Trust, Heuristic Decision-Making, and the Design of Behaviorally-Aware Online Privacy Regulation

Christopher Jon Sprigman[1] & Stephan Tontrup[2]

DRAFT: Please do not further circulate

I. Introduction

Between 2013 and 2015, about 270,000 Facebook users allowed an app called "thisisyourdigitallife" access to personal data from their Facebook profiles, including their Facebook comments and "likes." Developed by a University of Cambridge professor, the app was advertised to users as a personality test based on an analysis of the user's activities on Facebook. Individuals deciding to permit the app to access their data were also (whether they understood it or not) permitting the app to collect data about their Facebook friends. This data could be used in ways that put both their personal privacy and their friends' privacy at risk—analysis of Facebook "likes," for example, can provide information about an individual's personality traits, sexual orientation, political views, mental health status, and substance abuse history. And then there was a different risk that no user could possibly have anticipated. It turns out that a company called Cambridge Analytica later used data the app collected to build voter profiles that it gave to the Trump campaign, which used them to target political advertisements in the 2016 U.S. presidential election.

Why do so many people who claim to care about their privacy hand over so much information to online platforms like Facebook? In this Article we will argue that the answer has a lot to do with the way that the current U.S. privacy law framework of "notice and choice" frames privacy decisions.

---

[1] Murray and Kathleen Bring Professor of Law, New York University School of Law; Co-Director, Engelberg Center on Innovation Law and Policy. The authors wish to thank Florencia Marotta-Wurgler, Jeanne Fromer, Jason Schultz, Barton Beebe, Christoph Engel, Thomas Streinz, Benedict Kingsbury, Kathy Strandburg, Michael Weinberg, Scott Hemphill, Rochelle Dreyfuss, Jacob Victor, Stefan Bechtold, Alex Stremitzer, Eyal Zamir, Ben Depoorter, Andrew Kukilow and his team and participants at the 2022 Conference on Empirical Studies and at the 2022 annual meeting of the American Law and Economics Association , in a conference held by Guarani Global Law and Tech at the New York University School of Law, in the "Mapping the Online World" conference sponsored by ETH-Zürich and NYU and held at the New York University School of Law, and in workshops at NYU's Engelberg Center and at ETH-Zürich and the University of Zürich. [MORE]. We also thank Stephen Gray, Selma Guney, Sally Kang, and Max Krinsky for excellent research assistance. The authors also thank the Engelberg Center and the Filomen D'Agostino and Max E. Greenberg Research Fund for providing funding that supported this project. [OTHER ACKNOWLEDGMENTS]

[2] […]

The notice and choice framework, around which U.S. privacy law is based, requires that firms collecting and using individuals' private information disclose what they are collecting and what they will do with it, with the expectation that potential disclosers, having read and understood the notice, will choose whether to consent to the disclosure requested.[3]

One foundational problem with the notice and choice approach is that most people do not read privacy policies before deciding whether to disclose personal information.[4] People without notice of the information a requester is seeking, or what she plans to do with that information, cannot possibly be making the informed privacy decisions that the notice and choice framework envisions. But not reading notices alone does not explain the sort of privacy behavior that has hundreds of thousands of Facebook users handing over sensitive personal information to unknown online actors like "thisisyourdigitallife." People seem readily to reveal private information even where privacy risks should be apparent and users might be expected to take them into account. This Article shows that the notice and choice privacy framework is part of the problem. By framing privacy choices as if they were occurring within a bilateral relationship between the user and the platform or service requesting disclosure, the notice and choice framework triggers a form of heuristic decision-making that often is successfully employed in real-world bilateral relationships but which may often be maladaptive in digital worlds, both because the signals that the heuristic relies on to suggest trustworthiness are so easily exploited, and, in particular because the heuristic blinds users from considering the broader social risks of their personal privacy choices.

---

[3] *See* FED. TRADE COMM'N, PRIVACY ONLINE: A REPORT TO CONGRESS 7 (1998), https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf ("Consumers should be given notice of an entity's information practices before any personal information is collected from them. Without notice, a consumer cannot make an informed decision as to whether and to what extent to disclose personal information."); *see also* FED. TRADE COMM'N, INTERNET OF THINGS: PRIVACY & SECURITY IN A CONNECTED WORLD, at v (2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf ("The Commission staff believes that consumer choice continues to play an important role.").

[4] See Pew Research Center, *Americans And Privacy: Concerned, Confused and Feeling a Lack of Control Over Their Personal Information*, Nov. 15, 2019 (finding that only 22% of Americans report reading privacy policies "always" (9%) or "often" (13%), while 74% read them "sometimes" (38%) or "never" (36%)), https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/. And of the respondents who claim to read at all, 43% admit that they glance the policy over without reading it closely. Id. This data is self-reported and may be optimistic. Research on the actual internet browsing behavior of consumers shows that only 1 or 2 out of 1000 retail software shoppers access the online end user license agreement (EULA) that governs the terms on which they are permitted to access and use the software, and, of those who access the EULA, most read no more than a small portion. *See* Yannie Bakos, Florencia Marotta-Wurgler & David R. Trossen, *Does Anyone Read the Fine Print?: Consumer Attention to Standard Form Contracts*, 43 J. LEGAL STUD. 1 (2014). There is little reason to believe that consumer attention to online privacy policies is significantly greater than for EULAs.

This Article provides novel empirical evidence demonstrating that people often rely on a basic *trust heuristic* to make decisions about their privacy.[5] The individual considering whether to disclose uses the heuristic to form a belief about the trustworthiness of the party seeking disclosure—i.e., the "counterparty." The trust heuristic prompts the potential discloser to compare the counterparty's actual behavior with the possible alternative actions perceived to be within the counterparty's "action space"—i.e., the set of actions that the potential discloser perceives as available to the counterparty. If the chosen behavior appears to be beneficent relative to other options in the counterparty's action space, then the potential discloser is likely to view the counterparty's intentions as positive or kind, and the counterparty as trustworthy. That conclusion should, in turn, raise the potential discloser's willingness to divulge her private information. An example would be if the counterparty offers a privacy protection—say, by making a statement that the counterparty will protect all data disclosed to it by storing it only in encrypted form—that the potential discloser believes is not required but rather is an exercise of the counterparty's discretion. In such instance, the perception of beneficence should lead the potential discloser to be more likely to trust the counterparty, all else equal, and to be correspondingly more likely to disclose. On the other hand, if the counterparty's chosen behavior appears to be less beneficent than other behaviors that the potential discloser perceives to be within the counterparty's action space, then the counterparty is more likely to be perceived as untrustworthy, all else equal, and the potential discloser correspondingly less likely to disclose private information to that counterparty. An example would be if the counterparty is perceived by the potential discloser as being free to offer a privacy protection, like encrypting collected data, and yet fails to offer that protection.

Heuristic decision making is frugal. It focuses on a few cues, sometimes only on a single cue, and ignores other information that might be relevant. Heuristics can nevertheless be successful mental strategies in low-information environments when decision-makers do not have time or resources to evaluate all information available, when decision outcomes are uncertain, and the cue employed is substantially correlated to the predicted outcome.[6] However, applying a heuristic can also be dangerous, in particular when the cues the heuristic relies upon are not informative in the environment where the heuristic is used. We show in this Article that focusing on the trustworthiness of the counterparty in the digital privacy context will often be

[5] RALPH HERTWIG & ULRICH HOFFRAGE, SIMPLE HEURISTICS IN A SOCIAL WORLD (2013).

[6] The so-called "recognition heuristic," for example, exploits whether the decision maker recognizes an object or not. *See generally* Thorsten Pacher et al., *The Recognition Heuristic: A Review of Theory and Tests*, 2011(2) Frontiers in Psychology 147. Consider a situation in which the decision maker wants to buy a spare part for a car she knows nothing about. Two similarly-priced parts are offered; she has heard of one brand of part, but not the other. The recognition heuristic is likely to lead her to choose the spare part bearing the brand she has heard of. Other information that might be helpful to the choice, like the technical description of the products, is ignored. If there is a relationship between the recognition and the quality of the product, then the heuristic is ecologically valid; for example, firms with better products stay longer on the market and therefore are more likely to be known.

maladaptive. Online, people interact with algorithms and choice options that may be designed to reveal little or even mislead about the true motivations of the party requesting disclosure of private information. In such settings, the heuristic is based on the slimmest of evidence and subject, in many instances, to misinterpretation or manipulation. If applied in such a context the heuristic can cause people to ignore otherwise valuable information about the risk of disclosure—i.e., information about how personal data is created, altered, and could be shared or sold by the requesting party.

Moreover, we will show that the heuristic is leading people not only toward poor decisions about their own privacy, but also to ignore information about the social component of privacy—that is, about how their personal privacy decisions affect the privacy interests of others. Conventionally a privacy decision is framed as the choice of an individual: for example, the notice and choice framework makes it appear to potential disclosers as if they can each, as individuals, conclusively decide what to reveal and to whom. But that is an incomplete picture. In a social network such as Facebook, an individual's decision whether to disclose information will often affect others' privacy interests—for example, the privacy of other people whose image is captured in a photograph the user posts on his timeline. And the same is true in the other direction—any individual Facebook user's privacy is a product, in part, of other Facebook users' decisions.

We designed a series of three experiments to better understand how people make privacy decisions in online settings.[7] We will describe those experiments very briefly here, and then again in detail in Part III of this Article.

