartellrechts (FIW-Schriftenreihe 131). Köln 1989, in: ZHR 1990, 286–288.
1 Digest of Strasburg Case-Law relating to the European Convention of Human Rights I–VI, Köln 1984–1985 – P. van Dijk / G.J.H. van Hoof: Theory and Practice of the European Convention of Human Rights. Deventer 1984 – Jochen Abr. Frowein / Wolfgang Peukert: Europäische Menschenrechtskonvention. Kommentar. Kehl 1985 – Internationaler Kommentar zur Europäischen Menschenrechtskonvention, 1. Lieferung. Köln 1986, in: Rabels Z 1988, 402–411.

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

List of Authors

Prof. Dr. Elliott Ash
Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland

Prof. Dr. Oren Bar-Gill
School of Law, New York University, USA

Prof. Dr. Anna Baumert
School of Human and Social Sciences, University of Wuppertal, Germany

Prof. Dr. Stefan Bechtold
Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland

Prof. Dr. Florian Becker
Law School, Christian-Albrechts-University of Kiel, Germany

Prof. Dr. Morten Broberg
Faculty of Law, University of Copenhagen, Denmark

Dr. Renate Buijze
Erasmus School of Law, Erasmus University Rotterdam, The Netherlands

Prof. Dr. Yun-chien Chang
Law School, Cornell University, USA

Dr. Konstantin Chatziathanasiou
Faculty of Law, University of Münster, Germany

Dr. Marina Chugunova
Max Planck Institute for Innovation and Competition, Germany

Prof. Dr. Stefanie Egidy
Department of Law, University of Mannheim, Germany

Prof. Dr. Markus Englerth
Berlin School of Economics and Law, Germany

Prof. Dr. Susann Fiedler
Department of Strategy and Innovation, Vienna University of Economics and Business, Austria

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Dr. Hale A. Forster
Behavioral Science and Insights, Limetree, USA

Prof. Dr. Jens Frankenreiter
School of Law, Washington University in St. Louis, USA

Prof. Dr. Jonah B. Gelbach
School of Law, University of California at Berkeley, USA

Prof. Dr. Andreas Glöckner
Faculty of Human Sciences, University of Cologne, Germany

Prof. Dr. Sebastian J. Goerg
Nuremberg Institute for Market Decisions (NIM), Germany / Technical University of Munich, Germany

Prof. Dr. Werner Güth
Max Planck Institute for Behavioral Economics, Germany

Prof. Dr. Dr. Hanjo Hamann
Faculty of Law, EBS Universität für Wirtschaft und Recht, Germany

Prof. Dr. Klaus Heine
Erasmus School of Law, Erasmus University Rotterdam, The Netherlands

Prof. Dr. Eric Helland
Robert Day School of Economics and Finance, Claremont McKenna College, USA

Prof. Dr. Sigrid Hemels
Erasmus School of Law, Erasmus University Rotterdam, The Netherlands / Lund University School of Economics and Management, Sweden / Council of State, The Netherlands

Dr. Heike Hennig-Schmidt
Department of Economics, University of Bonn, Germany

Prof. Dr. Dr. Yoan Hermstrüwer
Law Faculty, University of Zurich, Switzerland

Prof. Dr. Svenja Hippel
Faculty of Law and Economics, University of Bonn, Germany

Prof. Dr. Daniel E. Ho
Law School, Stanford University, USA

Prof. Dr. Oliver Kirchkamp
Faculty of Economics and Business Administration, Friedrich-Schiller-University Jena, Germany

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Prof. Daniel M. Klerman
Gould School of Law, University of Southern California, USA

Prof. Dr. Hartmut Kliemt
Faculty of Economics and Business Studies, Justus-Liebig-University Gießen, Germany

Dr. Johannes Kruse
Max Planck Institute for Software Systems, Germany

Prof. Dr. Sebastian Kube
Department of Economics, University of Bonn, Germany

Prof. Dr. Michael Kurschilgen
Faculty of Business and Economics, UniDistance Suisse, Switzerland

Dr. Dr. Pascal Langenbach
Faculty of Law and Economics, University of Bonn, Germany

