Trust and ban
– a critical reassessment of debates on the regulation of innovations in democracies

Beatrice Brunhöber, Bernhard Jakl

A common belief, which also underpins the EU's current digital strategy, is that trust in normative orders can be fostered through the imposition of bans. The prevailing approach, informed in legal theory and in theoretical sociology, which we call a "communicative picture" tends to support such notions. This approach considers bans as a special form of communication that stabilizes expectations and thus generates trust, or at least functions as a latent reason for correct behavior. In contrast, our critical argument, derived from legal philosophy, is that the relationship between trust and ban varies in different fields of law. The transition to an institutional-argumentative justification of norms proposed here allows to critically reassess the questionable nexus of trust and ban.

The need for trust is increasingly apparent in today's intricate mass society. Individuals can only engage in action and collaboration within such a society if they place trust in others, even without knowing their identities or intentions. Economic, organizational, and technical complexities can only be navigated through trust in institutions such as the market, organizations, and large-scale technologies. Contrary to initial impressions, trust within modern, complex mass society is fundamentally strengthened through legal coercion, a fact particularly reflected in various forms of legal ban.

The utilization of bans is growing in significance for fostering trust in innovations. Legal bans frequently serve the purpose of instilling trust in innovations by shielding individuals from undesirable outcomes stemming from their use. For instance, the European Artificial Intelligence Act (Draft AI Act)[1], adopted by the EU Parliament in March 2024, categorizes AI ap-

1 Artificial Intelligence Act proposed by the European Council (21 April 2021), adopted by European Parliament (13 March 2024), awaits reading in the EU Council, COM(2021) 206 final (Draft AI Act), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206> accessed 26 April 2024.

plications based on risk and prohibits AI practices (such as social scoring) deemed to pose unacceptable risks. According to the EU Commission, the Act was formulated to "ensure that Europeans can trust what AI has to offer."[2]

In the realm of legal philosophy, the outlined discussions regarding trust in innovations within complex mass societies shed light on the fundamental interplay between trust and ban. This interplay is characterized by bans reacting to and at the same time shaping innovations.

This contribution provides a critical (re-)assessment of debates surrounding the regulation of innovations in democracies. Firstly (1.), we examine the widely accepted – as we put it – *communicative picture* of the relationship between trust and ban, as portrayed by legal theory and theoretical social science: the notion of it primarily functioning as a form of communication. In contrast, the subsequent two sections adopt a perspective rooted in the philosophy of law, offering an alternative picture of the relationship between trust and ban based on the specific legal doctrines developed within different areas of law: here we explore the argumentative standards of criminal law (2.) and private law (3.). In the final section (4.), we conclude that the delineated *institutional-argumentative picture* of the relationship between trust and ban, based on legal philosophy, holds greater appeal than the traditional communicative picture, as it exhibits more significant critical potential.

A. The communicative picture of the relationship between trust and ban

The interplay between trust and ban in the context of democratic control over innovations is conventionally examined through the lens of legal theory and theoretical sociology, often grounded in either the concept of communicative justification discourses or the concept communicative systems. We may therefore call this type of – otherwise very different – conceptions a "communicative picture" of the relationship between trust and ban.

2 See the Digital Strategy of the European Institutions, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 5 May 2024.

I. Adapting law to evolving realities

The evaluation of bans, emphasizing their communicative mediations and justifications, sheds light on the intricate relationship between a normative surplus generated through communication and the imperative of adapting to (normative) reality for generating trust. On one hand, approaches rooted in discourse theory often advocate for ban in pursuit of a normatively excessive conception of democracy and justice.[3] For example, scholars respond to the emergence of "surveillance capitalism" by advocating for democratic control over information.[4] Conversely, systems theory approaches aim at systemic adaptations.[5] Building upon these premises, authors argue that legal subjectivity ought to be dissociated from personal

3 For the realization of his concept of justice see John Rawls, *A Theory of justice* (revised edn, Cambridge, MA: The Belknap Press of Harvard University Press 1999) 5, 22, 113, 156 on defining fundamental rights and obligations and emphasizing the institutional framework. According to this, only the enforcement of a public system of penalties by the government removes the presumption that others do not obey the rules, ibid. 209. On the implementation of the equal originality of private and public autonomy, see Jürgen Habermas, *Between Facts and Norms. Contributions to a discourse theory of law and democracy*, transl. by William Rehg (Cambridge, MA: MIT Press 1996), 399-404, according to which the aim is to abolish privileges that are incompatible with the equal distribution of subjective freedoms demanded by this principle. Freedom thus depends essentially on state activities and direct specifications justifying in principle a priority of public over private autonomy. For *democratic trust* see below 2.3. (text to n 36) and Russel Hardin, Trust and Trustworthiness (New York: Russel Sage 2002), 151-172; Pippa Norris, "The conceptual Framework of Political Support" in Sonja Zmerli and Tom WG van der Meer (eds), Handbook on Political Trust (Cheltenham: Edward Elgar Publishing 2017) 19-32; Pippa Norris, In Praise of Skepticism. Trust but Verify (New York: Oxford University Press 2022); Mark E. Warren, "Trust and Democracy" in Eric M. Uslaner (ed), The Oxford Handbook of Social and Political Trust (Oxford: Oxford University Press 2018) 75-94; for the development in democratic theory see Beatrice Brunhöber, Die Erfindung "demokratischer Repräsentation" in den Federalist Papers (Tübingen: Mohr Siebeck 2010), 136-144.

4 E.g. Shoshana Zuboff, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power* (London: Profile Books 2019) 366-371, to overcome the danger of "instrumentarian power" and Carol Gould, "How Democracy can Inform Consent: Cases of the Internet and Bioethics" (2019) 36 (2) *Journal of Applied Philosophy* 173, for a democratic revision of consent by means of an "all-affected-principle".