In Experiment 1, we investigated whether subjects do in fact employ the trust heuristic we have detailed above to make these decisions. We designed our experiment as a sort of tag-along to an unrelated study that presented subjects with an effort task.[8] When subjects had finished the task and were about to be paid for their participation, they assumed the experiment was over and their behavior was no longer being observed by researchers. In fact, however, we randomly assigned participants into different treatments in which we offered them varying payment options—some requiring substantial disclosure of confidential financial information, others offering

[7] All experiments were conducted in our online laboratory with participants of the University of Münster in Germany. Germany was a good jurisdiction in which to conduct our experiment because, unlike in the United States, Germans are accustomed to making payments via bank transfer—a payment method that, as shall be described below in Section III.A.2 is an important feature of our experiment. Payment by bank transfer requires substantial disclosure of confidential information. As we shall describe, in our experiment we also offered a more privacy-protective payment method. Thus, by conducting our experiment in Germany, we were able to offer subjects a choice between different payment methods that presented very different privacy implications, and different trade-offs between privacy and convenience.

[8] See Stephan Tontrup & Christopher Jon Sprigman, *Self-Nudging Contracts and the Positive Effects of Autonomy—Analyzing the Prospect of Behavioral Self-Management*, 19 J. OF EMPIRICAL LEGAL STUDIES 594 (2022).

an anonymous cash payment option designed to preserve their privacy. We assumed that subjects would use the trust heuristic to determine whether they should disclose identity and financial information or remain anonymous and pick up their money in cash. Subjects employing the trust heuristic would compare the payment methods offered with what they perceived to be our action space (the choice set they perceived to be available to us). If we offered a payment method that preserved their privacy (like an anonymous cash pickup), while we might also have offered only a bank transfer, the heuristic would suggest that the subject should trust the experimenter, as we made a beneficent choice, and be more likely to disclose confidential information—that is, to forego the anonymous payment method and instead provide us with their bank details. If, by contrast, subjects were offered the same payment options (anonymous cash payment and bank transfer) but were told that we were *obliged* by university rules to offer them the anonymous payment choice, then subjects would be prevented from using the trust heuristic—in this scenario, we do not have choice within our action space except to provide the privacy-protective option—and these subjects should not perceive the protective option to be a product of our beneficence. They should therefore be less likely, all else equal, to trust us, and also less likely to provide their bank data.

Our results support these hypotheses. They show that when subjects perceived the provision of privacy protection in the form of an anonymous payment method to be a product of the counterparty's *choice*, subjects could apply the trust heuristic and accordingly were more likely to opt for the more convenient bank transfer and to reveal the requested sensitive information. However, when we made it impossible for subjects to use the trust heuristic by informing them that the identity protection offered to them was *mandated by law*, and not the counterparty's beneficent choice, they were significantly less likely to reveal personal information. The result shows that the participants employed the heuristic and that it leads to substantially more trust and privacy disclosure, compared with subjects who we prevented from employing the heuristic, thereby eliminating the subject's ability to frame the protection provided as the product of the requester's beneficent choice.

In Experiment 2 we investigated whether subjects employing the trust heuristic would be more likely to ignore valuable information that could improve the quality of their privacy decision. For this test, we recruited a new sample of subjects. During the registration process the subjects were asked to provide contact information that would allow us to invite them to participate in future studies, and we prompted them to reveal financial information to facilitate payment for those future studies. We designed two treatments. In one we provided subjects with credible background information, such as the researchers' university affiliations, past work history, record of academic publications, and other indicia of the researchers' trustworthiness that would allow them to make a more informed assessment of the privacy risk they would assume should they provide us their financial details. In the second we aimed to send a negative

trust signal by giving subjects no information except the Gmail (i.e., non-university) address of a research assistant. We tested whether subjects employing the trust heuristic would consider these positive and negative informational signals in their privacy choice.

Our results support our hypothesis that the trust heuristic tends to crowd out otherwise-credible information about privacy risk. When subjects employ the trust heuristic, they appeared to trust their counterparty and it did not matter whether subjects received a positive or a negative informational signal about their counterparty. In particular, the negative signal—i.e. no information was available about the researchers collecting the confidential data—did not reduce subjects' willingness to reveal that information. Only when we made it impossible for subjects to use the trust heuristic by informing them that their counterparty was *obliged* by law to offer them protection did subjects consider the informational signal they were given and the results changed substantially.[9] Subjects given no credible information about their counterparty's trustworthiness were significantly less likely to disclose their financial data. Conversely, subjects who received a positive signal—i.e. they learned about the researchers' affiliations, published research and projects—became more likely to disclose their private banking information, as should rationally be expected.

We conclude that participants who employ the trust heuristic base their decision primarily on the heuristic, and tend to ignore both negative as well as positive informational signals that would allow them to better estimate their true privacy risks. Moreover, our results suggest that blocking the trust heuristic benefits the subjects and allows them to consider more credible signals in their decision-making. The experiment thus demonstrates that the trust heuristic can be maladaptive in digital environments: the potential discloser's reliance upon the perception of the counterparty's beneficent intention can easily be manipulated, and at the same time information that may be more helpful in making the trust decision is ignored.

In our Experiment 3 we investigated whether the trust heuristic leads subjects to disregard the public good nature of their privacy choices. We show that the heuristic lulls subjects into ignoring the ways in which their privacy decisions can affect the privacy of others, and, conversely, how the privacy choices of third parties (people who they may not even know) might affect their privacy (for example by disclosing data that refers to them). We again recruited new subjects for participation in future studies and we asked these new subjects to reveal their Facebook data to us. We employed a web application (specifically, the Osint "Social Links" app, which we detail

[9] The heuristic can be applied only if the potential discloser perceives (correctly or not) that the counterparty has a choice between different options. The heuristic processes whether the decision-maker chooses the more benevolent of her options or the less benevolent. However, if the counterparty can only implement one option because she is obliged by law to do so, the heuristic cannot operate because it cannot assess the counterparty's benevolence against a perceived choice set.

below[10]) that allowed us to elicit all publicly accessible information associated with subjects' Facebook accounts, including postings that they made anywhere on Facebook and also postings others had made referring to them and their content, such as pictures where they tagged the subjects. We gave subjects a choice of whether to permit us access to this data and incentivized their decision, paying them € 4.00 if they revealed the information.

Collecting this Facebook information allows us to create a privacy risk that extends not only to the discloser, but also to the discloser's Facebook friends. We implement two treatments. In the first we offer subjects privacy protection, informing them that they can register for our studies while remaining anonymous and without providing any Facebook information. In the second treatment, we offered subjects the same privacy protection of remaining anonymous but blocked them from using the trust heuristic (as in Experiments 1 and 2). We find that in the first treatment more than half of the subjects allow us to harvest all public information available about them on Facebook. When we prevented subjects from using the heuristic, subjects were significantly less likely to reveal their Facebook data to us.

Of course, the intensity of subjects' use of Facebook varies substantially, and, as a consequence, some users who give us access reveal more personal information to us than others do. We exploit this variance and analyze the magnitude of the Facebook data subjects reveal to us. We find that when the trust heuristic is blocked, fewer subjects reveal their data, and also—crucially—the subjects who reveal their data have less Facebook data to reveal, on average, compared with those who decide not to reveal. By comparison, in the treatment where subjects can employ the trust heuristic, we see not only more subjects disclosing their data but also no difference in disclosure rate between subjects with a lot of Facebook data to disclose and those who have less. This result suggests that subjects who are blocked from applying the heuristic make decisions that more accurately reflect both their personal risk and also the *social* risk of disclosure. Thus Experiment 3 demonstrates that blocking subjects' heuristic decision-making can make their choices more sensitive to the privacy risks they impose on others, and that others impose on them, in digital ecosystems and social networks.

Taken as a whole, our results demonstrate a foundational weakness in the notice and choice privacy framework that goes beyond mere failure to read privacy disclosures. It appears that the framework, by conceptualizing privacy as a bilateral trust relationship, may facilitate a type of heuristic decision-making that is both ill-suited to this type of decision environment and exploitable by bad actors. Providing a privacy-protecting option signals to heuristic decision-makers that the platform or service she is using is trustworthy and may lead to individuals being less likely to avail themselves of protections and more likely to disclose their personal data than they

[10] See for more detail Section III.C.2.

would be if no means to preserve their privacy were provided. We see precisely this effect arising in our experiments.

Given what we see across our experimental findings, how can privacy law be reformed such that the legal regime nudges people to make more informed privacy decisions? The answer is the law must be designed in the light of how individuals actually make their privacy choices, or regulation may fail or even backfire when notice and choice with its bilateral framing induces users into applying the trust heuristic and ignoring both credible information and the social privacy risks caused by their individual choices. In our view, that behaviorally-aware privacy law re-design has two principal elements:

First, the trust heuristic's dysfunction suggests, at a minimum, that parties seeking confidential information should be obliged to disclose *that the law, and not the beneficent intent of the party seeking disclosure, is the source of the protection that is being offered.* If this sort of disclosure succeeds in blocking the trust heuristic, our findings suggest that people will be more receptive to credible institutional and technological cues in assessing their privacy risk. Note, however, that re-designing privacy law to block use of the trust heuristic will run into the same foundational problem that undermines the entire notice and choice framework: practically nobody reads the endless number of lengthy and legalistic notices people are presented with for each service they wish to use, and it seems correspondingly unlikely that people are going to read disclosures about the source of particular privacy protections.