Prof. Dr. Chang-Ching Lin
Department of Economics, National Cheng Kung University, Taiwan

Prof. Michael A. Livermore
School of Law, University of Virginia, USA

Prof. Dr. Jörn Lüdemann
Faculty of Law and Interdisciplinary Faculty, University of Rostock, Germany

Helena Lyng-Olsen
Law School, Stanford University, USA

Prof. Florencia Marotta-Wurgler
School of Law, New York University, USA

Dr. Dorothee Mischkowski
Faculty of Social and Behavioural Sciences, Universiteit Leiden, The Netherlands

Prof. Dr. Luigi Mittone
Department of Economics and Management, University of Trento, Italy

Prof. Dr. Peter Moffatt
School of Economics, University of East Anglia, United Kingdom

Prof. Dr. Dr. Alexander Morell
Faculty of Law, Goethe University Frankfurt, Germany

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Prof. Dr. Azzurra Morreale
Business School, LUT University, Finland

Dr. Claudia Möllers
Faculty of Business Administration and Economics, Heinrich-Heine-University Düsseldorf, Germany

Prof. Dr. Daniel S. Nagin
Heinz College of Information Systems and Public Policy, Carnegie Mellon University, USA

Prof. Dr. Andreas Nicklisch
Center for Economic Policy Research, University of Applied Sciences of the Grisons, Switzerland

Prof. Dr. Hans-Theo Normann
Düsseldorf Institute for Competition Economics (DICE), Heinrich-Heine-University Düsseldorf, Germany

Prof. Dr. Axel Ockenfels
Adenauer School of Government and Economics Department, University of Cologne, Germany /Max Planck Institute for Behavioral Economics, Germany

Dr. Laurence O'Hara
Institute of Public Law, Goethe University Frankfurt, Germany

Prof. Dr. Stefan Okruch
Andrássy University Budapest, Hungary

Prof. Dr. Niels Petersen
Faculty of Law, University of Münster, Germany

Prof. Dr. Jeffrey J. Rachlinski
Cornell Law School, USA

Prof. Dr. Crystal Reeck
Fox School of Business and Management, Temple University, USA

Prof. Dr. Julia Sasse
Media Faculty, Ansbach University of Applied Sciences, Germany

Prof. Dr. Kai-Uwe Schnapp
Faculty of Business, Economics and Social Sciences, University of Hamburg, Germany

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

Prof. Dr. Indra Spiecker genannt Döhmann
Faculty of Law, University of Cologne, Germany

Prof. Dr. Dr. Armin Steinbach
Federal Ministry of Finance, Germany /HEC Paris, France

Prof. Dr. Alexander Stremitzer
Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland

Prof. Dr. Emanuel V. Towfigh
Faculty of Law, EBS Universität für Wirtschaft und Recht, Germany

Prof. Dr. Christian Traxler
Hertie School, Germany

Prof. Dr. Anne van Aaken
Faculty of Law, Director Institute of Law and Economics, University of Hamburg, Germany

Prof. Dr. Carl Christian von Weizsäcker
Max Planck Institute for Behavioral Economics, Germany

Prof. Dr. Gari Walkowitz
Faculty of Business Administration, Technical University Bergakademie Freiberg, Germany

Dr. Lilia Wasserka-Zhurakhovska
J.E. Cairnes School of Business & Economics, University of Galway, Ireland

Prof. Dr. Elke U. Weber
Andlinger Center for Energy + Environment, School of Public and International Affairs, and Department of Psychology, Princeton University, USA

Prof. Dr. Keren Weinshall
Faculty of Law, Hebrew University of Jerusalem, Israel

Patrick C. Wu
University of California at Irvine, USA

Prof. Dr. Eyal Zamir
Faculty of Law, Hebrew University of Jerusalem, Israel

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -

T.T.-Prof. Dr. Frederike Zufall
Department of Informatics, Karlsruhe Institute of Technology, Germany
/Waseda Institute for Advanced Study, Waseda University, Tokyo

https://doi.org/10.5771/9783748947257 https://www.inlibra.com/de/agb - Open Access -