5 Niklas Luhmann, *Das Recht der Gesellschaft* (Frankfurt a.M.: Suhrkamp 1993) 277, according to whom stabilization can serve as a motivation for innovation. Therefore, Luhmann concludes, social theory has to change from "target formulas such as peace and justice to system analysis", ibid 438.

conceptions and that legal frameworks should be adjusted accordingly (e.g., by granting legal capacity to software agents).[6]

II. Bans as recognized political instrument

Given the prevalent acceptance of the communicative picture of the relationship between trust and bans, the latter are commonly regarded as effective tools to forestall undesirable outcomes of innovations in their early stages, thereby fostering trust in the respective innovation.[7] This highlights a unique interaction between bans and trust. Bans respond to the ongoing progression of innovations and their anticipated impacts on individuals and society. Concurrently, bans influence the future development of the regulated innovations, both technically and socially, shaping aspects such as the types of applications brought to market and their modes of utilization.[8]

III. The relationship between trust and ban

Within communicative approaches, there is controversy over whether modern law relies not only on communicative mediation[9] but also necessitates trust in public discourse for resolving social conflicts.[10] Irrespective of this controversy, the question remains to what extent bans must be issued, justified, and shaped to assert and foster socially effective trust within a pluralistic society of free individuals. From these communicative standpoints, legally institutionalized bans and the associated coercion not only mitigate uncertainty regarding the actions of others. Rather, institutionally enforce-

6 E.g. Gunther Teubner, "Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten" (2018) *Archiv für die civilistische Praxis* 155, 204.

7 E.g. from a European legal perspective Irina Orssich, "Das europäische Konzept für vertrauenswürdige Künstliche Intelligenz" (2022) *Europäische Zeitschrift für Wirtschaftsrecht* 254.

8 Cf. Wolfgang Hoffmann-Riem, *Innovation und Recht – Recht und Innovation. Recht im Ensemble seiner Kontexte* (Tübingen: Mohr Siebeck 2016) 698: The legislator should open up "options spaces" within which further development can take place.

9 Luhmann (n 5) 128 refers in this respect to repetitions of communication acts that constrain the scope for alternatives and thus have a stabilizing effect. Also Niklas Luhmann, *Soziale Systeme. Grundriß einer allgemeinen Theorie* (Frankfurt a.M.: Suhrkamp 1987) 498.

10 Habermas (n 3) 16 and 80 for the special case of legal argumentation.

able bans also facilitate the resolution of negative cooperative experiences by stabilizing expectations. According to the viewpoint of theoretical sociology, bans can ultimately cultivate trust.[11] In terms of justification, bans in this context are perceived less as instruments of power and more as means of communication.[12]

From this perspective, the potential for sanctions accompanying bans cannot be the main driver for cultivating trust,[13] as trust relies on communication. If sanctions take precedence, the individuals involved may no longer extend trust, nor is it necessary for them to do so. Instead, sanctions enforce proper behavior, not trust. The availability of sanctions has a limited impact, primarily influencing the motivation of the trustees and incentivizing them to act in a trustworthy manner in their own self-interest. The availability of sanctions afforded by bans thus serves as a subsidiary reason and possesses a latent function.[14]

From this justification-oriented perspective, bans should be measured less in terms on whether they are institutionally enforced and more in terms of whether they are justified in a manner that renders them perceived as binding by the individuals and groups they target. In the context of democracies, in line with discourse theory considerations, such perception is more likely to occur if individuals also perceive themselves as the creators of laws and can trust legislation and its application to be fair under the rule of law.[15]

11 Niklas Luhmann, *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität* (5th edn, Konstanz: UVK Verlagsgesellschaft mbH 2014) 27-38.

12 E.g. Klaus Günther, "Zwang/Sanktion und Vertrauen im Konflikt" (ConTrust Working Paper Series, manuskript 2020) 8-13.

13 Luhmann, (n 11) 39 sees trust as a kind of deception about the complexity of the world. Also see Georg Simmel, *Soziologie. Untersuchungen über die Formen der Vergesellschaftung* (10th edn, Frankfurt a. M.: Suhrkamp 1992) 263 who describes trust as "hypothesis of future behavior" ("Hypothese künftigen Verhaltens", translation by the authors).

14 Luhmann (n 11) 35 and Günther (n 13) by drawing on Joseph Raz's differentiation between operative and auxiliary reasons. According to Raz, *Practical Reasons and Norms* (Oxford: Oxford University Press 1999) 32-34, a reason is an operative reason if the belief in its existence implies that one adopts a practical critical attitude. A reason that is not an operative reason, on the other hand, is what Raz calls an auxiliary reason.

15 See Habermas (n 3) 119-120; Rawls (n 3) 52, 53 and 131.

IV. Open questions

The communicative picture delineated by legal theory and theoretical sociology underscores essential facets of the relationship between trust and ban. However, on the basis of legal philosophy, the question arises of whether the narrow emphasis on communication neglects the argumentative standards developed within the institutionally distinct areas of law. To address this question, the subsequent two sections scrutinize these institutional-argumentative standards concerning the relationship between trust and ban within the domain of criminal law on one hand and private law on the other.

B. Criminal law and innovations: from ban to trust

In this section we investigate the correlation between trust and ban in criminal law, illustrated through the lens of criminal liability pertaining to actions associated with innovations such as algorithms and/or artificial intelligence (AI).

I. Bans in criminal law: Penalizing conduct in the context of innovations

Criminal liability for such conduct typically falls under cybercrime legislation. This is because the majority of algorithms and/or AI applications are utilized within computer programs. Under the prevailing definition, cybercrime includes the utilization of a digital device, such as a computer, as an integral part of committing a crime or making a computer system the object of the crime.[16] A significant portion of global cybercrime regulation[17]

16 See e.g. Budapest Convention on Cybercrime (adopted 23 November 2001, entered into force 1 July 2004) 2296 UNTS 167 (Convention on Cybercrime) art. 1 (a).