Our second proposal is a more fundamental revision of privacy law, which we believe is required before we can have any realistic prospect of overcoming the influence of the trust heuristic. We describe a framework for more predictably and pervasively blocking use of the trust heuristic, one which works at the root by disrupting the particular framing of privacy decisions—i.e., as bilateral transactions—on which the heuristic depends. The re-framing we describe would preserve privacy law's commitment to individual autonomy and choice. But it would free individual privacy decisions from the heuristic by mandating that privacy preferences be expressed *in advance*, and using *a standardized form and set of terms relating to both collection and use.* In our re-setting of privacy law, requesters seeking disclosure of personal information would query a database containing a person's choices—choices made in advance of disclosure requests, according to a standardized menu of options for both types of information that might be disclosed and the uses to which that information may be put. Importantly, these ex ante choices would be made without the involvement of any particular counterparty. Once an individual's choices are recorded, information would be transferred without the need for further permissions if the request and the discloser's preferences align. Any collection or use of data that exceeds that consent would require additional permissions. In this way, the law could push back against strategies that firms use to activate the heuristic, and also create a new

equilibrium in privacy decisions such that firms that ask for more disclosure must do so against a default that the potential discloser has already established and which is likely to be sticky.

In Part II we describe our study's theoretical framework. In Part III we report experimental method and results. Part IV discusses the implications for privacy regulation, and for future privacy scholarship.

II. The Cognitive Background of Privacy Decision-making

A. Rational Privacy Preferences and Cognitive Errors

Much of the academic research on privacy has, like U.S. privacy law itself, been based on a rational choice framework. That framework assumes (1) that people act on their stable preferences to make value-maximizing trade-offs between privacy and other concerns,[11] (2) that decisions whether to disclose private information are made by balancing "the usefulness of privacy with the utility of openness,[12] and (3) that disclosure is made only when the individual considering disclosure expects a net benefit.[13] According to this perspective, individuals can be relied on to rationally decide between disclosing or secreting their personal information.

But people's actual privacy choices appear to contradict this rational choice account. In a poll conducted in 2013, Gallup recorded that 83% of American internet users surveyed reported either being "very concerned" (48%) or "somewhat concerned" (35%) about the protection of their personal information.[14] And yet, a growing body of empirical evidence shows that individuals who claim to care about their privacy will nonetheless voluntarily turn over volumes of personally identifiable information in a range of settings, including when using social media websites.[15] Indeed, some evidence suggests that when consumers are asked to pay for privacy protections, they will refuse. In one seminal field experiment, participants were given the choice to buy a DVD from one of two online stores.[16] One store required the disclosure of substantially more sensitive data than the other; otherwise the stores were

---

[11] VALERIAN J. DERLEGA ET AL., SELF-DISCLOSURE (1993); Richard A. Posner, *The Economics of Privacy*, 71 AM. ECON. REV. 405, 405–09 (1981); L. B. ROSENFELD, *Overview of the Ways Privacy, Secrecy and Disclosure are Balanced in Today's Society*, in *Balancing the Secrets of Private Disclosures* (2000); George J. Stigler, *An Introduction to Privacy in Economics and Politics*, 9 J. L. STUD. 623, 623–44 (1980).

[12] Sandra Petronio, *The Boundaries of Privacy: Praxis of Everyday Life, in Balancing the Secrets of Private Disclosures*, in Balance the Secrets of Private Disclosures 32, 37 (2000).

[13] Tiffany B. White, Consumer Disclosure and Disclosure Avoidance: A Motivational Framework, 14 J. Consumer Psychol. 41, 48 (2004).

[14] Computers and the Internet, Gallup (The data originates beyond 2013, but the exact year isn't clear), http://news.gallup.com/poll/1591/Computers-Internet.aspx

[15] JEFF H. SMITH, MANAGING PRIVACY: INFORMATION TECHNOLOGY AND CORPORATE AMERICA (1994).

[16] Alastair R. Beresford, Dorothea & Kübler & Sören Preibusch, *Unwillingness to Pay for Privacy: A Field Experiment*, (2010), https://www.ssoar.info/ssoar/bitstream/handle/document/23832/ssoar-2010-beresford_et_al-unwillingness_to_pay_for_privacy.pdf?sequence=1.

identical.[17] In a first treatment, DVDs were offered at a price one Euro cheaper in the store requesting more personal information; almost all buyers chose the cheaper store.[18] In a second treatment, prices were identical and participants bought from both shops equally often, suggesting that subjects placed no value on the increased privacy that one shop was offering.[19]

One interpretation of this inconsistency suggests that people may have unstable or simply weak preferences about their privacy. In particular, studies have shown that individuals' preferences often appear to be unstable; that is, affected by the framing of alternatives[20] and by the elicitation method employed.[21] Preferences tend to be most unstable when the costs of a choice are abstract and uncertain,[22] which may often be the case for privacy decisions, especially when the risks are difficult to assess.[23] In addition, the literature identifies a number of particular psychological distortions, including hyperbolic discounting (i.e., the tendency to choose smaller, immediate rewards, rather than larger ones that are realized later), problems with self-control, the tendency toward optimism bias (i.e., the mistaken belief that one's personal risks are lower than the risk faced by others who are similarly situated), the difficulty of estimating the weight of cumulative risks,[24] and a "control paradox" whereby people who experience more perceived control over some particular aspect of their privacy sometimes respond by mistakenly generalizing that perception of control and revealing more information than they would otherwise.[25] Together, these biases undermine rational choice as an explanatory framework for privacy decisions. They also suggest that privacy regulation based around notice and choice is likely to fail. In particular, if weak or unstable preferences are the major cause of the gap between reported preferences and behavior, then any effective privacy regulation might have to take privacy protection out of the users' hands. That is, privacy regulation would have to be more paternalistic, imposing mandatory privacy protections across a range of

---

[17] *Id.* at 3.

[18] *Id.*

[19] *Id.*

[20] Ravi Dhar & Itamar Simonson, *The Effect of the Focus of Comparison on Consumer Preferences*, 29 J. MARKETING RES. 430, 430–440 (1992); Tversky & Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1124–31 (1974).

[21] Amos Tversky, Paul Slovic & Daniel Kahneman, *The Causes of Preference Reversal*, 80 Am. Econ. Rev. 204, 204–217 (1990).

[22] Craig R. Fox & Amos Tversky, *Ambiguity Aversion and Comparative Ignorance,* 110 Q. J. Econ. 585, 585–603 (1995); Dale Griffin, Wendy Liu & Uzma Khan*, A New Look at Constructed Choice Processes Marketing Letters*, 16 Sixth Invitational Choice Symp. 321, 321–333 (2005); C. K. Hsee et al., *Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis*, 125 PSYCHOL. BULL. 576, 576–591. (1999); C. K. Hsee et al., *Lay Rationalism and Inconsistency Between Predicted Experience and Decision*, 16 J. BEHAV. DECISION MAKING 257, 257–272 (2003).

[23] Alessandro Acquisti, *Privacy in Electronic Commerce and the Economics of Immediate Gratification*, Proc. 5th ACM Conf. Electronic Comm. (2004).

[24] *Id.*

[25] Laura Brandimarte et. al., *Misplaced Confidences Privacy and the Control Paradox*, 4 SOC. PSYCHOL. & PERSONALITY SCI. 340–47 (2012).

settings.[26] This is in contrast to the current model, which is aimed at empowering people to act on their preferences.

But before we set off down the path of privacy paternalism, we want to test an alternative explanation that refers not to unstable preferences, but to the cognitive mechanisms of privacy decision-making. If cognitive mechanisms are part of the explanation for the variance between expressed privacy preferences and real-world privacy behavior, the picture may not be quite so dire. Cognitive errors can often be corrected by changing the decision-making environment. If cognitive errors underlie some peculiarities of privacy decision-making, and if those errors are subject to correction, then perhaps a less paternalistic regime of privacy regulation remains viable, albeit one built on different premises than the current regime of notice and choice.

B. Trust and the "Heuristic" Model of Privacy Decision-making

How do people make privacy choices? We can look for guidance to the substantial economic literature examining fairness, reciprocity, and trust in individual decision making. We will see that individuals' perceptions of others' intentions plays a crucial role in these different areas of decision-making. The economic literature suggests that individuals considering a transaction do not simply weigh expected payoffs, but instead make choices that take fairness considerations into account[27]—that is, perceptions of fairness or unfairness, and not just expected gains, appear to drive positive or negative reciprocity behavior.[28] And, importantly, when deciding whether an action was fair or unfair, individuals appear to base their judgment on the perceived intentions of their interaction partner.[29]

Many economic experiments investigating how people make decisions about what is fair or unfair employ a form of "ultimatum game." In this game, one player, the "proposer," is endowed with a sum of money. The proposer is empowered to propose a split of the endowment with a second player, the "responder." The proposer

---

[26] CITE

[27] Ernst Fehr & Armin Falk, *Wage Rigidity in a Competitive Incomplete Contract Market*, 107 J. POL. ECON. 106, 106–134 (1999); Truman F. Bewley, *Why Wages Don't Fall During a Recession*, 31 J. BEHAV. & EXPERIMENTAL ECON. 431, 431–432; Ernst Fehr, Simon Gächter & Georg Kirchsteiger, *Reciprocity as a Contract Enforcement Device: Experimental Evidence*, 65 ECONOMETRICA 833, 833–860 (1997).

