

# Experimentelle Ökonomik im Recht

Konstantin Chatziathanasiou und Monika Leszczyńska\*

I.	Einleitung .....	314
II.	Experimentelle Ökonomik .....	317
1.	Anliegen der experimentellen Ökonomik .....	317
2.	Experimentalökonomische Labore .....	318
3.	Beispiel: Ultimatumspiel .....	319
III.	Rechtlich relevante Forschungsfra- gen .....	320
1.	Rechtsgestaltung .....	321
2.	a. Probleme beschreiben .....	321
	b. Interventionsbedarf einschät- zen .....	323
	c. Lösungen evaluieren .....	324
2.	Rechtsanwendung .....	325
IV.	Theorie und Hypothese .....	326
1.	Ökonomische Theorie .....	326
	a. Homo oeconomicus .....	326
	b. Verhaltensökonomik .....	327
	c. Zum Verhältnis von Stan- dardtheorie und Verhaltens- ökonomik .....	329
2.	Hypothesenbildung .....	329
V.	Versuchsanordnung .....	330
1.	Manipulation unter kontrollier- ten Bedingungen .....	331
2.	Rolle des Zufalls .....	332
3.	Treatment an denselben Ver- suchspersonen? .....	333
VI.	Statistische Analyse der Ergeb- nisse .....	333
1.	Effektgröße .....	334
2.	Deskriptive und Inferenz-Statis- tik .....	334
3.	Nullhypothese und Alternativ- hypothese .....	335
4.	Interpretation des p-Werts .....	335
VII.	Diskussion von Resultaten .....	336
1.	Studenten als Teilnehmer und abstrakte Umgebung .....	336
2.	Mehrere Studien und juristisches Augenmaß .....	337
VIII.	Fazit .....	338

*Der Beitrag erläutert aus juristischer Sicht systematisch Anliegen und Methode der experimentellen Ökonomik. Ökonomische Laborexperimente bieten ein For- schungsinstrument, das Einsichten über menschliches Verhalten verspricht. Sie fin- den in einer abstrakten Umgebung statt, um eindeutige Kausalschlüsse über das Geschehen im Labor zu ermöglichen. Der Beitrag schafft eine Ausgangsbasis für die Beurteilung ihrer Aussagekraft. Zur Illustration werden aktuelle Studien heran- gezogen, die für das Recht von Interesse sind.*

## I. Einleitung

Die ökonomische Theorie hat Eingang in die deutsche Rechtswissenschaft gefun- den.<sup>1</sup> Auch ihre verhaltensökonomische Qualifizierung (*behavioral economics*)

\* Konstantin Chatziathanasiou ist Research Fellow am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern (Bonn); Monika Leszczyńska ist Visiting Research Fellow am selben Max-Planck-Institut und LL.M.-Kandidatin (Hauser Global Scholar) an der NYU School of Law. Wir danken Stefanie Egidy, Hanjo Hamann, Rafael Harnos, Rima-Maria Rahal und einem anonymen Gutachter für sehr hilfreiche Anmerkungen. Fehler gehen zu unseren Lasten.

1 Klassisch H.-B. Schäfer/C. Ott, Lehrbuch der ökonomischen Analyse des Zivilrechts, Berlin 2013; S. Magen, Konjunkturen der Rechtsökonomie als öffentlich-rechtlicher Grundlagenforschung, in: A. Funke/J. Krüper/J. Lüdemann (Hg.), Konjunkturen öffentlich-rechtlicher Grundlagenforschung, Tübingen 2015, S. 103 (108), diagnostiziert „zunehmend [...] auch im öffentlichen Recht ein wachsen- des genuin wissenschaftliches Interesse an rechtsökonomischer Forschung“; zur Rezeption in Straf- recht und Kriminologie M. Englerth, Der beschränkt rationale Verbrecher: Behavioral Economics in der Kriminologie, Münster 2010.

wird von Juristen mit Interesse wahrgenommen und rezipiert.<sup>2</sup> Die Verhaltensökonomik untersucht systematisch das Entscheidungsverhalten von Menschen in verschiedenen Situationen und versucht dabei der Tatsache gerecht zu werden, dass viele Annahmen des sog. Standardmodells der neoklassischen ökonomischen Theorie (*homo oeconomicus*) einer experimentellen Überprüfung nicht standhalten.<sup>3</sup> Juristen betrachten diese Entwicklungen in der Ökonomik mit Sympathie.<sup>4</sup> Gleichwohl begegnen auch diejenigen Juristen, die der Verhaltensökonomik grundsätzlich offen gegenüberstehen, der experimentellen Methodik mit einer gewissen Skepsis. So fragen *Fleischer, Schmolke* und *Zimmer* in ihrem Beitrag zur Verhaltensökonomik als Forschungsinstrument im Wirtschaftsrecht: „Welche Aussagekraft soll man Laborexperimenten beimessen, die auf weit gehender Dekontextualisierung beruhen, während juristische Entscheidungen immer kontextgebunden sind?“<sup>5</sup>

Aus juristischer Sicht liegt hier das Kernproblem für die Rezeption experimenteller Resultate.<sup>6</sup> Es geht um die Frage der externen Validität von Laborexperimenten: Inwiefern kann ein Experiment das Geschehen außerhalb des Labors vorhersagen? Das Problem der externen Validität darf nicht unterschätzt werden. Das Bewusstsein dafür sollte den Blick auf experimentelle Forschung und ihren Erkenntniswert aber auch nicht verstellen.<sup>7</sup> Der beste Weg, Über-Vorsicht zu vermeiden, liegt unseres Erachtens darin, in einem ersten Schritt ein Verständnis für die Eigenheiten und Bedingungen dieser Forschungsrichtung zu gewinnen. In jüngerer Zeit gehen Juris-

2 Repräsentativ die Beiträge in H. Fleischer/D. Zimmer (Hg.), Beitrag der Verhaltensökonomie (Behavioral Economics) zum Handels- und Wirtschaftsrecht, Frankfurt am Main 2011; programmatisch C. Jolls/C. R. Sunstein/R. Thaler, A Behavioral Approach to Law and Economics, Stanford Law Rev. (50) 1998, S. 1471.

3 Einführend M. Englerth, Verhaltensökonomie, in: E. V. Towfigh/N. Petersen, Ökonomische Methoden im Recht: eine Einführung für Juristen, Tübingen 2010, S. 165.

4 H. Eidenmüller, Der *homo oeconomicus* und das Schuldrecht: Herausforderungen durch Behavioral Law and Economics, JZ 2005, S. 216 (224), schlussfolgert, „dass Behavioral Law and Economics nicht nur den *homo oeconomicus* menschlicher, sondern auch den ökonomischen Effizienzkalkül „rechtsnäher“ gemacht hat. Der Akzeptanz dieser Forschungsrichtung unter Juristen kann dies, so ist zu hoffen, nur förderlich sein.“

5 H. Fleischer/K. U. Schmolke/D. Zimmer, Verhaltensökonomik als Forschungsinstrument für das Wirtschaftsrecht, in: H. Fleischer/D. Zimmer (Hg.), Beitrag der Verhaltensökonomie (Behavioral Economics) zum Handels- und Wirtschaftsrecht, Frankfurt am Main 2011, S. 1 (45 f.).

6 Statt vieler J. Wolff, Eine Annäherung an das Nudge-Konzept nach Richard H. Thaler und Cass R. Sunstein aus rechtswissenschaftlicher Sicht, RW 2015, S. 194 (221).

7 So auch C. Engel, Verhaltenswissenschaftliche Analyse: Eine Gebrauchsanweisung für Juristen, in: C. Engel et al. (Hg.), Recht und Verhalten: Beiträge zu Behavioral Law and Economics, Tübingen 2007, S. 363 (397): „Warnschilder können eine unerwünschte Wirkung haben.“; siehe auch die allgemeine Warnung bei A. Voßkuhle/T. Wischmeyer, The ‘Neue Verwaltungsrechtswissenschaft’ against the backdrop of traditional administrative law scholarship in Germany, in: P. Lindseth/S. Rose-Ackerman (Hg.), Comparative Administrative Law, Cheltenham u.a. 2017, im Erscheinen: „a naïve or uninformed “import” of theory etc. can be as dangerous as the refusal to engage with interdisciplinary insights“.

ten sogar einen Schritt weiter und betreiben selbst experimentelle Forschung.<sup>8</sup> Dies ist nicht ohne Risiko,<sup>9</sup> kann sich aber lohnen. Denn indem Juristen selbst das Experiment gestalten, wird das Problem der externen Validität frontal angegangen.

Neben der externen Validität stellt sich für Juristen ein weiteres Problem: viele Fragen, die für Rechtswissenschaftler von Interesse sind, werden in ökonomischen Experimenten nicht behandelt.<sup>10</sup> Juristen können hierauf durch die bereits genannte eigene Beteiligung an experimenteller Forschung reagieren. Sie können aber auch mit Experimentalökonomen in Dialog treten<sup>11</sup> und Denkanstöße an diese richten. Auch hierfür bedarf es aber eines Grundverständnisses der experimentellen Methoden.

Dieser Beitrag soll für die Beurteilung externer Validität und die Kommunikation mit Experimentatoren eine Ausgangsbasis schaffen. Es soll eine Gebrauchsanweisung<sup>12</sup> geboten werden, mit der experimentalökonomische Forschung leichter eingeordnet und bewertet werden kann.<sup>13</sup> Wir konzentrieren uns dabei auf die wichtigsten Aspekte der Methodik experimentalökonomischer Laborexperimente.<sup>14</sup>

8 Jüngere Beispiele mit zumindest einem juristischen Co-Autor: A. R. Dorrough/M. Leszczyńska/M. Barreto/A. Glöckner, Revealing side effects of quota rules on group cooperation, *Journal of Economic Psychology* (57) 2016, S. 136; C. Engel/L. Zhurakowska, You are in Charge: Experimentally Testing the Motivating Power of Holding a Judicial Office, *The Journal of Legal Studies* (46) 2017, S. 1; Y. Hermstrüwer/S. Dickert, Sharing is daring: An experiment on consent, chilling effects and a salient privacy nudge, *International Review of Law and Economics* (51) 2017, S. 38; S. Hoeppner/L. Lyhs, Behavior Under Vague Standards: Evidence from the Laboratory, Friedrich-Schiller-University Jena 2016; M. Kleine/P. Langenbach/L. Zhurakowska, Fairness and Persuasion, *Economics Letters* (171) 2016, S. 173; A. Morell/A. Glöckner/E. V. Towfigh, Sticky Rebates: Loyalty Rebates Impede Rational Switching of Consumers, *Journal of Competition Law & Economics* (11) 2015, S. 431; K. U. Schmolke/V. Utikal, Whistleblowing: Incentives and Situational Determinants, FAU – Discussion Papers in Economics No. 09, 2016. H. Spämann/L. Klöhn, Justice Is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges, *The Journal of Legal Studies* (45) 2016, S. 255.

9 Vgl. D. Adams, *Mostly Harmless*, London 2009, S. 142: „It can be very dangerous to see things from somebody else's point of view without the proper training.“; in diese Richtung auch H. Hammann, Evidenzbasierte Jurisprudenz, Methoden empirischer Forschung und ihr Erkenntniswert für das Recht am Beispiel des Gesellschaftsrechts, Tübingen 2014, S. 27f., der – neben fehlenden Anreizen – den Mangel einer Methodenausbildung als strukturelles Hindernis empirischer Forschung durch Juristen ausmacht.

10 Engel, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363 (374): „Juristen [wünschen sich] häufig verhaltenswissenschaftliche Befunde, die noch niemand erhoben hat.“

11 Für mehr *personelle* Interdisziplinarität plädiert Magen, Konjunkturen (Fn. 1) S. 103 (124).

12 Begriff bei Engel, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363; auch L. Klöhn, Der Beitrag der Verhaltensökonomik zum Kapitalmarktrecht, in: H. Fleischer/D. Zimmer (Hg.), Beitrag der Verhaltensökonomie (Behavioral Economics) zum Handels- und Wirtschaftsrecht, Frankfurt am Main 2011, S. 83 (98).