17 As of 1 May 2024, 68 countries have signed the Convention on Cybercrime, that is by all Council of Europe members as well as Canada, Japan, the United States and South Africa as well as Australia and quite a few further countries from Africa (e.g. Senegal, Ghana), Asia (e.g. Philippines) and South America (e.g. Argentina, Brazil, Chile, Columbia). It was estimated in 2017 that the Convention had already influenced the cybercrime regulation of more than 130 countries due to its policy measurements all over the world, see Alexander Seger, in Roderic Broadhurst et al. (eds.), *Cyber Terrorism* (Research Report of the Australian National University Cybercrime Observatory for the Korean Institute of Criminology 2017) Fig. 7.1.; also

is either harmonized or influenced by the Council of Europe Convention on Cybercrime[18] and, within the European Union, by corresponding Framework Decisions and Directives.[19] The Convention mandates that state parties criminalize various forms of conduct categorized as cybercrime and encourages interstate cooperation in law enforcement.[20] Its global impact is further augmented by capacity building initiatives in non-signatory states. The Convention advocates for the criminalization of access offenses (e.g. hacking), use offenses (e.g. cyberfraud), and content offenses (e.g. hate speech or child pornography).[21] Presently, many traditional cybercrimes are perpetrated through the use of algorithms and/or AI.[22] For instance, hacking may involve employing reverse engineering. Fraudulent phishing emails may be crafted utilizing machine learning to evade spam filters. Hate speech may be disseminated through social bots. Certain instances of child pornography may be produced using AI. This list could easily go on.

II. From ban to trust: Cultivating trust by pre-empting future risks associated with innovations

Criminal prohibitions in the context of innovations primarily seek to bolster trust in the utilization of new technologies by pre-empting potential future risks associated with the innovations. The drafting of the Cybercrime

see Neil Boister, *An Introduction to Transnational Criminal Law* (2th edn, Oxford: Oxford University Press 2018) 189; critical of the scope Marco Gercke, "10 years Convention on Cybercrime. Achievements and Failures of the Council of Europe´s Instrument in the Fight against Internet-related Crimes" (2011) 5 *Computer Law Review International* 142-43.

18 Draft AI Act (n 1).

19 Especially 2013/40/EU of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA; 2011/92/EU of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography and replacing Council Framework Decision 2004/68/JHA.

20 Convention on Cybercrime (n 16), chap. II, sec. 1 (substantive criminal law) and sec. 2 (procedural law).

21 Title I, II and III of the Convention on Cybercrime (n 16) and 2003 Additional Protocol concerning the criminalization of acts of a racist or xenophobic nature committed through computer systems (adopted 28 January 2003, entered into force 1 July 2004) 2466 UNTS 205.

22 See e.g. for hate speech with social bots Sabine Gleß and Thomas Weigend, "Intelligente Agenten und das Strafrecht" (2015) 123 (3) *Zeitschrift für die gesamte Strafrechtswissenschaft* 561.

Convention commenced in 1997,[23] a period when computers, the Internet, and digital devices such as mobile phones had yet to assume significant roles in daily lives of most individuals.[24] The objective of the Cybercrime Convention was to establish global control of cyberactivity in order to mitigate nascent risks to commerce, businesses, private communications and public institutions at an early stage.[25] These risks are associated with factors such as the widespread use of digital devices, which expands the potential number of affected individuals, the availability of anonymity and encryption options, which may incentivize engaging in particularly risky behavior and may be used for concealing responsibility for the commitment of a crime, and the transnational nature of cybercrime, hindering investigation, prosecution and adjudication processes.[26] Addressing these challenges is intended to facilitate individuals' ability to securely share data via cloud computing, communicate via email or to conduct banking transactions online without running the risk of exploitation or compromise. In essence, bans of (presumed) risky cyberactivity and corresponding law enforcement measures are generally aimed at fostering trust in cyberspace.

With the emergence of AI, concerns about mitigating anticipated risks associated with its utilization arose early on.[27] These concerns culminated in the Draft AI Act[28], marking the world's inaugural major legislation aimed at regulating AI to instil trust in its application.[29] The Draft AI Act governs

23 See Ryan M. F. Baron, "A critique of the International Cybercrime Treaty" (2002) 10 (2) *CommLaw Conspectus* 263, 265. In 1997 a Committee of Experts on Crime in Cyber-Space was set up by the Council of Europe (Specific Terms of Reference of the Committee of Experts on Crime in Cyber-Space, Council of Europe's Fight Against Corruption and Organised Crime, sec. 5 (c) 583rd Meeting) which eventually drafted the Convention on Cybercrime (n 16). The Convention was opened for signature in 2001 and came into force in 2004.

24 Jonathan Clough, "The Council of Europe Convention on Cybercrime" (2012) 23 *Criminal Law Forum* 363, 365.

25 For an overview of presumed damages from cybercrime see Nir Kshetri, *The Global Cybercrime Industry: Economic, Institutional and Strategic Perspectives* (Berlin: Springer 2010) 4-6.

26 Cf. Marco Gercke and Philipp Brunst, *Internetstrafrecht* (2th edn, Stuttgart: Kohlhammer 2023) para. 10; UNODC, *Comprehensive Study on Cybercrime* (Vienna: UN 2013) 226.

27 Thomas C King, Nikita Aggarwal, Mariarosaria Taddeo, Luciano Floridi, "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions" (2020) 26 (1) *Sci Eng Ethics* 89-120.