[28] Joyce Berg, John Dickhaut & Kevin McCabe, *Trust, Reciprocity, and Social-History*. 10(1) GAMES & ECO BEHAV. 122, 122–42 (1995); Alvin Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara & Shmuel Zamir, *Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study*, 81 AM. ECON. REV. 1068, 1068–95 (1991); Lisa Cameron, *Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia*, 37 ECON. INQUIRY. 47, 47–59 (1999).

[29] Armin Falk et. al., *On the Nature of Fair Behavior*, 41(1) ECON. INQUIRY 20–26 (2003). Martin Dufwenberg & Georg Kirchsteiger, *A Theory of Sequential Reciprocity*, 47(2) GAMES & ECON. BEHAV. 268–98 (2004); Matthew Rabin, *Incorporating Fairness into Game Theory and Economics*, 83 AM. ECON. REV. 1281, 1281–1302 (1993); Gary Charness & Matthew Rabin, *Understanding Social Preferences with Simple Tests,* 117 Q. J. ECON. 817, 817–869 (2002); Werner Güth & Oliver Kirchkamp, *Will You Accept Without Knowing What? The Yes-No Game in the Newspaper and in the Lab*, EXPERIMENTAL ECON. (2012).

communicates his or her proposed split to the responder, and the responder may accept it or reject it. If the responder accepts, the money is split according to the proposal; if the responder rejects, both players receive nothing. (Both players know in advance the consequences of the responder accepting or rejecting the offer.) If both proposer and responder are purely rational, the proposer should make an offer in which she keeps virtually all of the endowment, and the responder should accept any split that gives him more than zero. But over many experiments, participants rarely behave in this way; rather, fairness preferences typically lead proposers to make more generous offers, and responders to reject (even at cost to themselves) offers that they perceive as insufficiently generous.[30]

Assumed intentions play an important role in these fairness decisions: in ultimatum games we see that a low offer that might otherwise be rejected as being unfair is likely to be accepted if the responder assumes that the proposer could only have made a less fair offer but did not. In contrast, an offer with an identical payoff is far less likely to be accepted where the proposer is perceived by the responder as having been free to propose a payoff that was more fair.[31] In other words, identical actions by the proposer may signal entirely different information about the proposer's intent depending on the responder's perception of what actions were available to the proposer. As a consequence, responses to offers presenting identical payoffs may vary, even though responders' preferences about the payoff itself are stable.

How strongly do perceived intentions affect decision making independently from expected payoff? Cushman et al. find that expected payoffs dominate,[32] whereas Charness and Levine report that perceived intention dominates.[33] In a series of experiments that combines what we view as the desirable features of the Cushman and Charness and Levine studies, Schaechtele et al. report strikingly that the effect of intention on outcomes was at least as large as the effect of perceived likely payoff.[34] In other words, seemingly good intentions are so important for decision making that they can, and often do, outweigh an objectively bad payoff.

We can relate this decision-making literature in the fairness domain to individuals' privacy behavior, and, in particular, to the specific decisions people make about whether to disclose private information in the context of a potential transaction. In doing so, we can explain the apparent gap between preference and behavior without referring to a potential instability of individuals' privacy preferences. We show

---

[30] Richard Thaler, *Anomalies: The Ultimatum Game*, 2 J. ECON. PERSPECTIVES 195-206.

[31] Armin Falk et. al., *On the Nature of Fair Behavior*, 41(1) ECON. INQUIRY 20–26 (2003).

[32] Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "tremblinghand" game. PLOS One, 8, 1–7.

[33] Gary Charness and David I. Levine, Intention and Stochastic Outcomes: An Experimental Study. The Economic Journal, Vol. 117, No. 522 (Jul., 2007), pp. 1051-1072.

[34] Simeon Schaechtele, Tobais Gerstenberg & David Lagnado, *Beyond Outcomes: The Influence of Intentions and Deception*, Proc. 33d Ann. Conf. Cognitive Sci. Society 1860–1865 (2011).

evidence that people facing privacy decisions employ a decision-making strategy, which we label the "trust heuristic," that allows them to ascribe an intent to the party requesting disclosure. The potential discloser decides whether the intent of the party requesting information—who we label the "counterparty"—is benign or not in a very similar way as the subjects in the ultimatum game: they compare the terms the counterparty offers for their privacy protection with the terms that the potential discloser believes the counterparty's action space—i.e., the set of choices available to the counterparty--permits. If the counterparty's offer is the most beneficent her action space permits, then the potential discloser is likely to attribute a benign intent to the counterparty. The potential discloser is more likely to trust the counterparty and more likely to reveal information. If, by contrast, the counterparty fails to offer a beneficent action that is perceived to be within her action space, then the potential discloser is more likely to understand the counterparty's intention to be not benevolent, and the potential discloser will be correspondingly less likely to disclose information. Within this framework, what varies when potential disclosers make privacy decisions that appear to conflict is not the potential disclosers' privacy preferences, but their perception of the privacy risk to which they are exposed. That risk is lower if the party seeking disclosure is trustworthy. An individual may reveal personal information to a party she trusts, but may refuse to disclose the same information to a different party whom she does not trust.

That trust mediates willingness to disclose personal information may not be surprising. What *is* surprising is that people make these trust judgments based not on credible information that is reasonably accessible to them (regarding, for example, the technology that protects their data or the reputation of a platform they sue to store their data), but rather on a simple cognitive heuristic that considers very little information. For example, people employing the heuristic may simply fail to take up and process information about the technological aspects of privacy risk—i.e., how secure a particular social media platform or cloud service is, or the number of people that potentially have access to the confidential information if it is shared in a certain way. These relevant data may be ignored in favor of fast and frugal judgments focused on the trustworthiness of interaction partners.

This is not to suggest that the trust heuristic has no value. In the offline world, and especially in small communities based on face-to-face interactions, individuals employing the trust heuristic may be behaving adaptively. But the trust heuristic is vulnerable to manipulation by counterparties behaving strategically. In particular, the heuristic is less likely to be helpful in a situation that presents a pooling equilibrium in which both beneficent and bad-faith counterparties face incentives which lead to the same behavior. Consider, for example, a privacy transaction in which a beneficent actor offers a privacy protection in the expectation that it will be used, versus the same transaction in which a bad-faith actor offers the same protection yet in the expectation that it will induce trust, and, by doing so, is in fact *less* rather than more likely to be

used. In such cases, the heuristic leads to misinterpretation as individuals cannot draw reliable conclusions about their counterparty's intentions. In such cases, use of the heuristic can leave a party considering disclosure open to exploitation.

To investigate our theory, we designed experiments to analyze how individuals estimate privacy risks. We employ a model proposed by Schaechtele et al that explains how individuals ascribe positive and negative intentions to parties they interact with.[35] We assume that when deciding whether or not to disclose private information in a particular transaction, individuals classify their counterparty as falling into either a category of "trustworthy" or "not trustworthy". To make this classification the decision-maker constructs an "action space" of their counterparty—i.e., the set of choices that the decision-maker believes (correctly or incorrectly) that the counterparty is free to take in the context of their interaction. The action space includes actions the counterparty could have performed but did not (what we call the "counterfactual") as well as actions the counterparty in fact did select. Is a counterfactual action more beneficent (i.e., more privacy protecting) than the action that the counterparty actually took? If so, then the counterparty is identified as a "negative intention" type and is likely to be treated as not trustworthy. If, on the other hand, the counterfactual actions are less beneficent than the choice the interaction partner did in fact make, then the counterparty is classed as a "positive intention" type and is likely to be trusted. All else equal, we expect the decisionmaker to release more sensitive information to a counterparty she has classed as "positive intention" and therefore trusts.

The theory, if it is correct, would account for why people seem to act as if they have inconsistent privacy preferences. A person may express strong preferences for privacy protection and yet be willing to disclose sensitive information to a counterparty she perceives as manifesting positive intention. This trust judgment may drive seemingly inconsistent privacy choices in transactions that impose very similar objective privacy risks and without the individuals' privacy preferences changing. When operating within the framework of the trust heuristic, a person may reveal confidential information to one online banking service but not to another, or to one online merchant but not to another, depending on their perception of their counterparty's intent and trustworthiness.

III. Experiments

We designed three experiments to explore our theory that individuals' privacy choices are often based on heuristic decision-making, and, if indeed that is the case, to understand the implications for privacy in the online environment.

A. Experiment #1: The Trust Heuristic and Risk Perception

1. Motivation

[35] Supra note 35.

In the real world, trust may grow out of friendship, past interactions, mutual disclosure of information, or many other factors. But in the online world people often have no (long-term) relations with those they interact with and information about counterparties may be scarce or costly to access. If people interacting online base privacy decisions on trust, they must use less individualized and context-rich criteria to decide whether they should trust someone seeking private data. As explained, we assume that people employ a trust heuristic to make that decision, a strategy that tries to exploit revealed intent. If we can identify this heuristic as a mental strategy that people often use to make privacy decisions, then regulators could design interventions that would interfere with this decision-making process—most likely by blocking the heuristic and attempting to induce privacy holders to look for more objective information relevant to trust decisions.