13 Ähnliches Anliegen bei K. van den Bos/L. Hulst, On Experiments in Empirical Legal Research, *Law and Method* 2016, S. 1; in Bezug auf die Ökonomik allgemein E. V. Towfigh/N. Petersen, Ökonomische Methoden im Recht: eine Einführung für Juristen, Tübingen 2010; aus Sicht der Politikwissenschaft I. Kubbe, Experimente in der Politikwissenschaft: Eine methodische Einführung, Wiesbaden 2016.

14 Allgemeine Darstellung experimenteller Methodik bei P. Sedlmeier/F. Renkewitz, *Forschungsmethoden und Statistik in der Psychologie*, München 2008; zur experimentellen Ökonomik C. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton u.a. 2003.

Neben diesen gibt es auch andere empirische Forschungsinstrumente.<sup>15</sup> Zum Beispiel können Feldexperimente, also Experimente, die in einer natürlichen – oder „realen“ – Umgebung stattfinden, Laborexperimente sehr sinnvoll ergänzen.<sup>16</sup> Die experimentelle Ökonomik unterscheidet sich in einer Reihe methodologischer Eigenheiten auch von der psychologischen Forschung, etwa in der starken Fokussierung auf monetäre Anreize.<sup>17</sup> Neben diesen Eingrenzungen sei erwähnt, dass wir hier keine Übersicht über den Forschungsstand anbieten<sup>18</sup> und auch keine Klärung der philosophischen Fragen, die mit experimenteller Sozialwissenschaft verbunden sind.<sup>19</sup>

Nach einer allgemeinen Einführung zur experimentellen Ökonomik (II.) skizzieren wir anhand verschiedener Beispiele einen Überblick über rechtlich relevante Forschungsfragen, denen mit Experimenten nachgegangen werden kann (III.). Danach gehen wir näher darauf ein, wie in einer experimentellen Studie theoretisch hergeleitete Hypothesen (IV.) getestet (V.) und Resultate beschrieben (VI.) und diskutiert (VII.) werden.<sup>20</sup>

## **II. Experimentelle Ökonomik**

Gleich zu Beginn möchten wir einen Eindruck davon vermitteln, was überhaupt unter ökonomischen Experimenten und Laboren zu verstehen ist.

### **1. Anliegen der experimentellen Ökonomik**

Die Ökonomik widmet sich allgemein der Frage, wie Menschen in Anbetracht knapper Ressourcen auf Anreize reagieren.<sup>21</sup> Die Forschungsfragen sind dabei nicht auf wirtschaftliche Zusammenhänge beschränkt.<sup>22</sup> Ökonomische Methoden

15 Zu anderen quantitativen empirischen Methoden wie der Analyse von Felddaten etwa *F. L. Leeuw/H. Schmeets*, Empirical Legal Research, Cheltenham u.a. 2016; zum „Spektrum der Methoden“ *Hamann*, Evidenzbasierte Jurisprudenz (Fn. 9), S. 171 ff.

16 Zu Vor- und Nachteilen von Feldexperimenten siehe etwa *Sedlmeier/Renkewitz*, Forschungsmethoden (Fn. 14), S. 149.

17 Zu den Unterschieden *R. Hertwig/A. Ortmann*, Experimental practices in economics: A methodological challenge for psychologists?, Behavioral and Brain Sciences (24) 2001, S. 383; kritisch *F. Guala*, The Methodology of Experimental Economics, Cambridge u.a. 2005, S. 231 ff.

18 Hierzu etwa *J. H. Arlen/E. Talley* (Hg.), Experimental Law and Economics, Cheltenham 2008; *C. Camerer/E. Talley*, Experimental Study of Law, in: *Polinsky/Shavell* (Hg.), Handbook of Law and Economics 2007, S. 1619; *C. Engel*, Experimental Criminal Law: A Survey of Contributions from Law, Economics and Criminology, Preprints of the Max Planck Institute for Research on Collective Goods 2016; *S. Hoepfner*, Experimental Law and Economics, in: *J.G. Backhaus*, Encyclopedia of Law and Economics, New York 2014.

19 Eindringlich *A. Rosenberg*, Philosophy of Social Science, Boulder 2015, S. 11 ff.

20 Diese Struktur ist von der üblichen Darstellung einer experimentellen Studie inspiriert.

21 Statt vieler *Rosenberg*, Philosophy of Social Science (Fn. 19), S. 95.

22 Zum Vorwurf, es handele sich um „Imperialismus“, siehe *G. Kirchgässner*, Homo Oeconomicus, Tübingen 2013, S. 152 ff.

werden etwa auch bei der Analyse kriminellen Verhaltens<sup>23</sup> oder sogar zur Untersuchung von familienrechtlichen Regelungen fruchtbar gemacht.<sup>24</sup>

Die Experimentalökonomik geht historisch auf das Anliegen zurück, ökonomische Theorie zu testen und weiterzuentwickeln.<sup>25</sup> Dieser historische Anwendungsbe- reich ist mittlerweile erweitert worden. Denn nicht in allen Experimenten geht es im strengen Sinne darum, eine Theorie zu testen. Ein beträchtlicher Teil von Experiments zielt darauf ab, eine bestimmte soziale Entscheidungssituation nachzubilden.<sup>26</sup> Dann wird untersucht, wie sich eine gezielte Veränderung der Umwelt in dieser Entscheidungssituation auswirkt.<sup>27</sup>

Für die experimentelle Rechtsökonomik sind beide Dimensionen relevant. Da die Rechtsökonomik auf ökonomischer Theorie basiert, interessiert sie sich auch für deren Test und Weiterentwicklung. Da aber gerade das Recht auf die Steuerung komplexer sozialer Situationen abzielt, ist der zweite Aspekt von besonderer Bedeutung.

## 2. Experimentalökonomische Labore

Viele ökonomische Fakultäten und Institute unterhalten sogenannte Labore, in denen die Teilnehmer eines Experiments Geld verdienen können.<sup>28</sup> Ein Labor ist meist ein umfunktionierter Seminarraum, in dem die Versuchspersonen in Einzelkabinen an einem Computer Entscheidungen treffen. Dabei schützt oftmals ein Vorhang die Privatsphäre der Teilnehmer, die auch ansonsten während des Experiments anonym bleiben.

23 Englerth, Der beschränkt rationale Verbrecher (Fn. 1).

24 Zur ökonomischen Analyse des Scheidungsfolge- und Ehevertragsrechts K. U. Schmolke, Grenzen der Selbstbindung im Privatrecht: Rechtspaternalismus und Verhaltensökonomik im Familien-, Ge- sellschafts- und Verbraucherrecht, Tübingen 2014, S. 369 ff.

25 Pionierarbeit bei H. Sauermaier/R. Seltene, Ein Oligopolexperiment, Zeitschrift für die gesamte Staatswissenschaft (115) 1959, S. 427.

26 Eingehend zu dieser zweiten Kategorie K. M. Schmidt, The Role of Experiments for the Development of Economic Theories, Perspektiven der Wirtschaftspolitik (10) 2009, S. 14 (15): „Even though many experimental papers have a section in which they discuss whether the results of the experiment are consistent with standard theory or with a subset of alternative models they cannot be considered a rigorous test of any theory. For this the experiments are too complicated and leave too much room for different interpretations. These experiments have a different objective. They try to capture some essential aspects of a complex (social) decision process in the lab. Sometimes they uncover new phenomena, i.e. regularities in behavior that are typical in certain situations, and examine how robust these phenomena are.“

27 Schmidt, The Role of Experiments (Fn. 26), S. 14 (15) spricht von „materiellen Modellen“: „A “material” model is a model of flesh and blood the exogenous variables of which are controlled by the experimental design in order to see how the endogenous variables react to changes in the treatment variables.“

28 Für Photos des BonnEconLab: <http://www.bonneconlab.uni-bonn.de/ueber.uns/fotos-des-labors> (abgerufen am 12.8.2016).

Typischerweise<sup>29</sup> verteilen die Versuchsleiter die Instruktionen des Experiments in gedruckter Form, verlesen diese laut und stehen für Rückfragen bereit. Das richtige Verständnis wird durch Kontrollfragen geprüft. Dann trifft jeder Teilnehmer auf seinem Bildschirm Entscheidungen, die sich auf seine Auszahlung nach Abschluss des Experiments auswirken. Was die Entscheidungen für die Auszahlung bedeuten können, ist den Teilnehmern bekannt.<sup>30</sup> Die Auszahlung findet jeweils einzeln statt, ohne dass die übrigen Teilnehmer erfahren, welcher Versuchsteilnehmer im Experiment welche Rolle innehatte und wieviel er in dieser Rolle verdient hat. Die Vorgänge in einem ökonomischen Experiment sind also nicht hypothetisch: Die Teilnehmer kommen, um Geld zu verdienen, und ihre Entlohnung hängt von ihren Entscheidungen ab.<sup>31</sup>

### 3. Beispiel: Ultimatumspiel

Vielzitiertes<sup>32</sup> Beispiel für eine experimentelle Entscheidungsaufgabe ist das Ultimatumspiel *Werner Güths*.<sup>33</sup> Hier haben die Teilnehmer entweder die Rolle des Senders oder des Empfängers. Der Sender entscheidet, welchen Anteil von zum Beispiel 10 € er dem Empfänger anbieten möchte. Der Empfänger wird über die Entscheidung des Senders informiert und entscheidet sich, ob er das Angebot annimmt oder ablehnt. Nimmt er an, erhält er den angebotenen Betrag und der Sender den Restbetrag. Lehnt er ab, erhalten beide Spieler nichts. Das durchschnittliche Angebot liegt bei 40 bis 50 % des Gesamtbetrages.<sup>34</sup> Besonders interessant – und für Ökonomen überraschend<sup>35</sup> – ist aber das Verhalten der Empfänger: Diese lehnen Angebote ab, die ihnen zu niedrig – oder „unfair“ – erscheinen, obwohl dies bedeutet, dass sie mit leeren Händen dastehen.<sup>36</sup> Die Frage nach der Psycholo-

29 Ein Wort der Vorsicht: Zu allen aufgezählten Eigenschaften gibt es Ausnahmen. Auch in der Experimentalökonomik gibt es unterschiedliche Ansichten über das „richtige“ Vorgehen.

30 Die Regeln der Disziplin schließen eine Irreführung aus; zum methodologischen Problemkreis der „Täuschung“ siehe A. Ortmann/R. Hertwig, The costs of deception: Evidence from psychology, Experimental Economics (5) 2002, S. 111; R. Hertwig/A. Ortmann, Deception in experiments: Revisiting the arguments in its defense, Ethics & Behavior (18) 2008, S. 59.

31 Zu monetären Anreizen und dem Verhältnis der experimentellen Ökonomik zur Psychologie Guala, Methodology (Fn. 17), S. 231 ff.

32 Google Scholar zählt 3915 Zitationen (Suche vom 1.12.2016).

33 W. Güth/R. Schmittberger/B. Schwarze, An experimental analysis of ultimatum bargaining, Journal of Economic Behavior & Organization (3) 1982, S. 367.

34 S. Magen, Fairness, Eigennutz und die Rolle des Rechts. Eine Analyse auf Grundlage der Verhaltensökonomik, in: C. Engel et al. (Hg.), Recht und Verhalten, Tübingen 2007, S. 261 (283).

35 Binmore/Roth/Samuelson sprechen in E. van Damme et al., How Werner Güth's ultimatum game shaped our understanding of social behavior, Journal of Economic Behavior & Organization (108) 2014, S. 292 (297), sogar von damaliger „Feindseligkeit“ etablierter Ökonomen.