28 Draft AI Act (n. 1).

29 <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 5 May 2024.

the entry of specific AI products into the EU internal market.[30] This approach has sparked apprehension over whether certain particularly harmful AI activities should instead be subjected to EU-wide criminalization. Examples include the penalization of deepfakes, i.e. the use of applications to produce AI-manipulated political information or to create sexually explicit AI-fabricated images humiliating others.[31]

Typically, criminal bans seem to serve as notably effective methods for shielding individuals from undesirable outcomes and thereby bolstering trust in innovation.

III. Reassessing the relationship between trust and ban in criminal law regulation of innovations

The communicative picture of trust intersects with the developed relationship between trust and ban in a crucial domain: According to system theories, the penalties facilitated by corresponding criminal statutes can stabilize behavioural expectations[32] within the realm of innovations. They are intended to foster trust in the conduct of others when engaging with computers, the Internet, algorithms and/or AI. This outcome remains valid even when shifting the focus from the availability of sanctions to the norms permitting sanctions, as discourse theories suggest.[33] The efficacy of prohibitions then relies less on their enforcement and more on whether they are perceived as binding, a condition that is met when individuals see themselves as their authors.[34] Criminal provisions concerning innovation typically emerge from a (national) democratic process often implementing

30 See for the consequences for criminal product liability Victoria Ibold, *Künstliche Intelligenz und Strafrecht: zur strafrechtlichen Produktverantwortung in der Innovationsgesellschaft* (Baden-Baden: Nomos Verlagsgesellschaft 2024).

31 In early 2024 criminalization of sexually explicit deepfakes was introduced by both the UK ministry of Justice (Guardian, 16 Apr. 2024, <https://www.theguardian.com/technology/2024/apr/16/creating-sexually-explicit-deepfake-images-to-be-made-offence-in-uk> accessed 10 May 2024) as well as by the European Union (Directive of the European Parliament and of the Council on combating violence against women and domestic violence, PE-CONS 33/24 of 25 April 2024 [not yet published in the Official Journal], (19) and art. 5 (1)(b)).

32 Luhmann (n 11) 27-38; see above 2.3.

33 Günther (n 12); Raz (n 14); see above 2.3.

34 Habermas (n 3); Rawls (n 3); see above 2.3.

supra- and international guidelines[35] and thus must also adhere to the respective fundamental principles outlined in national constitutions. These include the requirements for democratic self-determination as well as human rights and principles such as the rule of law.

However, from the vantage point of the communicative picture, another aspect is somewhat overlooked, which can be highlighted more effectively from an institutional-argumentative perspective informed by the philosophy of law. The communicative picture of trust tends to underemphasize the fact that criminal law not only pertains to the trust relationship between citizens which must be upheld through criminal sanctions wielded by sovereign authority, but also encompasses the trust relationship between citizens and the sovereign authority itself. This relationship only comes into view in discussions about trust in government or democratic institutions.[36] From an institutional-argumentative standpoint grounded in the philosophy of law, it becomes evident that criminal law prohibitions serve as a direct mechanism from authority to control individual behavior. As observed, legislation on innovations often seeks to control individual activities in a manner that instils trust in the respective innovation.

The main objective is to avert future risks associated with the innovation, leading to a significant expansion of criminal law.[37] Firstly, unlike other domains of criminal law, regulations concerning innovations are frequently justified by the use of "risky" tools or the risk posed to targeted vulnerable objects. For instance, the Cybercrime Convention advocates for criminalizing the mere possession of hacking tools,[38] implying that they could be used

35 Beatrice Brunhöber, "Criminal Law of Global Digitality. Characteristics and Critique of Cybercrime Law" in Alexander Peukert, Matthias Kettemann, Indra Spiecker gen. Döhmann (eds.), *Law of Global Digitality* (London: Routledge 2022) 246-47; Allen Buchanan, 'The Legitimacy of International Law' in Samantha Besson and John Tasioulas (eds), *The Philosophy of International Law* (Oxford: Oxford University Press 2010) 79.

36 See Hardin (n 3) 151-172; Norris (2017, n 3) 19-32; Norris (2022, n 3); Warren (n 3) 75-94; for the development in democratic theory see Brunhöber (n 3) 136-144.

37 With regard to the following see Andrew Ashworth and Lucia Zedner, *Preventive Justice* (Oxford: Oxford University Press 2014) 95-118; Beatrice Brunhöber, "Von der Unrechtsahndung zur Risikosteuerung durch Strafrecht und ihre Schranken" in Roland Hefendehl et al (eds), *Festschrift für Bernd Schünemann* (Berlin: De Gruyter 2014) 3-15.

38 Convention on Cybercrime (n 16) art. 6 (1)(b).

in a detrimental manner.[39] Consequently, criminalization is not grounded in the violation of specific rights and legal concerns but rather alludes to some form of ambiguous risk. Secondly, given the objective of preventing any risks, criminal law regulation of innovations frequently necessitates penalizing behavior that facilitates harmful or dangerous conduct, enabling law enforcement to intervene at an early stage. Criminalizing the mere possession of hacker tools eliminates the necessity for evidence of actual computer system access to initiate an investigation. The presence of hacking tools on the suspect's computer alone suffices as evidence. Thirdly, owing to the goal of preventing any risks, corresponding offenses often do not require an intent to cause harm or substantive actions towards that end.[40] For instance, the Convention on Cybercrime calls for criminalizing "computer hacking" without requiring additional elements of a crime, such as breaching security measures.[41] Finally, unlike other domains of criminal law, penalization within the context of innovations often covers "neutral" every day behaviours that are deemed risky when undertaken with malicious intentions.[42] This broadens criminal liability from rare exceptional circumstances to embrace everyday life. For example, given that cybercrime regulation, in terms of its structure (computer systems as a tool or objective), potentially affects any use of information technology, many users are uncertain whether their actions fall under its purview (e.g. sharing music and movies, taking part in online protests via distributed denial of service attacks, and sharing explicit content images). At best, this uncertainty leads to indifference to the relevant offences; at worst, it induces self-restraint (a chilling effect).[43]

The trend toward expanding criminal law runs counter to the foundational principles of the criminal law system: Despite varying opinions on the specifics, legal scholars generally concur that the application of criminal law should be highly restrained. Moreover, democratic constitutions typically include specific provisions for criminal law to circumscribe its

39 Brunhöber (n 35) 245-46; see Andrew Ashworth, *Positive Obligations in Criminal Law* (Oxford: Hart Press 2013), 149-172 generally criticizing the "unfairness of risk-based possession offences".