Our first experiment aims to empirically establish that individuals indeed employ the trust heuristic when making privacy decisions. For a systematic test, we implement a set of three treatments: In the first we've constructed a setting in which we expect subjects if they employ the heuristic to conclude that their counterparty is trustworthy. In the second treatment use of the heuristic should suggest the counterparty is not trustworthy and subjects should therefore reveal less information. Finally, in our third treatment we switch the heuristic off by eliminating the counterfactual. Potential disclosers can apply the heuristic only if their counterparty has a choice between different options. The heuristic exploits whether the decision-maker chooses the more benevolent of her options or the less benevolent. However, if the counterparty can only implement one approach or behavior because she is obliged by law to do so the heuristic is, in effect, blocked from operating—it has no choice to process. Once deprived of the heuristic, we expect subjects to look for cues other than revealed intent to estimate their actual privacy risk (we test this expectation directly in Experiment 2). Since in Experiment 1 we provide very few cues (like the names of the experimenters or their affiliations or links to institutions or publications), we expect subjects to be less likely to reveal versus subjects in our first treatment, where subjects can use the counterfactual to ascribe positive intent. With this last treatment we also begin to describe a strategy for blocking the heuristic, which might be an effective yet simple and cheap adjunct to privacy regulation designed to push people toward making more fully-considered privacy decisions in settings where they would otherwise be able to do so.

2. Experimental Design

We used as a platform for Experiment 1 an unrelated online study, to which subjects had previously been invited. The unrelated study offered subjects the opportunity to enter into a contract and to complete a real effort task (counting

numbers in a table) in return for a payment.[36] After subjects had completed the real effort task the experiment was, in their perception, over.[37] They entered the experiment's payment screen unaware that the choices they make regarding their payment and privacy protection were the object of the study we report here.[38] This element of field work increases the reliability of our findings.[39]

We presented subjects with a choice set of different payment methods. Two methods required disclosure of confidential information while one fully preserved the subjects' anonymity. The first, the use of a bank transfer, required the subjects to indicate their full name and banking information to remit payment—we refer to this payment method as "bank transfer."[40] The second, use of a PayPal account ("PayPal") revealed the name of the participant and her PayPal address. When we recruited subjects, we made it a condition that they had a PayPal account (or were willing to open one before participation) associated with an email address that revealed their true name. The third payment method was fully anonymous ("anonymous payment"), requiring no disclosure of personal information at all. Subjects were asked to specify a 5-digit code after the experiment. With this code, but without revealing their name or any other information linked to their identity, subjects could pick up their payment in the office of the University's student government after the study was finished. This payment protocol ensured that experimenters do not learn the names of the subjects and cannot connect the subjects' appearance with their choices in the experiment.

The bank transfer payment option was fast and convenient, but the anonymous payment option was designed to impose significant transaction costs on the subjects. Subjects choosing the anonymous payment method as a means of preserving their privacy had to make a trip of fifteen minutes or more from their department to the student government's office and then invest an additional five minutes in order to be paid. They also had to make themselves available during one of only two time slots of three hours each that we offered to them to pick up their earnings. These transaction costs assure us that the participants did not randomly choose the anonymous payment option, but that they chose it in order to preserve their privacy and had a willingness to pay with their time for this protection.[41]

---

[36] The study demonstrated that people can employ their own loss aversion to improve effort and reach their work goals—an ability we refer to as "behavioral-self-management." See Stephan Tontrup & Christopher Jon Sprigman, Experiments on Self-Nudging, Autonomy, and the Prospect of Behavioral-Self-Management, 19 J. of Empirical Legal Stud. 594 (2022).

[37] See Tontrup & Sprigman, supra n. 37, at PAGE.

[38] In our data analysis we control for whether the real effort task has an influence on the subjects' payment choices. For example the amount subjects have earned may influence which payment option they select, or how the way the contractual threshold was set may have an impact on their choices.

[39] John List & Glen Harrison, *Field Experiments*, 42 J. OF ECON. LIT. 1009-1055 (2004).

[40] Subjects choosing bank transfer were required to indicate their name, the name of their bank, and their account number.

[41] Given that student jobs in Münster pay approximately €8 per hour the price for protecting their privacy they were willing to pay in the form of transaction was about €2.20.

Our experimental treatments vary the subjects' choice set of payment options. We randomly assigned participants to one of four treatments. All treatments present participants with two specified payment options to choose between. They also permit participants to opt out of these two fixed options. Subjects who opted out were required to provide an anonymous email address for experimenters to work out an individual payment solution with them (we provided a link to a service that offers anonymous email accounts to users at no charge).

While bank transfer was offered in each treatment, our manipulation varied the second payment option. This second option was either to anonymously pick up their earnings in the student government's office or to have the money transferred via PayPal, which required subjects to indicate their name and the email address registered with PayPal. The four treatments were structured as follows:

Negative Intent. Participants were offered to choose between (1) bank transfer and (2) PayPal. An anonymity-preserving payment was not offered; subjects had to indicate their full name and then either their bank details or an email address registered with PayPal. As in each treatment, subjects could opt out and reject the two offered payment options and arrange an alternative payment solution with our assistant. Because the counterparty could have chosen to offer an anonymous payment method but did not, the counterfactual action the counterparty did not choose appears to be more benevolent than the action the counterparty did offer. The treatment is therefore designed to suggest that the counterparty is *not* trustworthy. We therefore expect subjects to ascribe a negative intent to their counterparty and accordingly to be hesitant to disclose confidential information.

Positive Intent. In the second treatment, subjects were offered (1) bank transfer and (2) anonymous payment and were thus empowered to conceal their identity. Again, subjects could opt out of these two specified options and arrange an alternative solution. Because the experimenter's available action space in this treatment includes the possibility not to offer anonymous payment but this option is in fact offered, the counterfactual action should appear less benevolent to subjects than the offer the experimenters actually made. Therefore this treatment is designed to suggest that the counterparty is trustworthy, and subjects should more willingly disclose private information.

Legal Requirement. In the third treatment, *Legal Requirement*, we offered participants the option of (1) bank transfer and (2) anonymous payment—the same two payment options as in the *Positive Intent* condition—and subjects were also able to opt out of these specified payment methods (as in all conditions). The payment options offered should, as in *Positive Intent*, suggest that the counterparty is trustworthy. However, in the *Legal Requirement* treatment subjects were instructed that the data protection policy of the University of Münster obliges all experimenters to allow

participants in scientific studies to receive payment without revealing their identity.[42] This message constrains the action space of the counterparty; experimenters must select the benevolent action as it is their only choice in compliance with the policy. Consequently, subjects' use of the heuristic is blocked: They cannot ascribe an intention to their counterparty based on the relative benevolence of the selected behavior compared to a counterfactual, since there is no counterfactual option fitting the policy. Thus, the treatment is designed to strip the subjects of their strategy to estimate their counterparty's trustworthiness and we expect them to be become more hesitant to reveal privacy information.

Expressive Signal. The final treatment provides a robustness check aiming to rule out an alternative explanation for why subjects may be less likely to release sensitive information in the *Legal Requirement* treatment. It is possible that by informing subjects of the university's data policy, we are sending a signal that subjects should consider privacy protection to be important and desirable. This might reduce the willingness of subjects in the *Legal Requirement* treatment to disclose personal information.[43] To rule out this alternative cause, we designed an *Expressive Signal* treatment where subjects could again (as in both *Positive Intent* and *Legal Requirement*) choose between (1) bank transfer and (2) anonymous payment, or arrange an alternative solution. However, subjects in the *Expressive Signal* treatment were given a different message about the university's data policy than in the *Legal Requirement* treatment: subjects in *Expressive Signal* were told that a data policy had been enacted obliging experimenters to offer anonymous payment, but that the regulation did not yet apply to the study they are participating in "as this study is conducted before the provision comes into effect." Subjects are further told that the experimenters "decided to offer an anonymous payment option … even though the current policy does not oblige us to do so."[44]

The treatment should send the same signal about the importance of privacy as *Legal Requirement*—if indeed our messages are sending any such signal. Even though the provision has not yet come into effect, the policymaker has already expressed its intent and subjects were informed that the provision was enacted. At the same time, and in contrast to *Legal Requirement*, this treatment allows subjects to apply the trust heuristic. As the provision is not yet in effect, it cannot limit the experimenters' current action space, and therefore if they provide an anonymous payment method this choice can be compared to the less beneficent counterfactual of not offering privacy

[42] Typical lab policies in Germany require that data are anonymized such that names and experimental behavior can never be connected. This applies also for financial data. However, there are different ways to conform to this requirement: In the lab for example participants typically are paid in cash. Often the lab assistant who pays the participants is a different person than the assistant who has helped subjects with the study. For online studies payments can be transferred with the consent of the participants examples for such policies are … .

[43] We are indebted to Florencia Marotta-Wurgler for suggesting this point.

[44] We debriefed subjects after the session that the policy had already been in place. The subjects in this treatment were not included in the general subject pool.

protection. Thus, the treatment enables us to discriminate between whether subjects base their decision on the trust heuristic or on the potential policy signal: If subjects decide based on the trust heuristic, they should trust their counterparty and disclose private information more readily compared to subjects in the *Legal Requirement* treatment. Alternatively, if subjects respond to the expressive signal send by the privacy policy, their willingness to release sensitive content should decrease relative to the *Positive Intent* treatment.

3. Methods

Participants were either current or former students of the University of Münster in Germany (about 30% were professionals who had recently graduated) who we had recruited for our empirical legal studies subject pool. We sent potential participants an email invitation and a link in that email directed them to the study. We used the open-source LimeSurvey web app as the online platform to conduct the experiments.[45] As the main interest of our study is in privacy choices in a digitized world (i.e., in social media platforms, cloud computing, and the like), we chose to conduct the experiment online to improve the study's ecological validity. The interaction online is less personal than in the laboratory and better resembles the setting in which many privacy decisions on the internet are made. In a laboratory, by contrast, participants would learn who operates the lab and conducts the study and might build trust based on those personal contacts before and during their session.