36 Ab Angeboten von 30 % wird die Ablehnung wahrscheinlicher; vgl. Magen, Fairness (Fn. 34) S. 261 (283); interkultureller Vergleich bei J. Henrich et al., Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies, Oxford 2004, S. 10: „no society in which experimental behavior is consistent with the canonical model from economics textbooks“.

gie der Empfänger ist komplex und beschäftigt die experimentelle Forschung bis heute.<sup>37</sup>

### III. Rechtlich relevante Forschungsfragen

Bevor wir näher auf den Zusammenhang zwischen ökonomischer Theorie und experimenteller Methode eingehen, möchten wir skizzieren, inwiefern Experimente gerade für Juristen von Interesse sein können.

Experimentatoren sind an tatsächlichen (deskriptiven) und vor allem *kausalen* Zusammenhängen interessiert. Daher konzentrieren sie sich oft auf einen Zusammenhang und blenden auch Umstände aus, die sie grundsätzlich als relevant einstufen, die aber nicht zum Kernbereich ihrer Forschungsfrage gehören. Für Juristen ist dies ungewohnt. Denn Juristen stellen typischerweise normative Fragen („Wie ist die Rechtslage?“)<sup>38</sup> und möchten möglichst alle Eigenschaften einer Umwelt wahrnehmen, die für den zu beurteilenden Sachverhalt relevant erscheinen.<sup>39</sup> Teilweise stehen Juristen auch einer starken Fokussierung auf Kausalitätsaussagen kritisch gegenüber.<sup>40</sup> Experimentelle Forschung kann aber für Rechtsgestaltung und Rechtsanwendung<sup>41</sup> von Bedeutung sein.

*Colin Camerer* und *Eric Talley* fassen diese Bedeutung prägnant zusammen:<sup>42</sup> Erstens testen Experimente Theorien über menschliches Verhalten. Dies ist schon deshalb von Bedeutung, weil ökonomische Handlungsempfehlungen an die Politik und Wirtschaft selten auf einzelnen Resultaten experimenteller Forschung basieren, sondern vielmehr auf Theoriegerüsten, die mit Hilfe von Experimenten überprüft und weiterentwickelt wurden.<sup>43</sup> Zweitens können Experimente Daten zu einem regelmäßigen Verhalten generieren, das bisher noch nicht durch eine Theorie erklärt worden ist. In diesem Fall bieten experimentelle Befunde eine Basis, um

37 Dazu *van Damme* in *van Damme et al.*, Werner Güth's ultimatum game (Fn. 35), S. 292 (294): „The game is beautiful; it is trivial from the conventional game theoretic point of view, but psychologically it is very rich.“.

38 Zu diesen unterschiedlichen Interessen *N. Petersen*, Braucht die Rechtswissenschaft eine empirische Wende?, *Der Staat* (49) 2010, S. 435 (435).

39 Vgl. *Fleischer/Schmolke/Zimmer*, Verhaltensökonomik als Forschungsinstrument (Fn. 5), S. 1 (45 f.).

40 Skeptisch *A. von Bogdandy*, Zur sozialwissenschaftlichen Runderneuerung der Verfassungsvergleichung: Eine hegelianische Reaktion auf Ran Hirschls Comparative Matters, *Der Staat* (55) 2016, S. 103 (111): „Kausalität nur eine von vielen Kategorien menschlicher Erkenntnis“. Demnach wäre es ein Fehler, „ihr den höchsten oder gar einzigen Platz hierbei zuzuweisen.“; allgemein zu unterschiedlichen Erkenntnisinteressen in den Sozialwissenschaften *Rosenberg*, Philosophy of Social Science (Fn. 19).

41 Zur Uneindeutigkeit der Unterscheidung zwischen Rechtspolitik und Rechtsanwendung *Voßkuhle/Wischmeyer*, The 'Neue Verwaltungsrechtswissenschaft' (Fn. 7), in Erscheinen: „application of legal rules almost always entails both elements of interpretation and of law-making“.

42 *Camerer/Talley*, Experimental Study of Law (Fn. 18), S. 1619 (1624 ff.).

43 Siehe hierzu *M. Hellwig*, Neoliberales Sektierertum oder Wissenschaft? Zum Verhältnis von Grundlagenforschung und Politikanwendung in der Ökonomie, *Preprints of the Max Planck Institute for Research on Collective Goods* 2015.

neue Theorien zu entwickeln. Drittens können Experimente rechtspolitische Interventionen testen. *Camerer/Talley* beschreiben diesen für Juristen besonders relevanten Punkt eingängig: „[t]he inspiration here is very much akin to experimentation in the physical sciences, such as testing of airplane wing designs in wind tunnels, or testing ship designs in “tow tanks” with simulated oceanic waves. These experiments do not guarantee that a wing or ship which performs well in a wind tunnel or tow tank will be the best design in the air or at sea, but they can weed out bad designs at a low cost.“<sup>44</sup> Experimente sind also nützlich, um menschliche Reaktionen auf rechtliche Regeln zu untersuchen. Sie geben zwar keine abschließende Antwort, können aber zumindest helfen, einen offensichtlichen Fehlschlag zu vermeiden.<sup>45</sup>

Im Folgenden wollen wir das rechtlich relevante Potential experimenteller Forschung anhand von einigen Beispielen illustrieren. Zu beachten ist, dass wir zwar einzelne Studien heranziehen, tatsächliche Handlungsempfehlungen aber auf mehr Evidenz basieren sollten.

## 1. Rechtsgestaltung

Experimentelle Forschung kann rechtspolitisch relevant werden.<sup>46</sup> Sie gibt dem Gesetzgeber einen Hinweis darauf, wie er die Rechtsordnung künftig in Einklang mit den Erkenntnissen der Wirtschafts- und Sozialwissenschaften ausgestalten könnte.<sup>47</sup> Wir möchten unterschiedliche Anwendungsfälle vorstellen und mit Beispielen illustrieren. Dabei sind die Grenzen zwischen unseren Kategorisierungen sicherlich fließend.

### a. Probleme beschreiben

Experimentelle Forschung kann helfen, Probleme zu erkennen und zu beschreiben, bei denen womöglich<sup>48</sup> ein Interventionsbedarf besteht.

44 Camerer/Talley, Experimental Study of Law (Fn. 18) S. 1619 (1625); auch Schmidt, The Role of Experiments (Fn. 26), S. 14 (26) sieht das Labor als Windtunnel, in dem Anreizmechanismen getestet werden können.

45 Ähnlich Eidenmüller, homo oeconomicus (Fn. 4), S. 216 (223).

46 Überblick auch bei Engel, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363 (383 ff.).

47 Vgl. zum Nutzen von Behavioral Law and Economics bei der „Fundierung gesetzgeberischer Entscheidungen“ Eidenmüller, homo oeconomicus (Fn. 4), S. 216 (223).

48 Ob ein solcher Bedarf besteht, ist eine getrennt zu behandelnde normative Frage; siehe nur H. Eidenmüller, Liberaler Paternalismus, JZ 2011, S. 814 (814): „Tatsächlich zeigen empirische Forschungen der Kognitionspsychologie, dass individuelles, rechtsrelevantes Verhalten von systematischen Rationalitätsdefiziten geprägt wird. Aus diesen empirischen Erkenntnissen folgt jedoch keine normative Leitlinie für die Rechtspolitik.“.

## Altersvorsorge

Ein relevanter Problembereich ist die Altersvorsorge.<sup>49</sup> Menschen neigen oft dazu, zukünftige Ereignisse bei ihren Entscheidungen schwächer zu berücksichtigen als unmittelbar bevorstehende. Die Experimente von *Richard Thaler* gehören zu den ersten Studien, die sich diesem Entscheidungsverhalten widmeten. Auf die Frage, welchen Geldbetrag sie erhalten müssten, um auf eine Sofortzahlung von \$ 15 zu verzichten, antworteten die Teilnehmer im Median \$ 30 für eine Verzögerung der Auszahlung von 3 Monaten, \$ 60 für eine Verzögerung von einem Jahr und \$ 100 für eine Verzögerung von drei Jahren.<sup>50</sup> Dies legt nahe, dass individuelle Diskontierungsraten von Kosten und Nutzen in der nahen Zukunft viel höher sind als in einer entfernten Zukunft (*hyperbolic discounting*). Dieses Phänomen ist seither eingehend untersucht worden.<sup>51</sup> Und tatsächlich können im Labor gemessene Zeitpräferenzen unterschiedliches Verhalten in der echten Welt vorhersagen.<sup>52</sup> Rechtliche Regelungen zur Sicherung der Altersvorsorge können durch diese Problembeschreibung motiviert sein und auf ihr aufbauen.<sup>53</sup>

## Überschuldung

Hyperbolisches Diskontieren kann zur Folge haben, dass die zukünftigen Kosten von Schulden unterschätzt werden. Bei Kreditkartennutzern kann dies zur Überschuldung führen.<sup>54</sup> Hierauf hat der Gesetzgeber schon reagiert:<sup>55</sup> Gesetzliche Interventionen wie die Verbraucherkredit-RL 2008/48/EG oder der Credit Card Accountability Responsibility and Disclosure Act 2009 in den USA sollen eine Überschuldung verhindern. Experimentelle Forschung kann helfen, diese Interventionen laufend anzupassen, indem etwa bestimmte Eigenschaften von Kreditkarten wie Lock-Angebote (*teaser rates*) identifiziert werden, die individuelle Schwächen im Entscheidungsverhalten ausnutzen.<sup>56</sup>

49 Siehe etwa *A. Engert*, Die automatische Einbeziehung des Arbeitnehmers in die betriebliche Altersversorgung, Zeitschrift für Arbeitsrecht (35) 2004, S. 311.

50 *R. Thaler*, Some Empirical Evidence on Dynamic Inconsistency, Economics Letters (8) 1981, S. 201.

51 Übersicht bei *S. Frederick/G. Loewenstein/T. O'Donoghue*, Time discounting and time preference: A critical review, Journal of Economic Literature (40) 2002, S. 351.

52 Etwa *S. Meier/C. Sprenger*, Present-biased preferences and credit card borrowing, American Economic Journal: Applied Economics (2) 2010, S. 193, zum Leihverhalten.

53 Hierzu *Engert*, Die automatische Einbeziehung (Fn. 49), S. 311; im Kontext von „Nudging“: *L. A. Reisch/J. Sandrina*, Nudging in der Verbraucherpolitik: Ansätze verhaltensbasierter Regulierung, Baden-Baden 2015, S. 56 ff.

54 *O. Bar-Gill*, Seduction by Contract: Law, Economics, and Psychology in Consumer Markets, Oxford 2012, S. 83 ff.

55 *H. Eidenmüller*, Die Rechtfertigung von Widerrufsrechten, AcP (210) 2010, S. 67 (68) weist darauf hin, dass *Philipp Heck* schon 1891 ein Reurecht beim Abzahlungskauf vorgeschlagen habe: „Kunden würden zum Ankauf entbehrlicher und ihr Leistungsvermögen übersteigender Gegenstände verleitet. Das sei [Heck zufolge] .... psychologisch erklärlich, da eine erst in Zukunft zu erfüllende Verbindlichkeit im Austausch gegen einen augenblicklichen Genuss meist unterschätzt ...“ werde“.

56 *Bar-Gill*, Seduction by Contract (Fn. 54), S. 113.