40 Brunhöber (n 35) 246.

41 Convention on Cybercrime (n 16) art. 2. The parties to the Convention may include further elements of crime, but are not obliged to do so.

42 Cf. the debate on criminal liability for neutral assistance, e.g. Marcus Wohlleben, Beihilfe durch äußerlich neutrale Handlungen (Munich: CH Beck 1997) 7-10.

43 Neil Richards, *Why Privacy Matters* (Oxford: Oxford University Press 2022) 129.

scope (e.g. nulla poena sine lege, nulla poena sine culpa).[44] Criminal law represents an exceptionally severe, if not the most severe, instrument of sovereign authority: It not only authorizes monetary penalties (fines) but also entails deprivation of liberty (imprisonment) or even the loss of life (capital punishment, as in certain US-states). Furthermore, criminalizing particular behaviours signifies deeming them public wrongs (e.g. criminal records leading to job exclusion from crime-related professions, e.g. disqualification from teaching roles due to a history of child abuse), establishing a severe threat of an evil in order to give a pragmatic reason for not doing it, and to censure those who break the law.[45] Finally, criminalization grants law enforcement the power to conduct searches, surveillance, detentions, interrogations, and so forth. The exercise of such powers, which have significant consequences, necessitates a high standard of justification. That entails democratic decision-making regarding criminal provisions as well as theoretical justification based on substantial reasons for establishing such a rigorous control regime over individuals.[46] Regardless of the respective, quite different theoretical context, it is widely acknowledged that criminalization cannot be warranted solely by an imminent risk; rather it is essential that the penalized conduct causes harm to others (the harm principle[47]), violates legal interests (Rechtsgutstheorie[48]), or infringes upon concerns that outweigh individual liberty.[49] Consequently, criminalizing

44 E.g. nulla poena sine lege in art. 103 (2) German Basic Law (Grundgesetz); nulla poena sine culpa founded in human dignity (art. 1 (1) German Basic Law) or as prerequisite of the presumption of innocence (art. 6 (2) European Convention on Human Rights).

45 Andrew Ashworth and Jeremy Horder, *Principles of Criminal Law* (7th edn, Oxford: Oxford University Press 2013) 22-23.

46 Ibid 23.

47 John Stuart Mill, *On Liberty* (Harmondsworth Middlesex: Penguin Books 1979); Joel Feinberg, *Harm of Others* (New York: Oxford University Press 1984) 26.

48 Winfried Hassemer, „Grundlinien einer personalen Rechtsgutslehre (1989)" in Winfried Hassemer, *Strafen im Rechtsstaat* (Baden-Baden: Nomos Verlag 2000) 160, 167; first Winfried Hassemer, *Theorie und Soziologie des Verbrechens* (Frankfurt a.M.: Athenäum-Verlag 1973), 147, 221; Claus Roxin and Luis Greco, *Strafrecht Allgemeiner Teil*, vol. 1 (5th edn, Munich: CH Beck 2020) sec. 2 para. 7; first Claus Roxin, "Sinn und Grenzen staatlicher Strafe" (1966) *Juristische Schulung* 377, 381.

49 Beatrice Brunhöber, "Was ist freiheitlich-demokratische Strafrechtsbegrenzung? Stärkung des Blicks der Kriminalisierungstheorien für die Freiheit der Verbotsadressierten" in Beatrice Brunhöber, Christoph Burchard, Klaus Günther et al. (eds), *Strafrecht als Risiko, Festschrift für Cornelius Prittwitz* (Baden-Baden: Nomos Verlag 2023) 59-75; Antony Duff, *Answering for Crime. Responsibility and Liability in the Criminal Law* (Oxford and Portland: Hart Publishing 2007), 141-42.

conduct cannot be justified by merely alluding to potential risks. Criminalization can only be justified by at least the prospect of harm to others or violations of legal interests. The justification process thus necessitates precise identification of the rights and concerns that may be impacted by certain behaviours and their penalization.

C. Private law and innovations: from trust to ban

This third part explores the relationship between trust and ban in private law using as an example AI systems identified as a growth market for private algorithmic-based businesses and their regulation.

I. The approach of EU institutions: creating trust in the digital world through bans

In the field of innovations, the current risk-differentiated normative proposals from the EU Commission and the European Parliament aim to create trust in AI systems. The European legal framework is designed to ensure the reliability of AI systems, referred to as "trustworthy AI".[50] For instance, Article 11 of the Draft AI Act requires technical documentation and compliance assessment procedures for high-risk AI systems, while Article 14 stipulates human oversight and Articles 30-39 require notification procedures. This Draft AI Act is complemented by a Draft AI Liability Regulation,[51] which seeks to establish standards of liability beyond existing national private law. These standards are to correspond to the risks identified as inherent to the AI system by preventive technical prognosis according to Articles 8, 3, 4 and 5 of the AI Liability Regulation Draft. The implicit and explicit claim of these legislative initiatives is to create trust by ex-ante bans. This raises the question of the role of bans in private law.

50 Draft AI Act (Fn. 1).

51 European Parliament, Report with recommendations to the Commission on a civil liability regime for artificial intelligence, 20 Oct. 2020, P9_TA-PROV(2020)0276, <https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_EN.html> accessed 26 April 2024, followed by COM/2022/496 final, a Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (Draft AI Liability Directive), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496> accessed 26 Apr. 2024.