Our study design also offers the ecological validity advantage of field work, as subjects are not aware that they are still participating in an ongoing experiment when they make their payment choices. This design prevents demand effects—i.e., the phenomenon that subjects conform their behavior to what they assume is socially desired.[46] Also, in many lab experiments subjects perform artificial tasks that mimic realistic behavior and they are paid according to a payoff function. In our study, subjects are presented with a realistic privacy choice they often have to make outside the context of the experiment: they are asked to reveal actual personal information and the consequences of that action can be the same as in other social contexts where they may perform this action—when, for example, they purchase goods in an online shop and are asked to decide whether to pay with PayPal or with their credit card. At the same time our design also affords better control over the treatments than is typical in

[45] See http://www.limesurvey.org/. Note that we took precautions to ensure the validity of responses. After participants log in, the link becomes inactive, ensuring that the same participant can complete the study only once. The email informs the participants about the time a typical subject requires to complete the experiment, thereby ensuring that subjects do not discontinue participation because it is taking them longer to finish than expected.

[46] Field experiments are less likely to induce demand effects, as they are designed to unobtrusively assess the effects of realistic treatments on subjects who would ordinarily be exposed to them, often with "participants" not being aware that their behavior is subject to research.

field studies, which often must contend with a host of interfering factors which are difficult to control for.

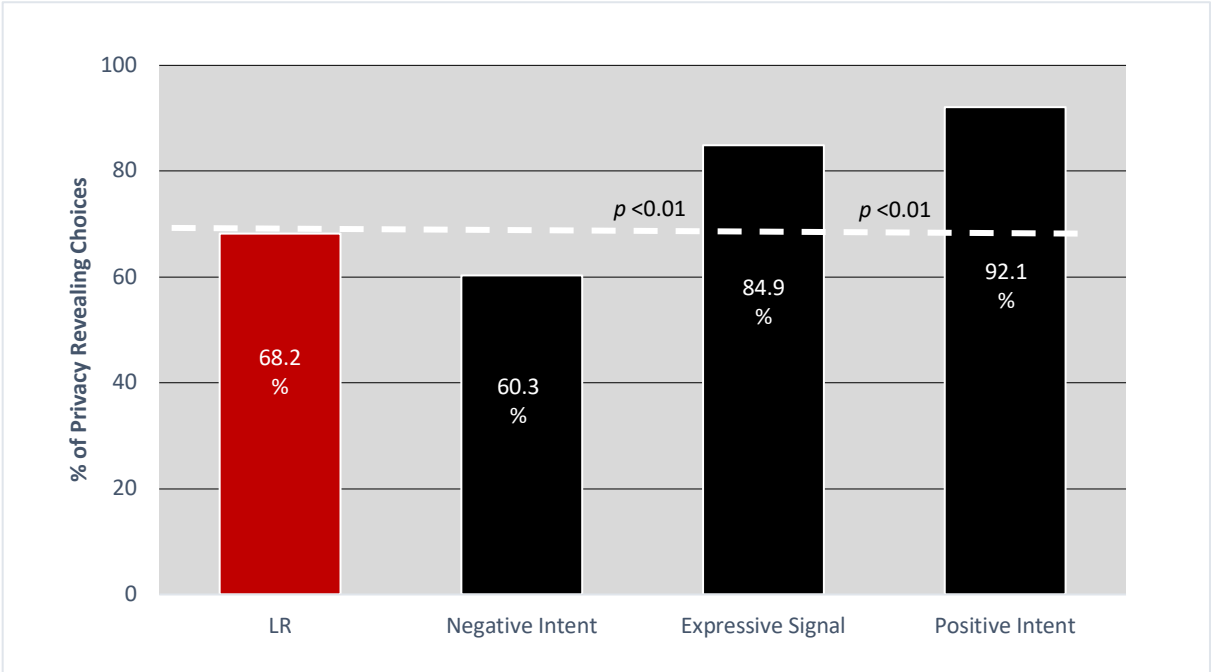
4. Hypothesis and Results

The goal of Experiment 1 is to establish that subjects rely on heuristic decision making to assess trustworthiness when they make privacy choices online. To examine whether that is true, we first compare subjects' behavior in the *Negative Intent* and *Positive Intent* treatments. If subjects indeed use the trust heuristic, then subjects in *Negative Intent* should make characteristically different privacy choices than those in *Positive Intent*. Recall, in the *Negative Intent* treatment subjects are offered only payment options that require them to reveal their identity and confidential information. If they apply the heuristic, they should therefore perceive the counterfactual—the offering of a privacy-protecting payment option—as comparatively beneficent. On that basis subjects should ascribe to the counterparty a negative intention and treat her as not trustworthy.

In the *Positive Intent* treatment by contrast subjects offered anonymous payment should perceive the counterfactual—the counterparty might have withheld any privacy-protecting payment option—as less beneficent. Thus, if subjects apply the heuristic they should ascribe a positive intent and trustworthiness to the counterparty.

This leads us to our first hypothesis: we expect subjects in the *Positive Intent* treatment as they should trust the counterparty more to choose the bank transfer significantly more often and, thereby disclose more personal information than subjects in the *Negative Intent* treatment.

Figure 1. Percentage of subjects who reveal confidential information across treatments



Our results support this hypothesis. In the second column of the graph you see the *Negative Intent* treatment, on the right side the *Positive Intent* treatment. The graph shows that 60.3% (32/53) of the subjects in *Negative Intent* choose the bank transfer, while 39.7% (21/53) chose to opt out of the offered payment methods bank transfer and PayPal and preferred to pick up their earnings in the student government's office.[47]

In the *Positive Intent* treatment 92.1% (47/51) of the subject chose bank transfer and revealed their name and banking details, compared to the 60.3% we observed in *Negative Intent*. The difference between the treatments is strongly significant (p=0.0002 Fisher Exact).

In a second step we investigate whether we can block subjects from relying on the trust heuristic, and, if so, how that affects their behavior. In our *Legal Requirement* treatment we offer subjects the same payment options as in *Positive Intent*, but we manipulate their perception of the counterparty's action space by instructing them that the law requires the counterparty to offer a privacy-protecting payment option. By restricting the counterparty's perceived action space our manipulation eliminates the counterfactual, and should preclude subjects from employing the heuristic to decide whether they can trust their counterparty. Our second hypothesis therefore predicts that subjects in the *Legal Requirement* treatment should be less willing than subjects in the *Positive Intent* treatment to reveal personal information and opt for the bank transfer payment, even though subjects in both treatments are offered the same payment options.

Our results strongly support our second hypothesis: while we have seen that in the *Positive Intent* treatment 92.2% (47/51) of subjects choose bank transfer and only 7.8% wish to conceal their identity and confidential information, disclosure drops significantly in the *Legal Requirement* condition, where only 68.6% (35/51) of the subjects select bank transfer and 31.4% conceal. This treatment difference is strongly significant (p=0.005; Fisher Exact).

The results show, as we had theorized, that many subjects indeed appear to use the trust heuristic to determine their counterparty's trustworthiness and that they base their privacy decision on this information. When the heuristic suggests subjects should trust their counterparty, as in *Positive Intent*, participants are indeed much more likely to disclose private information compared to when the heuristic suggests not to trust the counterparty, as in the *Negative Intent* treatment. And when we block subjects' use of the heuristic in *Legal Requirement*, subjects become more cautious, even though they are offered the same payment options as in the *Positive Intent* treatment.

[47] Recall, that subjects experienced a real cost in opting out of the bank transfer and PayPal options: the fully-anonymous method was costly in terms of time and effort required to collect payment.

Finally, we want to confirm our theory using the robustness check provided by the *Expressive Signal* treatment. In this treatment, we aim to reject the alternative explanation that the results in *Legal Requirement* may be driven by a normative message sent by the laboratory policy that a privacy-protecting payment option should always be made available to participants. The *Expressive Signal* treatment is identical to the *Legal Requirement* treatment except that it instructs the subjects that even though the data protection rule was enacted it is not yet in effect. If subjects are responding to the information the law's expressive signal seems to convey, the fortuity of timing should not change that information. But unlike in *Legal Requirement*, subjects can apply the trust heuristic: as the provision does not yet limit the counterparty's action space, subjects should ascribe to the counterparty who voluntarily offers a privacy-protecting payment positive intent and trustworthiness.

Thus, our third hypothesis assumes that subjects in *Expressive Signal* can use the trust heuristic and will disclose more personal information than subjects in *Legal Requirement,* whose use of the heuristic is blocked, even though subjects in both treatment should have received the same express signal. The results support our hypothesis: As Figure 1 shows, 84.9% (135/159) of the participants choose the convenient and costless bank transfer in the *Expressive Signal* treatment, and only 24 subjects or 15.1% opt for the privacy-protecting anonymous payment method. In *Legal Requirement* by contrast we find only 68.6% choosing bank transfer. The treatment difference is significant (p=0.02 (Fisher Exact)). So the data rejects that subjects in *Legal Requirement* were responding to an expressive signal.[48] As predicted by our theory, the frequency of subjects in *Expressive Signal* choosing bank transfer (84.9%) is statistically indistinguishable from the frequency we observe in *Positive Intent* (92.1% (p=0.27; Fisher Exact), suggesting that subjects in both treatments apply the trust heuristic ascribing positive intent and trustworthiness to their counterparty.