## Anlageentscheidungen

Das Zusammenspiel von Problembeschreibung und Intervention zeigt sich auch gut im Bereich des Verbraucherschutzes<sup>57</sup> bei Anlageentscheidungen.<sup>58</sup> Beachtlich ist hier der 2010 erschienene Report von *Chater/Huck/Inderst*, in dem verhaltensökonomische Forschung über Verbraucherverhalten im Bereich von Finanzanlageprodukten, aber auch eigens durchgeführte Experimente berichtet werden.<sup>59</sup> Ein wichtiger Aspekt ihrer Diagnose war, dass Anleger die Produkte, in die sie investieren, nicht richtig verstehen.<sup>60</sup> Der Bericht war bei der Entwicklung der Anforderungen an die sog. Basisinformationsblätter<sup>61</sup> von Bedeutung.<sup>62</sup> Akzeptiert man also Anlegerschutz als legitimes Ziel, kann experimentelle Forschung helfen, die richtige Informationsgestaltung ausfindig zu machen.<sup>63</sup>

### b. Interventionsbedarf einschätzen

Experimentelle Forschung kann auch helfen, den tatsächlichen Bedarf für eine (zusätzliche) rechtliche Intervention besser einzuordnen. Der Befund, dass „Menschen oft viel mehr verlangen, um ein Objekt aufzugeben, als das, was sie zu zahlen bereit wären, um es zu erlangen“,<sup>64</sup> wurde von *Richard Thaler* als Besitzeffekt (*endowment effect*) bezeichnet.<sup>65</sup> Dieser Effekt führt im Entscheidungsverhalten zu einer systematischen Verzerrung zugunsten der Dinge, wie sie sind (*status quo bias*).<sup>66</sup> Rechtliche Interventionen sollen verhindern, dass Unternehmen diese Verzer-

- 57 Die Generaldirektion Justiz und Verbraucher der Europäischen Kommission beruft sich auf „Behavioral research“: [http://ec.europa.eu/consumers/consumer\\_evidence/behavioural\\_research/index\\_en.htm](http://ec.europa.eu/consumers/consumer_evidence/behavioural_research/index_en.htm) (zuletzt abgerufen am 29.6.2016).
- 58 Zu *behavioral finance* einführend *C. Schmies*, Behavioral Finance und Finanzmarktregelung, in: *C. Engel et al. (Hg.)*, Recht und Verhalten, Tübingen 2007, S. 165; *L. Klöhn*, Kapitalmarkt, Spekulation und *Behavioral Finance*, Berlin 2006, S. 80 ff.
- 59 *Chater/Huck/Inderst*, Consumer Decision-Making in Retail Investment Services: A Behavioural Economics Perspective 2010.
- 60 Siehe die Bezugnahme bei *Chater/Huck/Inderst*, Consumer Decision-Making (Fn. 59), S. 50, auf die Studie zum Umgang mit Informationen zu Index-Fonds von *J. J. Choi/D. Laibson/B. C. Madrian*, Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds, *Review of Financial Studies* (23) 2010, S. 1405.
- 61 Eingeführt mit Verordnung (EU) Nr. 1286/2014 des europäischen Parlaments und des Rates vom 26. November 2014 über Basisinformationsblätter für verpackte Anlageprodukte für Kleinanleger und Versicherungsanlageprodukte (PRIIPs-VO).
- 62 Vgl. den 17. Erwägungsgrund der PRIIPs-VO, wo auf „ongoing research into consumer behaviour“ Bezug genommen wird.
- 63 Hierzu *Eidenmüller*, Liberaler Paternalismus (Fn. 48), S. 814 (819): „Den Anlegern sollten möglichst wenige, hervorgehobene Informationen gegeben werden. Eine Informationsgestaltung mit dem Ziel der Ausnutzung bestimmter Rationalitätsdefizite ist als rechtswidrig anzusehen.“.
- 64 *D. Kahneman/J. L. Knetsch/R. H. Thaler*, Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias, *The Journal of Economic Perspectives* (5) 1991, S. 193 (194): „people often demand much more to give up an object than they would be willing to pay to acquire it“.
- 65 *R. Thaler*, Toward a positive theory of consumer choice, *Journal of Economic Behavior & Organization* (1) 1980, S. 39 (44).
- 66 Dies ist für die Rechtsökonomik bedenklich. Denn wichtiger Ansatzpunkt ist das Coase'sche Theorem, wonach Verfügungsrechte bei fehlenden Transaktionskosten ihren Weg zu demjenigen finden,

rung ausnutzen.<sup>67</sup> *Jennifer Arlen* und *Stephan Tontrup* untersuchen, inwiefern der sog. Besitzeffekt durch bereits bestehende Institutionen abgeschwächt wird.<sup>68</sup> Sie zeigen, dass das Recht schon Institutionen bereithält, die für eine „Entzerrung“ (*debiasing*) des Entscheidungsverhaltens sorgen können: Zu diesen Institutionen gehören etwa Abstimmungen oder die Delegation von Entscheidungen auf Vertreter und Organe. Während es also tatsächlich ein Problem gibt, das grundsätzlich einen Interventionsbedarf auszulösen vermag, gibt es bereits Instrumente, um ihm beizukommen. *Arlen/Tontrup* mahnen insofern zu regulatorischer Zurückhaltung.<sup>69</sup>

### c. Lösungen evaluieren

Mit Experimenten lassen sich mögliche Lösungen von Problemen evaluieren. Besonders interessant ist diese Möglichkeit in Hinblick auf alternative Lösungen, zu denen „in der echten Welt“ keine Beobachtungen gemacht werden konnten.<sup>70</sup> Dazu lassen sich gesetzliche Interventionen auf ungewollte Nebenwirkungen prüfen.<sup>71</sup> Dieses Potential illustriert die Studie von *Georg Borges* und *Bernd Irlenbusch*, die an die (damalige) Ausgestaltung von Widerrufsrechten in Fernabsatzverträgen im BGB anknüpft.<sup>72</sup>

*Borges/Irlenbusch* fragen, ob sich die Gefahr, dass der Verkäufer ausgenutzt wird, erhöht, wenn ein Widerrufsrecht nicht mehr freiwillig angeboten, sondern gesetzlich festgeschrieben wird. *Borges/Irlenbusch* verstehen es als Ausnutzung, wenn ein Gegenstand bestellt, genutzt und der Kauf dann widerrufen wird. Beispiel ist der Kauf von Fernsehern vor Fußballweltmeisterschaften in der Absicht, das Gerät nach Benutzung zurückzugeben.<sup>73</sup> Ein solches Verhalten ist zwar auf Seiten des

der ihnen den höchsten Wert beimisst; s. *R. H. Coase*, The Problem of Social Cost, *Journal of Law and Economics* (3) 1960, S. 1.

67 Als mögliche Maßnahmen diskutiert *Schmolke*, Grenzen der Selbstbindung (Fn. 24), S. 868 ff., für den Bereich des Verbraucherkreditrechts gesetzlich zwingende verbraucherfreundliche „default rules“ und ein gesetzlich zwingendes Optionsmodell „nach dem der Verbraucher das Recht zur Wahl zwischen einem Vertrag mit und ohne die in Rede stehende Vertragsbedingung hat“; *Eidenmüller*, Widerrufsrechte (Fn. 55), S. 67 (95 f.), führt den Besitzeffekt als möglichen Bestandteil einer Begründung von Widerrufsrechten an; zur Bedeutung des Besitzeffekts in Zusammenspiel mit dispositivem Recht *Eidenmüller*, Liberaler Paternalismus (Fn. 58) S. 814 (817): „wichtigster Ausgangspunkt des liberalen Paternalismus“.

68 *J. Arlen/S. Tontrup*, Does the Endowment Effect Justify Legal Intervention? The Debiasing Effect of Institutions, *The Journal of Legal Studies* (44) 2015, S. 143.

69 *Arlen/Tontrup*, Does the Endowment Effect Justify Legal Intervention? (Fn. 68), S. 143 (145 f.), erklären dies damit, dass durch Abstimmungen und Delegationen die Verantwortung für die Entscheidung geteilt und mögliche Reue abgemildert werde.

70 *C. Engel*, Legal Experiments: Mission Impossible? Den Haag 2013, S. 14.

71 Vgl. zum normativen Problem der „Kosten guter Absichten“ *Engel*, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363 (396).

72 *G. Borges/B. Irlenbusch*, Fairness Crowded Out by Law: An Experimental Study on Withdrawal Rights, *JITE* (163) 2007, S. 84; auf diese Studie bezieht sich auch *Eidenmüller*, Widerrufsrechte (Fn. 55), S. 67 (95).

73 *Borges/Irlenbusch*, Fairness Crowded Out by Law (Fn. 72), S. 84 (88).

Verkäufers mit erhöhten Kosten verbunden, dürfte aber mangels rein schikanöser Absicht des Käufers grundsätzlich nicht als rechtswidrig zu beanstanden sein.<sup>74</sup> Tatsächlich beobachten sie in ihrem Experiment einen Effekt, der nahelegt, dass die Käufer eher bereit sind, das Widerrufsrecht auszunutzen, wenn dieses aufgrund einer Vorgabe und nicht aufgrund einer freiwilligen Entscheidung des Verkäufers gewährt wurde.<sup>75</sup> Den Verkäufern ist so eine Möglichkeit genommen, Kooperationsbereitschaft zu signalisieren.<sup>76</sup>

## 2. Rechtsanwendung

Experimentelle Forschung kann sich auch auf die Anwendung bereits bestehender Normen beziehen, etwa um zu bestimmen, welches bestehende rechtliche Instrument der Behandlung eines Problems angemessen ist.<sup>77</sup> Ein beachtliches Beispiel ist die von *Emanuel V. Towfigh* und *Andreas Glöckner* in der JuristenZeitung veröffentlichte empirische Studie zur angemessenen Behandlung von Sportwetten, in der die Autoren nachweisen, dass die Unterscheidung zwischen Glück- und Geschicklichkeitsspielen dem Gefährdungspotential der einzelnen Spiele nicht gerecht wird.<sup>78</sup>

Ein weiteres Beispiel liefern *Christoph Engel* und *Lilia Zhurakhovska*, die experimentell der Frage nachgehen, wie es sich auswirkt, wenn einem unabhängigen Entscheidungsträger für seine Entscheidung eine Begründungspflicht auferlegt wird.<sup>79</sup> Danach greift ein Entscheidungsträger, der über eine Sanktion zu befinden hat, zu einer mildernden Sanktion, wenn dem Sanktionierten auch die Gründe für die Entscheidung übermittelt werden.<sup>80</sup> Tatsächlich gilt die Begründung einer Entscheidung als ein geeignetes Instrument, um zwischen Sender und Empfänger einen Konsens herzustellen. Im Experiment zeigt sich der Empfänger aber nur dann geläutert, wenn die Begründung nicht nur an ihn, sondern an alle Teilnehmer des Ex-

74 Hierzu und zu Nutzungsvergütungspflichten in § 357 Abs. 7 BGB *P. Mankowski*, Verbraucherschützendes Widerrufsrecht und Rechtsmissbrauch, JZ 2016, S. 787 (789): „Der Unternehmer bekommt nur Kompensationsschutz, aber keinen Bestandsschutz.“.

75 *B. Rockenbach*, Fairness Crowded Out by Law: An Experimental Study on Withdrawal Rights: Comment, JITE (163) 2007, S. 106 (107), beurteilt den Effekt als nicht stark und sieht weiteren Forschungsbedarf.

76 Zur Erhöhung der Kooperationsbereitschaft durch eigene Verletzbarmachung nun auch *A. Kopányi-Peuker/T. Offerman/R. Sloof*, Fostering cooperation through the enhancement of own vulnerability, Games and Economic Behavior (101) 2017, S. 273.

77 *E. V. Towfigh*, Empirical Arguments in Public Law Doctrine: Should Empirical Legal Studies Make a “Doctrinal Turn?”, International Journal of Constitutional Law (12) 2014, S. 670, erläutert, inwiefern die Dogmatik Anlass zu empirischer Forschung geben kann.

78 Siehe *Glöckner/Towfigh*, Geschicktes Glücksspiel (Fn. 8), S. 1027 (1035), wobei es sich hier um eine Online-Studie und nicht um ein Labor-Experiment handelt.

79 *C. Engel/L. Zhurakhovska*, Words Substitute Fists: Justifying Punishment in a Public Good Experiment, Preprints of the Max Planck Institute for Research on Collective Goods 2013.