II. Where to find normative experiences with ban in existing private law?

In the realm of contractual agreements, which can also be assessed through the lenses of tort and unjust enrichment law, it becomes apparent, that private autonomy is restricted. This applies particularly to mass transactions governed by the law on general terms and conditions. This occurs both in public debates surrounding innovations and through legislative revisions, often converging on an ambiguous notion of contractual fairness[52] or intricate risk disclosure and liability assignments, notably extending to product liability law.[53]

The most recent instances of mandatory legislative adjustments within the detailed contract law framework in German law entail new conceptual classifications, stemming partly from European law imperatives for digitalization.[54] These have been incorporated into the German law of obligations through sec. 327a-q German Civil Code (Bürgerliches Gesetzbuch, BGB) for package contracts and contracts for goods with digital elements, alongside a revised concept of deficiencies for digital products (sec. 434, 475b et seq. BGB). Nevertheless, from these individual rules, characterized as "specific measure acts" (Maßnahmegesetze), it is hardly possible to discern foundational normative experiences applicable for identifying a general standard. Nevertheless such a standard, independent of the contingencies of a business model under consideration in each instance, is essential for establishing trust through bans within the domain of innovations.

In both civil law and common law systems, contracts remain binding based on generally accepted principles, except where they contravene principles of good morals, bona fide protections, public order or other mandatory regulations.[55] The determination of what constitutes a breach of good

52 See Heike Schweitzer, "Digitale Plattformen als private Gesetzgeber: Ein Perspektivwechsel für die europäische ‚Plattform Regulierung'" (2019) *Zeitschrift für Europäisches Privatrecht* 2019, 1, 8 and 12 uses the concept „Richtigkeitsgewähr" (assurance of correctness) even as an alternative concept for private autonomy.

53 E.g. from a German perspective Gerhard Wagner, "Liability Rules for the Digital Age – Aiming for the Brussels Effect" (2022) *Journal of European Tort Law* 191.

54 On the implementation and an overview on some consequences of the der Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning "contracts for the supply of digital content and digital services" into national law in the case of the German Civil Code (BGB) see Thomas Riehm, "Verträge über digitale Dienstleistungen" (2022) *Recht Digital* 209.

55 See as an example instead of multiple national norms art. 4:109 Principles of European Contract Law (PECL) and art. 4:110 PECL.

morals, bona fide protection, or public order, thereby permitting deviation from a contractual agreement as an exception, may vary between legal systems and occasionally change over time.[56]

In German law, for example, there exists a specific provision incorporating a general clause on good morals, as stipulated in sec. 138 BGB, as in Austrian Law with sec. 879 General Civil Code (Allgemeines Bürgerliches Gesetzbuch), in French Law with Art. 1131 and 1133 Code civil, and in Swiss law with Art. 20 Code of obligations (Obligationenrecht). Art. 138 BGB is open to interpretation and holds significant promise for examining normative experiences concerning the relationship between trust and ban. The legal concept of good morals outlined in sec. 138 BGB imposes certain constraints on all contractual agreements, some of which are not explicitly made positive law. Examples include adhesion contracts, usury, or contracts relating to organ donation and surrogate motherhood. The legal consequence of the nullifying the contractual agreement is prescribed here, rendering the contract unenforceable as well.

Although subject to debate, sec. 138 BGB can be understood structurally as a ban insofar as it withholds legal protection from the corresponding intentions of the parties.[57] Despite being a classic dogmatic reference point, which has thus far received little attention in the discourse on digitalization, the interpretation of good morals within a legal system nonetheless enables the identification of normative experiences regarding the relationship between trust and ban.

---

56 E.g. Hein Koetz, „Sitten- und Gesetzeswidrigkeit von Verträgen" in Jürgen Basedow, Klaus J. Hopt, Reinhard Zimmermann (eds), *Handwörterbuch des Europäischen Privatrechts* (Tübingen: Mohr Siebeck 2009) 1404-1407; in order to limit legal transaction risks arising from trust in the declarations of the contracting parties, the term "liability based on trust" ("Vertrauenshaftung") is sometimes used in German Privat Law, partly in accordance with Roman law, see Claus-Wilhelm Canaris, *Die Vertrauenshaftung im deutschen Privatrecht*, (Munich: CH Beck 1971, reprint 1981); Claus-Wilhelm Canaris, *Gesammelte Schriften*, edited by Hans Christoph Grigoleit and Jörg Neuner (Berlin, Boston: De Gruyter, 2012) 3-656. For a general strenthening of such a de-individualized trust see also Claus-Wilhelm Canaris, "Wandlungen des Schuldvertragsrechts. Tendenzen zu seiner Materialisierung", (2000) Archiv für civilistische Praxis 273-364, 276.

57 On Nullity as a sanction in the sense of its behavior-controlling effect Herbert L A Hart, *The Concept of Law* (3rd ed, Oxford: Clarendon Press 2012) 33–35. See also Bernhard Jakl, *Handlungshoheit. Die normative Struktur der bestehenden Dogmatik und ihrer Materialisierung im deutschen und europäischen Schuldvertragsrecht* (Tübingen: Mohr Siebeck 2019) 129.

When exploring the potential to elucidate the essence of good morals inherent in the law, which can be conveyed through principles within the framework of contract law and constitutional requirements, a key jurisprudential insight into the relationship between trust and ban emerges: The argumentative and dogmatic path of private law begins with trust even under extreme scenarios, ultimately culminating in ban on certain contractual provisions in strictly limited cases.

This normative experience of the good morals provision can serve as a model for creating trust in innovations through the mechanisms of private law.