In sum, the results of Experiment 1 establish that subjects often make online privacy decisions using a trust heuristic. And the heuristic is impactful: when the application of the heuristic suggested that the subject should trust the counterparty, it increased the rate of disclosure by almost 30%. So the heuristic drives decision-making. But heuristic decision-making by definition does not consider all relevant information. So will the heuristic lead subjects to ignore relevant information about technological or organizational privacy risks? That is the question we will analyze in Experiment 2.

B. Experiment #2: "Crowding Out" Credible Signals

[48] Indeed, it might be argued that in the *Expressive Signal* treatment we *amplify* whatever expressive signal is sent by the legal mandate. That is because we indicate, through our provision of privacy protections, that we agree with the expressive content of the legal rule even though we are not bound by it. As a result, our *Expressive Signal* treatment is essentially conservative—if the effect on subjects' behavior is driven by the law's expressive content, we would be more likely to have picked it up.

1. Motivation

Heuristic decision making is frugal; it ignores part of the relevant information that might otherwise be considered. It may nevertheless lead to efficient decision making[49] if it can exploit a match between cues used and the decision environment. The recognition heuristic is an example of a heuristic that is adaptive at least in some contexts. The recognition heuristic holds that "if one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion." Recognition can be a valid cue if there is a high correlation between the recognition and a criterion that predicts the outcome well—for example firms with better products often survive longer and are therefore more likely to be recognized by the consumer. In such cases, the correlation of recognition and criterion can make the heuristic ecologically adaptive. On the other hand the recognition heuristic is likely to fail when there is a bad fit between recognition and criterion. Consider an American using the heuristic to decide whether Santa Barbara, California or the Chinese city Chongqing is larger. Chongqing is one of the largest cities in the world, yet the average American is likely not to have heard of it. The heuristic fails because China is so distant from the United States that even large cities may not be recognized. In these circumstances, the connection between recognition and criterion is too loose for the heuristic to work reliably.[50]

In the same way the trust heuristic may also be prone to failure in the online world. The perception that the counterfactual action of the disclosure-seeking counterparty may have been more or less benevolent must be predictive of the potential discloser's true privacy risks. Yet, in a digital world where the privacy risk is diverse and affected by the choices of many and not just an identifiable counterparty the heuristic seems to be just as loosely connected to true privacy risks as the recognition of a Chinese city in America with its size.

That alone may not render the use of the heuristic problematic, if other relevant information was not crowded out. Yet, heuristic decision making is meant to be frugal, fast and efficient in uncertain environments. Therefore, heuristic thinking economizes by limiting the information that is processed;[51] for example the recognition heuristic ignores in our above example whether the technical description of one product suggests more reliability than the product of a recognized brand, or whether the portfolio of one company may suggest more expertise in the product's area. In the

---

[49] Gerd Gigerenzer, Wolfgang Gaissmaier, Heuristic Decision Making, ANNU, REV. OF PSYCHOL., Vol. 62, 451-482, 2011.

[50] Czerlinski, Jean; Goldstein, Daniel G.; Gigerenzer, Gerd (1999). "How good are simple heuristics?". Simple Heuristics that make us smart. New York: Oxford University Press. pp. 97–118.

[51] Simon points out the trade-off between rational decision making and heuristics that may perform well under uncertainty but that also can lead to failures when most relevant information is ignored. Simon speaks of bounded rationality: Simon HA (1977) The logic of heuristic decision making. In: Models of discovery and other topics in the methods of science. Boston studies in the philosophy of science book series (BSPS), vol 54, pp 154–175

same way, the trust heuristic might lead individuals to ignore information that would help them to better estimate their true privacy risks.[52]

That possibility is what Experiment 2 investigates: do subjects applying the heuristic ignore valid information that might otherwise inform their privacy decision?

2. Design and Methods

In Experiment 2 we analyze whether signals of high or low credibility affect subjects' privacy decisions when they apply the trust heuristic. We recruited new subjects via email and invited them to participate in future studies. When invitees entered the registration website, they were instructed how to register for the future experiments and prompted to provide their payment information. Invitees were also instructed that if they decided not to participate, they could discontinue the registration process by clicking "No".

The design of Experiment 2 varies three different payment options for subjects to choose among: bank transfer and PayPal, which were offered in all treatments and either anonymous pick up or mail, which were varied across treatments.[53] We implemented three treatments—*Positive Intent, Negative Intent* and *Legal Requirement*—that are structured similarly to those treatments in Experiment 1. In *Positive Intent* and *Legal Requirement* subjects were offered bank transfer, PayPal, and the anonymous payment method allowing them to use a code to pick up their earnings, as described above.[54] And just as in Experiment 1, subjects in the *Legal Requirement* treatment were informed that the University's data policy required experimenters to enable anonymous participation in the study, while by contrast in *Positive Intent* the experimenters appeared to provide the anonymous payment method voluntarily. Subjects in the *Negative Intent* treatment were offered only payment options that required them to indicate their full name and additional personal information: they could choose between bank transfer or PayPal, or they could indicate their home address to receive payment by mail.

Experiment 2 deviates from Experiment 1 by manipulating the availability of other information that subjects could utilize to make an informed privacy decision. We implement for each of the treatments (*Positive Intent, Negative Intent* and *Legal Requirement*) two informational conditions: *positive signal*, in which we provide credible information sending a positive signal of the counterparty's trustworthiness, and *negative signal*, in which we do not provide any information about the counterparty at all, which should send a negative signal (relative to the other condition) about the counterparty's trustworthiness. In *positive signal*, we presented subjects with biographical information

---

[52] Gerd Gigerenzer, Wolfgang Gaissmaier, Heuristic Decision Making, ANNU, REV. OF PSYCHOL., Vol. 62, 451-482, 2011.

[53] Subjects were not given an "opt out" choice to arrange an alternative payment solution, as they were in Experiment 1. Instead, subjects were free not to register for the future studies.

[54] See Section III.A.3.

about the researchers conducting the experiment, including our ties to academic institutions (New York University, and University of Münster), and displayed the emblems of the researchers' universities on top of each screen presented to the subjects. We also provided online links to our academic work and gave subjects the University of Münster email address of our research assistant, informing them that they were free to contact her. By contrast, in *negative signal*, we provided only our RA's *non-university* email address as contact, and gave no information about us (that is, nothing about affiliations, past research, etc.).[55]

We recruited the subjects via email sent over a server of the University of Münster that reaches more than 40,000 students. Invitees interested in participation clicked on a link and were directed to our experimental website hosted by LimeSurvey. As in Experiment 1 it was crucial that we conducted Experiment 2 online. The condition in which we provide little or no information is foundational to our design because it sends to subjects a negative signal suggesting that the counterparty's trustworthiness is low (for a real-world analogue, imagine a website which does not allow the user to trace who is running it and who is responsible for the platform and its content. When prompted to disclose private data, this lack of transparency should send a negative signal to anyone using the website). Such a low-information condition is easily implemented online; it is very difficult, however, to reproduce in the laboratory, where institutional information (the university responsible for the study; the researchers who operate the lab, etc.) is a given, and where personal face-to-face interactions tend to build trust.

Experiment 2 shares some of the same field elements as Experiment 1. Subjects make decisions in a realistic environment where they decide whether or not to disclose actual personal information. And when subjects make their decision, they are likely to perceive the solicitation of their choice of payment method to be an administrative act, and not an experimental task of relevance for our research.

Note that subjects (just as in Experiment 1) bear transaction costs if they decide to preserve their privacy; selecting anonymous payment forces them to invest more time and effort to pick up their earnings from future experiments compared to accepting payment by convenient bank transfer or PayPal.

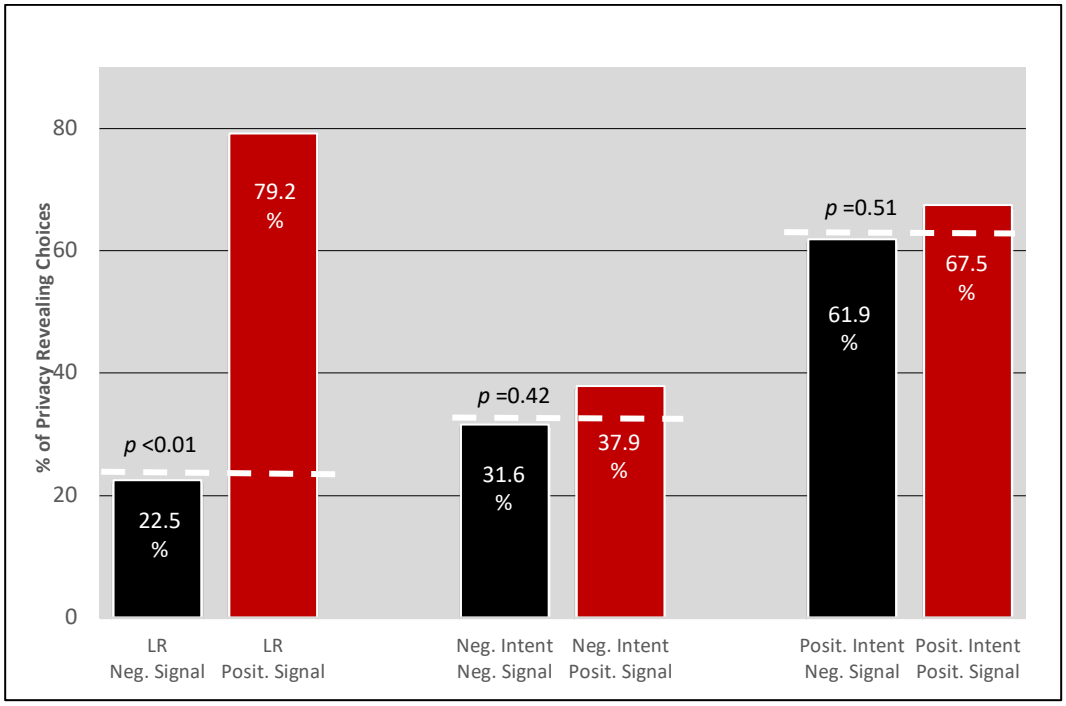
3. Hypothesis and Results

We measure first whether subjects using the trust heuristic respond to signals suggesting either low or high credibility of their counterparty by changing their privacy choices. Intuitively, one would expect that subjects are less likely to disclose personal information if they receive a negative signal suggesting low credibility and are more