80 *Engel/Zhurakhovska*, Words Substitute Fists (Fn. 79), S. 2, berufen sich für Beispiele für Sanktionen ohne Begründung – vor allem im US-amerikanischen Recht – auf *F. Schauer*, Giving Reasons, Stanford Law Rev. (47) 1995, S. 633 (634).

periments übermittelt wird.<sup>81</sup> Im Verwaltungsverfahren ist bei Verwaltungsakten grundsätzlich eine Begründung erforderlich.<sup>82</sup> Zu der Vielzahl an Funktionen,<sup>83</sup> die Begründungen erfüllen sollen, gehört auch die Förderung von Akzeptanz beim Empfänger zugunsten der Verfahrenseffektivität.<sup>84</sup> Die Studie von *Engel/Zhurakhovska* gibt zumindest Anlass zum Zweifel, ob diese Funktion tatsächlich erfüllt wird, wenn die Beweggründe nicht an einen weiteren Adressatenkreis kommuniziert werden. Eine Schlüsselfunktion für die Steigerung von Akzeptanz beim Empfänger kann hingegen eine vorhergehende Anhörung des Empfängers spielen, wie die Studie von *Marco Kleine, Pascal Langenbach* und *Lilia Zhurakhovska* zeigt.<sup>85</sup>

#### IV. Theorie und Hypothese

Nachdem wir verschiedene rechtlich relevante Anwendungsfälle skizziert haben, wenden wir uns nun dem Zusammenhang zwischen ökonomischer Theorie und experimenteller Methode zu.

##### 1. Ökonomische Theorie

Ökonomische Theorie wird eingesetzt, um die im Experiment zu testenden Hypothesen zu formulieren. Bevor wir anhand eines Beispiels illustrieren, wie Hypothesen gebildet werden, möchten wir auf die theoretischen Grundlagen eingehen, von denen auch in rechtsökonomischen Experimenten ausgegangen wird. Diese Grundlagen bestehen in der bereits eingangs erwähnten ökonomischen Standardtheorie und ihrer verhaltensökonomischen Qualifizierung.

###### a. Homo oeconomicus

Die ökonomische Standardtheorie geht vom berühmt-berüchtigten *homo oeconomicus*, dem voll-rationalen, eigeninteressierten Akteur aus.<sup>86</sup> Nach dem *Rational Choice*-Paradigma<sup>87</sup> wählt der *homo oeconomicus* aus den ihm zur Verfügung ste-

81 *Engel/Zhurakhovska*, Words Substitute Fists (Fn. 79), S. 22: „It seems that giving reasons is not necessarily a good idea. If these reasons are not made public, the authority may overly focus on educating the addressee, whereas bystanders become skeptical that others who are tempted to misbehave are effectively disciplined. By contrast, if the authority is transparent about the reasons, words may indeed partly substitute acts, to everybody's benefit.“.

82 Nicht aber bei Realakten, Rechtsverordnungen und Verwaltungsvorschriften, s. *J. Saurer*, Die Begründung im deutschen, europäischen und US-amerikanischen Verwaltungsverfahrensrecht, Verwaltungssarchiv (100) 2009, S. 364 (366 ff.), zur Ausrichtung des VwVfG auf Handlungsformen.

83 *Saurer*, Begründung (Fn. 82), S. 364 (382): „Funktionspluralismus“.

84 Statt vieler Stelkens/Bonk/Sachs/Stelkens VwVfG, 8. Auflage 2014, § 39 Rn. 1.

85 *M. Kleine/P. Langenbach/L. Zhurakhovska*, How Voice Shapes Reactions to Impartial Decision Makers: An Experiment on Participation Procedures, Journal of Economic Behavior & Organization (143) 2017, S. 241; eingehend hierzu *P. Langenbach*, Der Anhörungseffekt, Tübingen 2017.

86 Aus juristischer Sicht *J. Lüdemann*, Die Grenzen des homo oeconomicus und die Rechtswissenschaft, in: *C. Engel et al. (Hg.)*, Recht und Verhalten, Tübingen 2007, S. 7.

87 Welche Inhalte zur „rational choice theory“ gehören, ist umstritten; siehe hierzu *R. B. Korobkin/T. S. Ulen*, Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics, California Law Review (88) 2000, S. 1051 (1055): „considerable debate within both the eco-

henden Optionen diejenige aus, die ihm den meisten Nutzen bringt.<sup>88</sup> Was dem *homo oeconomicus* Nutzen bringt, bestimmt sich nach einer subjektiven Nutzenfunktion.<sup>89</sup> Um seinen Nutzen zu mehren, wählt der *homo oeconomicus* gemäß der Rationalitätsannahme rational und systematisch und nicht etwa willkürlich. Es wird dabei unterstellt, dass seine Vorlieben für bestimmte Güter stabil und transitiv sind.<sup>90</sup> Wenn also jemand Bananen Äpfeln und Äpfel Birnen vorzieht, dann wird erstens angenommen, dass dies immer der Fall ist (stabile Präferenzen) und zweitens, dass er Bananen auch Birnen vorzieht (Transitivität). Es handelt sich beim *homo oeconomicus* nicht um ein Menschenbild, sondern um ein ökonomisches Modell, das ein Verhalten positiv erklären und für bestimmte Zwecke Voraussagen generieren soll.<sup>91</sup>

### b. Verhaltensökonomik

Diese Theorie ist mit Hilfe von Experimenten weiterentwickelt worden. Ein ganzheitliches Modell individuellen Verhaltens, das die gesamte Bandbreite experimenteller Befunde einfängt, ist hierbei nicht entstanden.<sup>92</sup> Vielmehr gibt es eine Reihe von Ansätzen, wie die Annahmen der Standardtheorie zu modifizieren sind.<sup>93</sup>

Ein einflussreicher Ansatz ist die von *Daniel Kahneman* und *Amos Tversky* formulierte *Prospect Theory*,<sup>94</sup> mit der menschliche Entscheidungen angesichts eines be-

nomics and law-and-economics communities about precisely what rational choice theory is and is not.”.

88 Einführend *Towfigh/Petersen*, Ökonomische Methoden (Fn. 13), S. 27ff.

89 Ob auch altruistisches Verhalten als nutzenbringend anzusehen ist, variiert in der Literatur. „Schwache“ Konzeptionen von Eigennutz erlauben dies; *Korobkin/Ulen*, Law and Behavioral Science (Fn. 87), S. 1051 (1055) führen zu unterschiedlichen Verwendungen des *Rational Choice*-Paradigmas aus: „As it is applied implicitly or explicitly in the law-and-economics literature, however, it is understood alternatively as a relatively weak, or “thin,” presumption that individuals act to maximize their expected utility, however they define this, or as a relatively strong, or “thick,” presumption that individuals act to maximize their self-interest.“; kritisch zu „schwachen“ Konzeptionen *M. Englerth*, Behavioral Law and Economics – eine kritische Einführung, in: C. Engel et al. (Hg.), Recht und Verhalten, Tübingen 2007, S. 60 (68); „definitorische Korrektur [...] um den Preis der Nicht-Falsifizierbarkeit und der Beschränkung auf Ex-post-Aussagen“; hierzu auch *Rosenberg*, Philosophy of Social Science (Fn. 19), S. 100.

90 Die Annahme vollständiger Information findet sich seltener in modernen Modellen; siehe *Towfigh/Petersen*, Ökonomische Methoden (Fn. 13), S. 27ff.

91 *Lüdemann*, Die Grenzen des *homo oeconomicus* (Fn. 86), S. 7 (17), mag aber nicht ausschließen, „dass die methodologische Annahme rationaler Nutzenmaximierung schließlich auch die Einstellungen der Menschen prägt, die mit diesem Modell arbeiten“; ähnlich *Englerth*, Behavioral Law and Economics (Fn. 89), S. 60 (74).

92 *Towfigh/Petersen*, Ökonomische Methoden (Fn. 13), S. 23; hierzu auch *Schmidt*, The Role of Experiments (Fn. 26), S. 14 (21): „The Holy Grail of behavioral economics is a universal theory of human behavior that applies to all circumstances. It may be that such a theory exists and that we will eventually find it. However, at present I do not think that we should strive for such a universal theory.“.

93 W. *Güth*, On ultimatum bargaining experiments—A personal review, Journal of Economic Behavior & Organization (27) 1995, S. 329 (342), spricht vom „neoclassical repair shop“.

94 D. *Kahneman/A. Tversky*, Prospect Theory: An Analysis of Decision Under Risk, Econometrica (47) 1979, S. 263.

kannten Risikos<sup>95</sup> beschrieben werden sollen. Die wichtigste Innovation gegenüber dem Rationalitätsmodell ist dabei, dass die möglichen Optionen nicht absolut bewertet werden, sondern als Gewinne und Verluste relativ zu einem Referenzpunkt.<sup>96</sup> Mögliche Verluste werden dabei entsprechend der experimentell beobachteten Abneigung gegen Verluste (*loss aversion*) stärker gewichtet als mögliche Gewinne. *Prospect Theory* kann etwa den bereits angesprochenen Besitzeffekt erklären. In einem vielzitierten<sup>97</sup> Experiment von Kahneman/Knetsch/Thaler<sup>98</sup> erhielt eine Teilnehmergruppe Kaffeetassen. Die Teilnehmer wurden gebeten, den Preis anzugeben, zu dem sie bereit wären, die Tasse aufzugeben. Die übrigen Teilnehmer sollten angeben, zu welchem Preis sie bereit wären, eine Tasse zu kaufen. Die Teilnehmer in Besitz einer Tasse gaben hierbei den höheren Preis an. Nach der *Prospect Theory* bildet der Besitz der Tasse einen Referenzpunkt, von dem aus die Aufgabe der Tasse als Verlust wahrgenommen wird. Als solcher wird er stärker gewichtet als der mögliche Gewinn eines Gutes, wenn man nicht in dessen Besitz ist. Für den Besitzeffekt kann es aber auch andere Gründe geben. So geht die Theorie der Bereuens-Aversion (*regret aversion*) davon aus, dass Menschen Güter, die sie besitzen, nicht tauschen, weil sie das Gefühl vermeiden wollen, eine Entscheidung zu bereuen.<sup>99</sup>

Andere Ökonomen haben unter Berücksichtigung der Tatsache, dass Menschen auch danach entscheiden, was sie als „fair“ empfinden, Modelle entwickelt, die Fairness und Reziprozität berücksichtigen.<sup>100</sup> Weitere ökonomische Theorien modellieren Präferenzen so, dass sie eine Abneigung gegen „Ungleichheiten“ abbilden (*inequity aversion*).<sup>101</sup> Diese so fortentwickelten verhaltensökonomischen Theorien werden wiederum eingesetzt, um Verhaltensvorhersagen zu treffen.

95 Risiko meint hier der Fall, dass die Wahrscheinlichkeit des Eintritts eines bestimmten Ereignisses bekannt ist. Kann das Risiko nicht bestimmt werden, spricht die Ökonomik von Unsicherheit.

96 Einführend mit weiteren Nachweisen Englert, Behavioral Law and Economics (Fn. 89) S. 60 (83 ff.); zur Relevanz von Prospect Theory für die Finanzmarktregulierung Schmies, Behavioral Finance und Finanzmarktregulierung (Fn. 58), S. 165 (171).

97 Google Scholar zählt 3704 Zitationen (Suche vom 12.10.2016).

98 D. Kahneman/J. L. Knetsch/R. H. Thaler, Experimental Tests of the Endowment Effect and the Coase Theorem, *Journal of Political Economy* (98) 1990, S. 1325.

99 Siehe die Nachweise bei Arlen/Tontrup, Does the Endowment Effect Justify Legal Intervention? (Fn. 68), S. 143 (145); und den Übersichtsartikel von R. B. Korobkin, Wrestling with the Endowment Effect, or How to Do Law and Economics without the Coase Theorem, in: Zamir/Teichman (Hg.), *The Oxford Handbook of Behavioral Economics and the Law*, Oxford 2014, S. 300.