III. Trust as starting point for private law

In the legal-philosophical and institutional-argumentative assessment of the relationship between trust and ban in private law, the perspective initially shifts from the relationships between the state and its citizens to those among citizens themselves. Secondly, trust emerges here as an exemplar of interpersonal or intersubjective relationships, which remains also the prevailing paradigm in social and philosophical theories of trust.[58] Consequently, some scholars posit that the underlying reason for the binding force of contracts lies in the moral intuition that promises of performance inherently possess a uniquely compelling quality.[59] Others go so far as to invoke the notion that contractual obligations as a manifestation of human autonomy unfold within a framework of trust and respect akin to Kantian principles.[60]

Private law, particularly contract law, relies not foremost on state sanctions but on contractual agreements. Their binding nature and enforceability stem from mutual trust in individual freedom of choice and the fulfilment of performance promises by the parties involved. This entails the

58 See Carolyn McLeod, "Trust" in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), <https://plato.stanford.edu/archives/fall2023/entries/trust/> accessed 26 Apr. 2024.

59 E.g. Hanoch Dagan and Michael Heller, *The Choice Theory of Contracts* (Cambridge: Cambridge University Press 2017) 25-32.

60 Charles Fried, *Contract as Promise. A Theory of Contractual Obligation* (Cambridge, MA: Harvard University Press 1981) 5, 13-14, 17, 21. To this point, a critical interpretation of the legal philosophy of classical German philosophy from an action-oriented perspective cf. Jakl (n 57) 37 and 120-126.

risk for one contracting party of not only relying on the other contracting party but also of incurring losses if the latter fails to perform as expected.

Beginning with the mutual trust of contracting parties, bans in private law have only an indirect impact on governing social behavior, unlike criminal law and public law.[61]

The potency of mutual trust in contract law is exemplified, not least, by the success of the digital mobility service provider Uber and its algorithmic-based business model. Despite regulatory protections safeguarding the taxi industry throughout Europe through public law and administrative law including threats of fines, consumers were willing to use the service en masse. They willingly shared their location and payment data even in contravention of extensive data protection regulations in favour of what they perceived as a more user-friendly transportation alternative compared to taxis under public supervision. As a consequence, state regulations governing taxis across Europe were subsequently adjusted in favour of Uber.[62]

To explore the relationship between trust and ban, it is crucial to consider the potential justifications for limitations, restrictions, and even bans that exceptionally permit interventions into the freedom of trust-based contracts. However, the general rule is that contracts are binding. It is even acknowledged that mutual contractual obligations can override value judgements under the law of unjust enrichment (enrichment without cause) and tort law in civil law systems as well as in common law systems.[63] Further-

61 For examples of the broader European Terminology of Horizontal and indirect effects see Christian Timmermans, "Horizontal Direct/Indirect Effect or Direct/Indirect Horizontal Effect: What's in a Name?" (2016) 24 Issue 3/4 *European Review of Private Law* 673.

62 On the changes of the German Passenger Transportation Act (Personenbeförderungsgesetz) as an adjustment to reality for the needs of mobility services like Uber see Benjamin von Bodungen and Martin Hoffmann, "Digitale Vermittlung, Pooling, autonomes Fahren. Rechtsrahmen plattformbasierter Mobilitätsangebote vor dem Hintergrund der PBefG-Novelle" (2021) *Recht Digital* 93, 100.

63 E.g. in German Law for the overriding priority of the contract and its interpretation over the law on general terms and conditions, statutory prohibitions and enrichment law in the case of swap contracts the decision of the Federal Court of Justice (Bundesgerichtshof - BGH) (2023) *Neue Juristische Wochenschrift – Rechtsprechungs-Report* 1021 para. 22, 23. See for Britain making clear, that a claim in unjust enrichment could not succeed because unjust enrichment is excluded where the benefit conferred is dealt with by a contract, Supreme Court's Decision Barton and others vs. Morris and another in place of Gwyn Jones (deceased), 2023, UKSC 3 (Barton vs. Morris), <https://www.supremecourt.uk/cases/docs/uksc-2020-0002-judgment.pdf> accessed 26 Apr. 2024.

more, consent to a violation of a legal interest, even in cases involving bodily harm,[64] is conceivable, as is the preservation of the legal foundation in unjust enrichment law. For instance, in scenarios such as family guarantees, where a party's legitimate interest is subjectively acknowledged despite the contract being objectively disadvantageous.[65]

Drawing from normative experiences within private law lets us conclude that trust ought to be based, to some extent, in the individual freedom of choice of the contracting parties and their reciprocal trust in the fulfilment of mutual contractual obligations. Bans should be considered only as a well-grounded and insofar filtered exception that may follow.

IV. A comprehensive ban on social scoring?

The unique alteration in the dynamic between social trust and ban in private law can also be exemplified through the concept of social scoring. Social scoring pertains to mechanisms utilizing algorithmic data processing in application software, aiming to evaluate and incentivize positive conduct by individuals to govern or influence their behavior. Social scoring augmented by AI systems denotes the assessing of people's social behavior for the purpose of predicting or managing behavior.[66] Illustrations include associating infrequent sick leave with higher salaries in labour law or other incentives, as well as linking regular subscription upgrades to additional benefits or access to other advantages within bonus systems, which many workers and consumers often appreciate.[67]

64 E.g. for Germany: consent according to sec. 630 (d) BGB in the context of medical treatment involving bodily injury excludes other claims based on tort or unjust enrichment.

65 E.g. for Germany: even if a contract is unusually burdensome for the weaker party, the contract is binding, if the weaker party has a self-interest or the stronger party has an accepted interest in a specific advantage, e.g. to prevent shifts in assets to the disadvantage of the stronger party, see the decision of the Federal Court of Justice (Bundesgerichtshof - BGH) (2013) *Neue Juristische Wochenschrift – Rechtsprechungs-Report* 1258 para. 21.