100 M. Rabin, Incorporating Fairness into Game theory and Economics, *The American economic review* (83) 1993, S. 1281; A. Falk/U. Fischbacher, A Theory of Reciprocity, Games and economic behavior (54) 2006, S. 293; aus rechtlicher Sicht Magen, Fairness (Fn. 34).

101 E. Fehr/K. M. Schmidt, A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics* (113) 1999, S. 817; G. E. Bolton/A. Ockenfels, ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review* (90) 2000, S. 166.

### c. Zum Verhältnis von Standardtheorie und Verhaltensökonomik

Die Kritik seitens der Verhaltensökonomik hat dem Standardmodell keine generelle Absage erteilt. Als Modell ist es weiterhin nützlich.<sup>102</sup> Dies lässt sich am Beispiel des Verbraucherschutzrechts zeigen: Die Verhaltensökonomik gibt wichtige Anstöße, um einzelne Risiken im Verbraucherrecht zu identifizieren.<sup>103</sup> Für das Problem aber, dass Verbraucher regelmäßig kleine Forderungen nicht gegen Unternehmen geltend machen und hierdurch ein Vollzugsdefizit entsteht (Problem der Streuschäden), gibt nach wie vor die Standardtheorie die beste Erklärung. Es liegt ein Fall „rationaler Apathie“ vor: Es lohnt sich – auch angesichts der mit Geltendmachung des Schadens verbundenen Risiken – für den einzelnen Verbraucher nicht, diese Schäden geltend zu machen.<sup>104</sup> Umgekehrt besteht aber für Unternehmen ein Anreiz diese Apathie auszunutzen.<sup>105</sup> Von dieser Problembeschreibung ausgehend lassen sich Lösungsvorschläge beurteilen, wie etwa Verbandsklagen oder behördliche Lösungen als alternative Mittel der Rechtsdurchsetzung.<sup>106</sup>

## 2. Hypothesenbildung

Eine Theorie wird getestet, indem mittels der Theorie Hypothesen über das Verhalten der Versuchspersonen in einer bestimmten Situation formuliert und dann im Experiment geprüft werden. Die deduktiv aus einer allgemeinen Theorie abgeleiteten Hypothesen können dabei nicht bestätigt, sondern nur widerlegt werden (Prinzip der Falsifikation).<sup>107</sup> Die Bestätigung einer Theorie kann allenfalls auf eine „schwache“ Art erfolgen, etwa wenn sich die Theorie in einer ganzen Reihe von Fällen und unter verschiedenen Umständen behauptet.<sup>108</sup> Auch wenn nicht in einem strengen Sinne Theorie getestet, sondern vielmehr eine Entscheidungssituation modelliert werden soll, werden die Hypothesen auf Grundlage von ökonomi-

102 Laut *Toufigh/Petersen*, Ökonomische Methoden (Fn. 13), S. 23, ist das Rationalmodell „nach wie vor auch im Hinblick auf die empirische Fundierung und mangels ganzheitlicher theoretischer Alternativen Grundlage [...] der ökonomischen Analyse des Rechts“; *Lüdemann*, Jenseits von Rationalität und Eigennutz, in: Engel et al. (Hg.), Recht und Verhalten, Tübingen 2007, S. 1 (2): „überaus fruchtbare intellektuelle Element [...], doch kein theoretischer Generalschlüssel“.

103 Überblick bei *Reisch/Sandrini*, Nudging in der Verbraucherpolitik (Fn. 53).

104 Siehe *F. Weber*, Ökonomische Analyse verschiedener Rechtsdurchsetzungsmechanismen im Verbraucherrecht, in: B. Heiderhoff/R. Schulze (Hg.), Verbraucherrecht und Verbraucherverhalten, Baden-Baden 2016, S. 187; *G. Wagner*, Neue Perspektiven im Schadensersatzrecht-Kommerzialisierung, Strafschadensersatz, Kollektivschäden: Gutachten A für den 66. Deutschen Juristentag, München 2006, S. 106 ff.

105 Wettbewerb verstärkt diesen Anreiz; siehe *Bar-Gill*, Seduction by Contract (Fn. 54), S. 2: „competition forces sellers to exploit the biases and misperceptions of their customers“.

106 Vgl. *Engel*, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363 (377), der darauf hinweist, dass das Verhalten der Nutzenmaximierer rechtspolitisch der eigentliche Grund für die Intervention sein kann.

107 Zur Falsifizierbarkeit als Akzeptanzkriterium empirischer Wissenschaft grundlegend *K. Popper*, The Logic of Scientific Discovery, London u.a. 2002.

108 Siehe *Schmidt*, The Role of Experiments, (Fn. 26), S. 14 (17), der auf die Nützlichkeit einer Theorie und ihrer Anwendbarkeit unter extremen Bedingungen verweist.

scher Theorie gebildet.<sup>109</sup> Hypothesen zwingen zur eindeutigen Formulierung eines Wirkzusammenhangs zwischen zwei (oder mehreren) Variablen, der dann widerlegt oder *ex negativo* bekräftigt werden kann.

Wir illustrieren die Herleitung einer rechtlich relevanten Hypothese anhand der bereits angesprochenen Studie von *Arlen/Tontrup*. Die Studie widmet sich der Frage, inwiefern ein Besitzeffekt durch rechtliche Institute ausgeglichen werden.<sup>110</sup> Diese Forschungsfrage impliziert bereits einen verhaltensökonomischen Befund. Dementsprechend stützen sich die Autoren für ihre Hypothesen über das Verhalten der Teilnehmer im Experiment nicht einfach auf die ökonomische Standardtheorie. Sie gründen ihre Hypothesen vielmehr auf der Theorie, die den Besitzeffekt mit einer Bereuens-Aversion (*regret aversion*) erklären möchte. Danach vermeiden Menschen es, Güter, die sie gegenwärtig besitzen, zu tauschen, weil sie antizipieren, dass sie ihre Entscheidung möglicherweise bereuen könnten. Auf vorhergehender Forschung<sup>111</sup> aufbauend gehen *Arlen/Tontrup* davon aus, dass die Verantwortlichkeit für eine Entscheidung notwendige Bedingung ist, damit Menschen auch Reue für eine Entscheidung empfinden. Dementsprechend werden sie in solchen Situationen auch die mögliche Reue antizipieren.

*Arlen/Tontrup* stellen die Hypothese auf, dass ein Besitzeffekt in weniger Fällen zu beobachten sein wird, wenn die Teilnehmer Güter, die sie gegenwärtig besitzen, nicht selbst tauschen, sondern von einem Stellvertreter tauschen lassen – wenn also ein Agent handelt, der für die Entscheidung über den Tausch verantwortlich ist.<sup>112</sup> *Arlen/Tontrup* sagen weiter voraus, dass diejenigen, die mittels eines Agenten gehandelt haben, weniger Verantwortungsgefühl für die Entscheidung zum Tausch eines Gutes empfinden. Deshalb sollten sie auch weniger Reue antizipieren. Wie diese Hypothesen in einer Versuchsanordnung getestet werden, erläutern wir im nächsten Abschnitt.

## V. Versuchsanordnung

In diesem Abschnitt gehen wir näher darauf ein, wie in einem Experiment Hypothesen, die anhand (verhaltens-)ökonomischer Theorie formuliert wurden, getestet werden. Es geht also um die Versuchsanordnung. In einer Studie ist sie idealerweise so klar dargestellt, dass es möglich wäre, die Versuchsanordnung nachzustellen und zu prüfen, ob das Experiment repliziert werden kann.<sup>113</sup>

109 Siehe hierzu auch unter II. 1.

110 *Arlen/Tontrup*, Does the Endowment Effect Justify Legal Intervention? (Fn. 68).

111 M. Zeelenberg/W. W. van Dijk/A. S. R. Manstead, Reconsidering the Relation between Regret and Responsibility, *Organizational Behavior and Human Decision Processes* (74) 1998, S. 254.

112 *Arlen/Tontrup*, Does the Endowment Effect Justify Legal Intervention? (Fn. 68), S. 143 (153).

113 Zur Replizierbarkeit von Experimenten aus ökonomischer Sicht C. F. Camerer et al., Evaluating replicability of laboratory experiments in economics, *Science* (351) 2016, S. 1433; aus psychologischer Sicht *Open Science Collaboration*, Estimating the reproducibility of psychological science,

## 1. Manipulation unter kontrollierten Bedingungen

Die Hypothesen treffen eine Aussage über die Auswirkungen einer manipulierten unabhängigen Variablen (auch erklärende oder exogene Variable genannt) auf eine abhängige Variable (auch erklärte oder endogene Variable genannt). In der Abstraktheit des Labors kann der Experimentator diese Auswirkungen untersuchen und Kausalbeziehungen feststellen.<sup>114</sup> In der Studie von Arlen/Tontrup liegt die manipulierte unabhängige Variable darin, ob ein Teilnehmer selbst entscheiden muss oder die Entscheidung auf einen Stellvertreter delegieren kann. Die erste abhängige Variable liegt in der Anzahl der Fälle, in denen ein Gut, das ein Teilnehmer besitzt, getauscht wird; die zweite abhängige Variable in der Frage, welches Maß an Verantwortungsgefühl die Teilnehmer aufweisen, die ihre Entscheidung delegiert haben.

Um seine Hypothese zu testen, bildet der Experimentator jeweils eine Kontrollgruppe („Baseline“) und eine Experimentalgruppe und nimmt zwischen diesen beiden Gruppen eine einzige Veränderung vor („Treatment“). Beobachtet man nun einen Unterschied zwischen Baseline und Treatment, können dank den kontrollierten Bedingungen des Labors andere Erklärungen (Störfaktoren) als die vorgenommene Manipulation ausgeschlossen werden.<sup>115</sup>

In der Studie von Arlen/Tontrup erhielt in der Kontrollgruppe, der Bedingung *Baseline*, jeder Teilnehmer eines von zwei Losen.<sup>116</sup> Jedes Los hatte die gleiche Gewinnchance. Die Teilnehmer hatten die Wahl, entweder das Los, das sie anfangs bekommen hatten, zu behalten oder es gegen das andere auszutauschen. Entschieden sie sich für den Tausch, erhielten sie einen Bonus von € 0,25. Da beide Lose die gleiche Gewinnchance hatten, hätte ein rationaler Entscheider sich für einen Tausch der Lose entschieden: Die Chancen zu gewinnen wären unverändert, aber der Teilnehmer würde etwas dazuverdienen. Ein Besitzeffekt ist zu beobachten, wenn sich der Teilnehmer dennoch dafür entscheidet, das Los zu behalten. 70% der Teilnehmer behielten ihr Los. Die Autoren führten dann eine Serie an Treatments durch, um zu testen, ob die Teilnehmer im Vergleich zum Baseline rationaler agieren, wenn sie die Verantwortung teilen – durch eine Vertretung oder durch eine Abstimmung.

Science (349) 2015, S. 943; Hinweise aus juristischer Perspektive bei Hamann, Evidenzbasierte Jurisprudenz (Fn. 9), S. 94 f.

114 Zur „Logik des Experiments“ einführend Sedlmeier/Renkewitz, Forschungsmethoden (Fn. 14), S. 126 ff.

115 Eingehend zu unterschiedlichen Kontrolltechniken Sedlmeier/Renkewitz, Forschungsmethoden (Fn. 14), S. 134 ff.

116 In den Worten der Autoren: „Each ticket had a 50 percent chance of winning. Winners earned a substantial payoff. Each subject was offered the opportunity to exchange his ticket for the other ticket plus a monetary bonus of €.25. Trading to obtain the bonus is the rational decision. Participants who keep their tickets exhibit an endowment effect. Consistent with prior evidence, more than 70 percent of the laboratory subjects exhibited an endowment effect.“.