66 See e.g. Martin Wiener, W. Alec Cram and Alexander Benlian, "Algorithmic Control and Gig Workers: A Legitimacy Perspective of Uber Drivers" (2023) 32 (3) *European Journal of Information Systems* 485.

67 See e.g. Emma McDaid, Paul Andon and Clinton Free, "Algorithmic management and the politics of demand" (2023) 103 *Accounting, Organizations and Society*, <https://www.sciencedirect.com/science/article/pii/S0361368223000363> accessed 26 Apr. 2024.

According to the Draft AI Act, social scoring is banned from the EU market due to its potential to interfere with trust in the use of AI applications. Specifically, AI applications for behavior or emotion recognition in workplaces or schools are to be disallowed due to their deemed unacceptable risk.[68] Additionally, political or religious profiling ought to be banned. AI applications in general are not allowed to directly influence or exploit people's behavior.[69]

With regard to trust building, existing and forthcoming regulations within public law at the European level, including the Draft AI Act, are not very convincing. There remains a concern that social trust in the legal system could be significantly undermined if it were revealed that the proposed ban on social scoring under European law, and thus under public law, could firstly be rendered ineffective by contractual agreements, as seen in the Uber case with taxi regulations. Secondly, the current proposal lacks an argumentative approach to this social issue rooted in individual freedom of choice, making it challenging to justify such a broad ban to the individual contracting parties in a comprehensible or plausible manner under civil law. Consequently, citizens may even lose trust in AI systems, because they are regulated under the Draft AI Act at this point. This concern is further compounded by the absence of a distinction between the risks posed by state-run and private social scoring systems and their respective potential benefits.[70]

Given the normative experience in private law, particularly in contract law, where trust serves as the foundation and bans are infrequent exceptions requiring solid justification, the transition from ban to trust adopted by the executive and legislative branches for the digital realm seems at least bold, if not improbable. For instance, it seems unlikely that all bonus systems, initially regarded as instances of social scoring, will be eliminated upon the latter's prohibition. However, this raises the spectre of the Draft AI Act inadvertently silencing an essential discourse on the rejection of

68 Cf. Reason 31 and Art. 5 (1f) and (1g) of the Draft AI Act (n 1).

69 Cf. Reason 29 and Art. 5 (1c) of the Draft AI Act (n 1) esp. for the (normatively not completely convincing) description of a risk of deceiving natural persons by nudging through AI Systems.

70 E.g. critical on state-run social scoring systems and their ability to improve social situations so far Anja Geller, *Social Scoring durch Staaten. Legitimität nach europäischem Recht – Mit Verweisen auf China* (Munich: Ludwigs-Maximilians University Munich 2022) 99, <https://edoc.ub.uni-muenchen.de/31151/1/Geller_Anja.pdf> accessed 26 Apr. 2024.

welfare augmentation through social scoring for valid reasons, thus stifling public debate in Europe.

Irrespective of the future of social scoring beyond the Draft AI Acts, the predominantly state-centric European approach currently adopted in the policy field of digitalization with its intended path from standard bans to creating social trust, stands in a remarkable contrast to the contentious yet tested and established normative experiences in private law. These normative experiences typically involve a path from trust to ban, a path that appears compelling, if not plausible, particularly in the context of regulating upcoming algorithms and AI systems.

D. Conclusion

We have (re-)evaluated the debates surrounding the regulation of innovations in democracies, drawing on legal philosophy and considering the various argumentative standards across different areas of law. This approach has allowed us to discern the rationales behind issuing certain bans, not only by analysing public debates but also by interpreting and reconstructing the law. By expanding the prevailing communicative picture of the relationship between trust and ban, we have introduced an institutional-argumentative picture.

Moving beyond the communicative picture, we elucidated that the relationship between trust and ban exhibits a distinct directional structure in the realms of criminal law and civil law. In criminal law, bans with sanctions are intended to foster trust, whereas in private law, trust in individual decisions serves as the starting point, with bans utilized in exceptional cases to secure trust.

Regarding the criminal law path from ban to trust, the communicative perspective demonstrates how trust can be cultivated between interacting citizens, with sanctions potentially stabilizing behavioural expectations in the context of innovations. However, the emphasis on generating trust through banning untrustworthy behavior sidelines the equally crucial principle of limiting criminal law to exceptional circumstances – leaving room for potentially unlimited use of criminal law. Innovation debates often prioritize the prevention of any risks associated with innovations without considering the specific rights and concerns intended to be protected by bans or the freedoms these restrict. For example, the UN Comprehensive

Study on Cybercrime[71] focuses primarily on the risks of cyber activities without addressing the different legal interests being protected, e.g. the prohibition of cyberfraud serves the protection of asset rights whereas the prohibition of hate speech serves the protection of personal rights.

In contrast, the private law trajectory from trust to ban reveals a gap in the communicative understanding from a legal-philosophical and insofar institutional-argumentative perspective. The mutual trust between contract parties is underestimated and the potential path from trust to ban is neglected. This tendency to neglect is particularly concerning as regulations in the digital sphere strive to maintain effectiveness by offering justifications for bans that influence the everyday behavior and use of digital opportunities by contract parties. Consequently, there is a risk that crucial public debates will be overshadowed by bans, including for example discussions on which welfare gains from state or privately organized social scoring we may want to give up for good reasons.

Contrary to the communicative picture in the legal-philosophical and institutional-argumentative picture trust no longer appears as an independent normative concept with unique analytical or explanatory power. Instead, trust derives its significance in relation to bans across various legal domains, such as criminal law or private law in different ways. This differentiation enables a nuanced and thus well-founded critique of debates on trust-building bans to regulate innovations in democracies. It is this approach that opens our eyes to the issues that we should be discussing.

71 UNODC (n 26) passim.