So wurde in der Bedingung *Mandatory* jeder Teilnehmer (Prinzipal) mit einem weiteren Teilnehmer (Agent) zusammengeführt. Wie in der Bedingung *Baseline* erhielt jeder Teilnehmer in der Rolle des Prinzipals ein Los. Anders als im *Baseline* traf nun aber der Teilnehmer in der Rolle des Agenten die Entscheidung über den Tausch des Loses mit Wirkung für den Prinzipal. Wie im *Baseline* erhielt der Prinzipal im Fall eines Tausches einen Bonus in Höhe von € 0,25. Für die Teilnehmer in der Rolle des Agenten bestand ein Anreiz, sich für den Tausch zu entscheiden. Sie erhielten in jedem Fall € 2, wenn sie sich für den Tausch entschieden. Die Prinzipale konnten aber gegen die Durchführung des Tauschs ihr Veto einlegen (nicht gegen die Zahlung). Im Kontrast zum *Baseline* behielten diesmal nur 31 % aller Teilnehmer ihr Los.<sup>117</sup>

Sogar, wenn die Delegation oder die Abstimmung optional war, entschieden sich die Teilnehmer dafür, diese Institutionen einzusetzen, um ihre Entscheidungen zu rationalisieren (*to debias*). Die Autoren kommen zu dem Schluss, dass die Einführung neuer Institutionen daher oft unnötig ist, weil das Recht schon genügend Instrumente bereitstellt.<sup>118</sup>

## 2. Rolle des Zufalls

Über die Frage, ob und in welcher Rolle ein Teilnehmer an einem Experiment teilnimmt, entscheidet der Zufall. Die Teilnehmer werden zufällig aus einer Datenbank ausgewählt und zum Experiment eingeladen und zufällig auf die veränderten Bedingungen des Experiments verteilt. Solches *random assignment* ist für einen Experimentator von entscheidender Bedeutung.<sup>119</sup> Damit wird ein Auswahlfehler (*selection bias*) bei der Bildung der Stichprobe ausgeschlossen. Die potenziellen Teilnehmer haben unterschiedliche Eigenschaften (zum Beispiel einen unterschiedlichen Sinn für Fairness). Deshalb sollte für jeden potentiellen Teilnehmer in der Datenbank dieselbe Wahrscheinlichkeit bestehen, zum Experiment eingeladen zu werden. Weiter sollte für alle Teilnehmer, die zum Experiment eingeladen wurden, dieselbe Wahrscheinlichkeit bestehen, im Experiment eine bestimmte Rolle auszufüllen. Dies betrifft sowohl die Wahrscheinlichkeit, innerhalb einer experimentellen Bedingung eine bestimmte Rolle einzunehmen (z. B. in der Studie von Arlen/Tontrup die Rolle des Prinzipals oder des Agenten) als auch die Wahrscheinlichkeit Teilnehmer in einer bestimmten Bedingung zu sein, also in der Kontroll- oder Experimentalgruppe. Erst dann ist es möglich, eine Aussage darüber zu treffen, ob sich der beobachtete Unterschied im Verhalten der Teilnehmer in zwei Bedingun-

117 Weitere Kontroll-Treatments schlossen alternative Erklärungen – wie eine Neigung zum Unterlassen (*omission bias*) – aus.

118 Auf einem anderen Blatt steht natürlich, ob eine Vertretung oder Kollegialentscheidung in jedem Fall greifbar ist. Schließlich sind mit der Einrichtung einer Delegation weitere Transaktionskosten verbunden.

119 Sedlmeier/Renkewitz, Forschungsmethoden (Fn. 14), S. 132: „wichtigste Technik“.

gen tatsächlich auf die Veränderung, also das im Fall der Experimentalgruppe vorgenommene Treatment, zurückführen lässt. Ansonsten ließe es sich nicht ausschließen, dass ein bestimmtes Merkmal der Teilnehmer in der Experimentalgruppe für den Effekt maßgeblich war.<sup>120</sup>

### 3. Treatment an denselben Versuchspersonen?

Jede experimentelle Studie gibt darüber Auskunft, ob das Treatment an denselben Versuchspersonen („within-subjects design“) oder an einer zweiten Gruppe durchgeführt wurde („between-subjects design“). Der Experimentator muss sorgfältig abwägen, wofür er sich entscheidet. Im within-subjects-Design lassen sich die Reaktionen derselben Person auf verschiedene Bedingungen beobachten. Ein solches Design hat den Vorteil, dass mehr Teilnehmer an jedem Treatment die statistische Belastbarkeit der Ergebnisse erhöhen.<sup>121</sup> Dazu gibt es weniger individuelle Unterschiede, die die Ergebnisse „stören“ könnten, schließlich werden die Teilnehmer samt ihren Eigenschaften „konstant gehalten“. Andererseits besteht die Gefahr, dass es Auswirkungen auf die Teilnehmer hat, dass sie in der vorherigen Runde schon einem anderen Treatment ausgesetzt waren (*carry over-Effekt*).<sup>122</sup> Dies gibt oft den Ausschlag zugunsten eines between-subjects-Designs (wie es z.B. auch bei *Arlen/Tontrup* für die Zuteilung zu den Treatments *Mandatory* und *Baseline* eingesetzt wurde).

## VI. Statistische Analyse der Ergebnisse

Experimentatoren beschreiben die gewonnenen Daten und berichten die Ergebnisse der statistischen Analyse. Eine Besprechung der möglichen Analysemethoden kann an dieser Stelle nicht geleistet werden.<sup>123</sup> Wir möchten aber anhand der soeben herangezogenen Studie von *Arlen/Tontrup* näher auf eine Reihe wichtiger Punkte eingehen, um Verständnis und Interpretation experimenteller Ergebnisse zu erleichtern.

*Arlen/Tontrup* stellen die Hypothese auf, dass Agenten das Verantwortungsgefühl vermindern: Teilnehmer fühlen sich in der Bedingung *Baseline*, in der sie selbst entscheiden, ob sie ihr Los tauschen, verantwortlicher für ihre Entscheidung als in der Bedingung *Mandatory*, in der die Entscheidung für sie von einem Agenten getroffen wurde. Um diese Hypothese zu testen, wurden die Teilnehmer gebeten, ihr Gefühl der Verantwortung auf einer Skala von 1 (sehr geringe Verantwortung) bis 10

120 Ganz ausschließen kann man den Einfluss von Störvariablen nie (z.B. Tages-, Wochen- und Jahreszeit). Man kann aber versuchen, ihn so gering wie möglich zu halten; hierzu auch *Schmidt, The Role of Experiments* (Fn. 26), S. 14 (17): „experimental control is never perfect.“.

121 Siehe zu den Vor- und Nachteilen eines within-subjects-Designs auch *Sedlmeier/Renkewitz, Forschungsmethoden* (Fn. 14), S. 152 ff.

122 *Sedlmeier/Renkewitz, Forschungsmethoden* (Fn. 14), S. 164 f.

123 Einführend *Sedlmeier/Renkewitz, Forschungsmethoden* (Fn. 14).

(sehr starke Verantwortung) anzugeben. Wurde das Los getauscht, lautete die durchschnittliche Antwort in der Bedingung *Mandatory* 4,72, in der Bedingung *Baseline* hingegen 6,35. Die Autoren berichten, dass die Teilnehmer sich in der Bedingung *Mandatory* „signifikant weniger verantwortlich“ fühlen als in der Bedingung *Baseline* ( $p < 0,01$ ).

### 1. Effektgröße

Der quantifizierte Unterschied zwischen den Bedingungen ist die Effektgröße.<sup>124</sup> Grundsätzlich kann sie selbsterklärend sein. Etwa wenn der Unterschied zwischen zwei Bedingungen mit einem Maß beschrieben wird, dass eine allgemeine Bedeutung hat. Zum Beispiel, wenn bei einer binären Entscheidung der Prozentsatz der Teilnehmer angegeben wird, die sich für eine Option entschieden haben. Anders ist dies, wenn der Experimentator ein Maß benutzt, das keine bestimmte Bedeutung in der realen Welt hat. So liegt der Fall in unserem Beispiel, in der das Verantwortungsgefühl auf einer Skala von 1 bis 10 gemessen wurde. In diesem Fall berechnet man die standardisierte Effektgröße. Je nach der Berechnungsmethode wird die bestimmte standardisierte Effektgröße als klein, mittel oder groß interpretiert.<sup>125</sup> Effektgrößen sind für Juristen interessant. Im Falle eines signifikanten aber kleinen Effekts kann bezweifelt werden, inwiefern die Rechtfertigung einer rechtlichen Intervention hierauf gestützt werden kann.<sup>126</sup>

Wie aber ist der p-Wert (auch Signifikanzwert genannt) zu interpretieren? Was ist darunter zu verstehen, dass Ergebnisse statistisch signifikant sind?

### 2. Deskriptive und Inferenz-Statistik

Die Durchschnittswerte, die in dem Artikel berichtet werden, beschreiben nur das Verhalten der Personen, die an dem Experiment teilgenommen haben (beschreibende oder deskriptive Statistik).<sup>127</sup> Es ist deutlich zu erkennen, dass der Durchschnittswert in der Bedingung *Mandatory* kleiner ist als in der Bedingung *Baseline*. Ein Experimentator ist aber nicht nur daran interessiert zu beschreiben, wie sich die Teilnehmer eines bestimmten Experiments verhalten haben. Er möchte generalisieren und Schlüsse über das Verhalten der Population ziehen, also über das Verhalten der Grundgesamtheit, aus der die Stichprobe gezogen wurde (schließende oder Inferenz-Statistik). Dazu „testet“ der Experimentator seine Hypothese, um

124 Die Effektgröße wird auch „Effektstärke“ genannt. Um die Verwechslung mit der Stärke eines Belegs zu vermeiden, sprechen wir von Effektgröße.

125 Bei Mittelwertunterschieden berechnet man z.B. den sog. Cohens d. Bei einem Wert zwischen 0,2 und 0,5 wird der Effekt als klein, bei 0,5 bis 0,8 als mittel und bei >0,8 als groß angesehen.

126 Hierzu Engel, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363 (367): „Wenn der erwartete Erfolg gering ist, könnte die rechtliche Intervention sogar unverhältnismäßig und damit verfassungswidrig sein.“.

127 Mehr hierzu aus juristischer Sicht bei Hamann, Evidenzbasierte Jurisprudenz (Fn. 9), S. 73 ff.

festzustellen, ob sie Aussagen über das Vorhandensein eines bestimmten Effekts in der Population zulässt.

Bei Arlen/Tontrup besagte die Hypothese, dass Menschen in der Bedingung *Mandatory* weniger Verantwortung für die Entscheidung empfinden als in der Bedingung *Baseline*. In der Gruppe der Menschen, die an dem Experiment teilgenommen haben, ist das, wie sich aus den Durchschnittswerten ergibt, tatsächlich der Fall. Inwiefern lässt sich diese Aussage aber auf eine größere Population ausweiten, also als Aussage über das Verhalten von Menschen allgemein verstehen?

### 3. Nullhypothese und Alternativhypothese

Die Inferenzstatistik, die zur Beantwortung dieser Frage eingesetzt wird, verlangt vom Experimentator, zwei sich gegenseitig ausschließende Hypothesen zu spezifizieren: die Nullhypothese und die Alternativhypothese. In unserem Beispiel besagt die Alternativhypothese, dass durch die Vertretung des Verantwortungsgefühls verringert wird, also der durchschnittliche Wert der Antwort auf die Frage nach dem Verantwortungsgefühl in der Bedingung *Mandatory* niedriger als in der Bedingung *Baseline* ausfällt. Die Nullhypothese besagt, dass die Durchschnittswerte in den beiden Bedingungen gleich ausfallen.<sup>128</sup> Die Daten aus dem Experiment werden genutzt, um zu testen, ob diese Hypothese zugunsten der Alternativhypothese verworfen werden kann. In anderen Worten: ob man es anhand der vorhandenen Daten ablehnen darf, dass sich die Teilnehmer in der Bedingung *Baseline* ebenso verantwortlich wie in der Bedingung *Mandatory* fühlen. Dazu führt der Experimentator einen statistischen Test durch. Der Test berechnet die Wahrscheinlichkeit, dass – gegeben, dass die Nullhypothese wahr ist – ein Unterschied zwischen den Durchschnittswerten zu beobachten ist.<sup>129</sup> Diese Wahrscheinlichkeit findet sich im p-Wert wieder.<sup>130</sup> Die statistische Konvention geht dahin, die Nullhypothese zu verwirfen und die Alternativhypothese bestätigt zu sehen, wenn der p-Wert unterhalb von 0,05 liegt.

### 4. Interpretation des p-Werts

Was bedeutet es nun, wenn der p-Wert bei 0,06 liegt? Ist die Nullhypothese wahr und die Alternativhypothese falsch? Das ist ein häufiges Missverständnis. Ein p-

128 Dies ist eine Vereinfachung, die p-Wert und statistische Signifikanz zugänglich machen soll. Da es sich bei der Alternativhypothese in unserem Beispiel um eine gerichtete Hypothese handelt, die über die Richtung des zu erwartenden Unterschieds eine Aussage macht, sollte die richtige Nullhypothese lauten, dass die Teilnehmer in der Bedingung *Mandatory* sich *gleich oder mehr verantwortlich* als in der Bedingung *Baseline* fühlen. Dass Nullhypotesen oftmals tatsächlich schlicht dahingehend formuliert werden, dass kein Unterschied zu beobachten ist, und bei der Alternativhypothese keine Richtung angegeben wird, sind Bestandteile des von G. Gigerenzer, Mindless statistics, The Journal of Socio-Economics (33) 2004, S. 587, scharf kritisierten „Null-Rituale“.

129 Man kann die Überlegung vergleichen mit einer Strafsache, in der man einschätzen muss, wie wahrscheinlich es ist, dass man so starke Beweise beobachtet, obwohl der Angeklagte unschuldig ist.

130 Daher auch das „p“ für probability.

Wert größer als 0,05 bedeutet, dass es nicht genug Belege gibt, um die Nullhypothese zu verwerfen. Auch die Alternativhypothese ist dadurch keinesfalls widerlegt. Ein p-Wert von 0,06 sagt aus, dass, wenn man das Experiment 100 Mal wiederholte, man etwa sechs Mal zufällig den gleichen oder einen größeren Unterschied zwischen den Durchschnittswerten beobachten würde gegeben, dass die Nullhypothese wahr wäre. Je niedriger der p-Wert, desto geringer ist also die Wahrscheinlichkeit, dass der im Experiment beobachtete Unterschied nur zufällig beobachtet wurde. Ein niedriger p-Wert und ein signifikanter Effekt bedeuten aber auch nicht, dass der Effekt groß ist, sondern nur, dass es starke Belege gegen die Nullhypothese gibt.

In der Beispielstudie von *Arlen/Tontrup* bedeutet ein p-Wert von <0,01, dass starke Belege dafür vorliegen, dass die Teilnehmer in der Bedingung *Mandatory* sich nicht bloß zufällig weniger verantwortlich für ihre Entscheidung fühlen als in der Bedingung *Baseline*.<sup>131</sup> Dies lässt den Schluss zu, dass der Befund, im Falle einer Stellvertretung werde weniger Verantwortlichkeit empfunden, auch für die Grundgesamtheit gilt, aus der die Stichprobe der Teilnehmer gezogen wurde.

## VII. Diskussion von Resultaten

In experimentellen Studien diskutieren die Autoren mögliche Anwendungsfelder ihrer Forschung und teils auch die Übertragbarkeit ihrer Ergebnisse auf Sachverhalte außerhalb des Labors. Im Folgenden stellen wir zwei relevante Diskussionsschwerpunkte vor.

### 1. Studenten als Teilnehmer und abstrakte Umgebung

Häufiger Kritikpunkt ist, dass die meisten Versuchspersonen Studenten sind und Experimente nur etwas über deren Verhalten aussagten, noch dazu in einer stark stilisierten Umgebung. Wie wir dargestellt haben, speist sich aber die Leistungsfähigkeit eines Experiments aus eben dieser Stilisierung, schließlich wird es so möglich, Effekte zu identifizieren. Außerdem spielen die Teilnehmer um echtes Geld. Die Beträge, um die es geht, sind – gerade für Studenten – nicht irrelevant. Die hauptsächliche Teilnahme von Studenten ist wenig bedenklich, wenn es darum geht, eine allgemeine Theorie zu testen.<sup>132</sup> Anders liegt dies, wenn eine Aussage über eine spezielle Gruppe getroffen werden soll. Immerhin zeigt aber eine zuneh-

131 Ein „starker Beleg“ liegt damit wohlgernekt nur hinsichtlich eines statistischen Effekts in einem bestimmten Experiment vor. Der Blick auf die Statistik genügt aber nicht für allgemeine Schlüsse über eine Kausalbeziehung. Die Güte der Versuchsanordnung muss mitberücksichtigt werden: Können die Experimentatoren Störvariablen und andere mögliche Erklärungen für ihre Resultate ausschließen?

132 *Falk/Heckman, Lab Experiments Are a Major Source of Knowledge in the Social Sciences*, Science (326) 2009, S. 535 (537): „For the purpose of testing theories, this is not a problem because most economic models derive predictions that are independent of assumptions concerning participant pools.“

mende Zahl an Studien, dass die Ergebnisse aus Experimenten mit Studenten durchaus den Ergebnissen aus Experimenten mit bestimmten spezifischen Berufsgruppen ähneln.<sup>133</sup> Dennoch wäre es wünschenswert, den Beteiligtenkreis zu erweitern.<sup>134</sup> Online-Experimente erscheinen hier als veritable Alternative oder Ergänzung.<sup>135</sup> Diese bieten zwar weniger Kontrolle über das Verhalten der Teilnehmer, geben aber Zugang zu einem größeren und vielfältigeren Teilnehmer-Pool.

## 2. Mehrere Studien und juristisches Augenmaß

Man sollte sich der Grenzen der experimentellen Methode bewusst sein. Diese Grenzen lassen sich ausweiten, wenn man für rechtliche Schlussfolgerungen die Resultate einer Reihe von Studien heranzieht, die nicht nur Befunde replizieren, sondern diese auch unter anderen Bedingungen bestätigen.<sup>136</sup> Idealerweise werden experimentelle Ergebnisse durch Resultate ergänzt, die mit anderen Methoden generiert wurden (etwa mit den eingangs erwähnten Feldexperimenten). In der Diskussion all dieser Resultate ist dann aber wiederum juristisches Augenmaß und auch eine gewisse Skepsis gefragt.<sup>137</sup> Es gilt, die Bedingungen, unter denen ein Ergebnis generiert wurde, mit den Umständen des realen Problems zu kontrastieren.<sup>138</sup> Zu diesen Bedingungen gehören sowohl die theoretischen Prämissen, als auch deren experimentelle Umsetzung.

Aus juristischer Sicht bietet sich an, experimentelle Resultate im Lichte juristischer Dogmatik zu diskutieren. Schließlich bietet diese einen Zugriff auf systematisiertes Erfahrungswissen.<sup>139</sup> Auch die Beurteilung, ob ein experimentell identifizierter Mechanismus für das vorliegende juristische Problem relevant ist, beruht letztlich auf

<sup>133</sup> G. Charness/P. Kuhn, Lab Labor: What Can Labor Economists Learn from the Lab?, in: A. Orley/C. David (Hg.), *Handbook of Labor Economics* 2011, S. 229 (254).

<sup>134</sup> Falk/Heckman, Lab Experiments (Fn. 132), S. 535 (537): „On some aspects of behavior, however, students are not representative of the general population or a target population of interest, and we would agree that a richer variation in context, populations, and environments should be used in future lab experiments.“.

<sup>135</sup> Kombination von Labor- und Online-Experiment etwa bei Arlen/Tontrup, Does the Endowment Effect Justify Legal Intervention? (Fn. 68).

<sup>136</sup> Meta-Studien wiederum verarbeiten eine Vielzahl an Studien zu einem bestimmten Thema, siehe zum Beispiel C. Engel, Dictator Games: A Meta Study, *Experimental Economics* (14) 2011, S. 583.

<sup>137</sup> Vgl. Engel, Verhaltenswissenschaftliche Analyse (Fn. 7), S. 363 (366): „Aber wenn auch nach der sorgfältigen Lektüre der einschlägigen psychologischen oder ökonomischen Veröffentlichungen Zweifel bleiben, sollte der Jurist keine normativen Entscheidungen auf den Befund stützen.“.

<sup>138</sup> Hellwig, Neoliberale Sektierertum oder Wissenschaft? (Fn. 43), S. 10, plädiert im Kontext ökonomischer Politikberatung für verstärkte kritische Überprüfung der Annahmen, die Modellen und Experimenten zugrunde liegen: „Im Zentrum der Annahmenkritik muss die Frage stehen, welche Rolle die Annahmen für die Schlussfolgerungen spielen bzw. inwiefern die Schlussfolgerungen gegenüber Verallgemeinerungen robust sein könnten.“.

<sup>139</sup> Kritisch zum Stand der Dogmatik, aber mit Hinweisen auf ihre Möglichkeiten O. Lepsius, Themen einer Rechtswissenschaftstheorie, in: M. Jestaedt/O. Lepsius (Hg.), *Rechtswissenschaftstheorie*, Tübingen 2008, S. 1 (18).

juristischem Expertenwissen. Die normativen Folgerungen, die an ein Experiment geknüpft werden, sind dem juristischen „Hoheitsgebiet“ zuzurechnen.<sup>140</sup>

### VIII. Fazit

Das eingangs erwähnte Problem der externen Validität, also der Frage, inwiefern Befunde aus dem Labor etwas über die Welt außerhalb des Labors aussagen, hat sich auch am Ende unserer Betrachtung nicht in Luft aufgelöst. Das bessere Verständnis für die experimentelle Forschung erlaubt es aber, die Frage der externen Validität von Experimenten für bestimmte Anwendungsfälle besser zu bewerten. Es ist eine Ausgangsbasis geschaffen, um über die Übertragbarkeit des experimentellen Resultats nachzudenken. Dies ähnelt einem juristischen Analogieschluss.<sup>141</sup> Ein rechtlicher Analogieschluss setzt eine vergleichbare Interessenlage voraus, also dass zwei unterschiedliche Fälle sich in einer Weise gleichen, die es rechtfertigt, eine Norm anzuwenden, die für den ersten Fall geschaffen wurde.<sup>142</sup> Die Frage, inwiefern die Befunde aus dem Labor auf die Welt außerhalb des Labors übertragbar sind, stellt sich in einer ähnlichen Weise. Ihre Beantwortung setzt einen sorgfältigen Vergleich zwischen der Versuchsanordnung und dem tatsächlich interessierenden Sachverhalt voraus. Wir hoffen, mit unserem Beitrag diesen Vergleich ein Stück weit einfacher gemacht zu haben.

140 So auch *Hamann*, Evidenzbasierte Jurisprudenz (Fn. 9), S. 89.

141 Den Vergleich zwischen dem Vorgehen bei einem Analogieschluss und der Diskussion der externen Validität eines Experiments verdanken wir einer Bemerkung *C. Engels*.

142 Zur Beurteilung der Gleichheit zweier Fälle *I. Puppe*, Kleine Schule des Juristischen Denkens, Göttingen 2008, S. 99: „eigentliche[r] Knackpunkt jedes Analogieschlusses“; siehe auch *R. Alexy*, Theorie der juristischen Argumentation, Frankfurt am Main 2008, S. 345, der die Analogie als formale Struktur *inhaltlicher Argumente* versteht, „die erst in dieser Struktur ihre Wirkung entfalten können“.