[55] Since the invitations were sent over the university server, the subjects may assume that we had to meet some minimum privacy standards to be given access.

likely to disclose if they receive a positive signal. However, like many heuristics[56] the trust heuristic bases judgments on a single criterion that is perceived to discriminate effectively between the choice options: whether the counterparty choses from the options available to her the one most favorable for the potential discloser. All other cues are not considered. Therefore, individuals relying on the trust heuristic can be expected to ignore both the negative and positive signals they receive. The two informational conditions in Experiment 2, *positive signal* and *negative signal*, test whether these signals are ignored, leading us to a first hypothesis: when subjects apply the trust heuristic, neither negative nor positive signals will substantially affect their privacy choices.

Figure 2. Subjects Disclosure Choices by Type of Signal Received and Treatment



Our results support this hypothesis. In Figure 2 you see for each treatment *Legal Requirement*, *Negative Intent* and *Positive Intent* two bars reporting the two informational conditions, *negative signal* and *positive signal*. The more similar the bars within each treatment are, the smaller is the differential impact of the informational signals on subjects' behavior. We focus first on *Negative Intent* and *Positive Intent*, because in both treatments we expect subjects to apply the trust heuristic. Percentages reported in Figure 2 refer to the frequency of decisions to disclose personal information—i.e., to

[56] Gigerenzer, G. & Goldstein, D. G. (1996). "Reasoning the fast and frugal way: Models of bounded rationality". Psychological Review, 103, 650-669.

choose the payment methods (bank transfer, PayPal) that require disclosure. For our analysis, we employ a Fisher test (2x2 contingency table) for the *Negative Intent* treatment; that is, we test whether the two informational variables (*negative signal* and *positive signal*) and the subjects' privacy choices (i.e., disclosure=bank transfer, PayPal as opposed to non-disclosure=the anonymous payment method) influence one another, that is whether the informational signals affect subjects' disclosing behavior. The result is insignificant (p=0.45 Fisher Exact), indicating that the informational cues do not change the likelihood of disclosure in the *Negative Intent* treatment. When we perform the same analysis for the *Positive Intent* treatment, we get the same result: the relationship is again insignificant (p=0.24 Fisher Exact), which means that informational signals appear to not affect privacy choices in *Positive Intent* either.

These results support our first hypothesis: subjects applying the trust heuristic seem to give negative as well as positive informational signals little or even no weight in their decision making. That subjects ignored negative informational signals is perhaps most concerning in the context of the *Positive Intent* treatment, for there the heuristic suggests disclosure and subjects are ignoring signals that are at odds with the heuristic's conclusion that their counterparty is trustworthy. In the opposite case—subjects in the *Negative Intent* treatment fail to heed the positive informational signals and remain reluctant to disclose, the heuristic can strip the individual of opportunities for beneficial disclosures and/or lead to unnecessary transaction costs, as we see in our study, where subjects who ignore positive information invest time and effort to pick up their earnings in person using the anonymous payment method.

We next investigate whether subjects will heed valid positive or negative informational signals when the trust heuristic is unavailable to them. In the *Legal Requirement* treatment we prevent subjects from using the trust heuristic, as we have seen above. If the heuristic is blocked the subjects need to search for and rely on other information in order to make their privacy decisions. This may allow our participants to process and respond to the credibility cues they were given. In our second hypothesis we posit that if subjects cannot rely on the heuristic, they will consider signals we provide them with in their decision making and their privacy behavior will change, according to what type of signal—positive or negative—they are presented with.

Our results support this second hypothesis. For subjects in the *Legal Requirement* treatment we observe a strong impact of both the negative and positive signals on privacy choices; in Figure 2 we can see that 79.2% of the subjects in the *Legal Requirement* treatment who receive a positive signal disclose private data, but only 22.5% who are presented with a negative signal disclose. The test for statistical relationship is highly significant: $p < 0.00001$ (Fisher Exact)—indicating that the informational signals clearly had an effect on the privacy behavior of the subjects in *Legal Requirement*, who were precluded from employing the heuristic.

This is an important finding: our *Legal Requirement* intervention that made using the trust heuristic impossible was able to open up the decision-making process and make effective other informational cues which appear to be ignored when subjects' decision-making is based on the heuristic. As a consequence, blocking use of the heuristic appears, at least in settings where otherwise credible informational signals are available, to lead individuals to better estimate their privacy risks.

C. *Experiment #3: The Heuristic and the Public Good Characteristics of Individual Privacy Decision*

1. Motivation

In Experiments 1 and 2 subjects could disclose strictly-defined types of information: their name and bank information or their PayPal address. They had full awareness and control over what information was transmitted if they decided to give us their data. In contrast, the risks of privacy decisions in social media and other social networks are inherently less predictable: *data is typically not in the control of the discloser alone*, but can be accessed, altered, or even created in the first place by other users of the network, and even beyond by the wider audiences with which the data is shared, often without the discloser's knowledge. A concise way to say this is that in social networks, data (and privacy) are a product of user *interactions*.

We capture this dynamic of social networks in Experiment 3. In this experiment, the Facebook information subjects can reveal to us is created and altered by a potentially unlimited number of other users, who may link their own content to the subjects' accounts, make public formerly private interactions with the subject, or share subjects' Facebook postings with other audiences who may then add content themselves. As a consequence, from the perspective of any individual discloser, the scope of the data that individuals disclose if they allow us access to their account is fundamentally uncertain. Whether people realize it or not, this is a typical scenario for making privacy decisions on social media. An individual may decide to give a company access to their Facebook data in exchange for being allowed to use a web service. Or she may log into a service using their Facebook login for convenience, and as a result the company and Facebook will own her user data and likely will share it with other companies or services. Because the data disclosed may include content of the people the discloser has been interacting with, disclosure *imposes a privacy risk on these other users of the network*. And the risk flows in the other direction as well: data produced or modified by other users on the network but linked to the discloser's account—e.g., a Facebook comment on the discloser's post made by a Facebook friend—*presents a privacy risk to the discloser*. Thus, individual privacy decisions in a networked environment are always intertwined and likely to create externalities.

The story of "thisisyourdigitallife" is an apt illustration of the social dimensions of individual privacy decisions in the context of online networks and platforms.[57]

The privacy implications of "thisisyourdigitallife" may be unusual in scale, but the story is, in an important way, entirely typical: the app didn't steal data; rather, people consented (at least notionally) to disclosure.[58] Lax Facebook rules attribute all content associated with a posting to the user who made that posting; this includes all comments or likes or even photos. The authors of those reactions need not consent when their information is disclosed alongside with the original posting. Under these rules, Facebook permitted the "thisisyourdigitallife" app to access and collect a wealth of information not only about consenting users themselves but also about their Facebook friends, who had never given their consent[59] This led to a huge multiplier effect. The app collected data from Facebook not only about the 270,000 people who downloaded it, but also from up to 87 million of the downloaders' Facebook friends—people who had never themselves downloaded "thisisyourdigitallife".[60]

This story shows us something fundamental about privacy: privacy is not private, it is a *public good*.[61] That is, in settings where people are connected in networks, an individual's decision to protect her privacy can lead to a "positive externality" whereby others' privacy is protected as well. And conversely, one individual's decision to sacrifice her own privacy can give rise to a *public harm*—i.e., it can result in others suffering "negative externalities" in the form of damage to their privacy. These externalities strip the individual of the power to protect her privacy alone. Privacy protection is a group effort. [62]

---

[57] See Zeynep Tufekci, *Facebook's Surveillance Machine*, N.Y. TIMES, March 19, 2018, https://www.nytimes.com/2018/03/19/opinion/facebook-cambridge-analytica.html.

[58] The app's user agreement stated specifically that the app might gather information on the user's Facebook friends. Lauren Etter & Sarah Frier, Facebook App Developer Kogan Defends His Actions with User Data, BLOOMBERG, Mar. 21, 2018, 5:22 PM) https://www.bloomberg.com/news/articles/2018-03-21/facebook-app-developer-kogan-defends-his-actions-with-user-data. And yet it is likely, as is the case with many online user agreements containing privacy terms, that virtually none of the people who used the app actually read it. See Yannos Bakos, Florencia Marotta-Wurgler & David R. Trossen, *Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts*, 43 J. L. Studies 1 (2014) (study of Internet browsing behavior finding that only 1 or 2 of every 1000 online software shoppers access software license agreements, and most of those access only a small portion).

[59] Alvin Chang, *The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram*, VOX May 2, 2018, 3:25 PM) https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram.

[60] Id.

[61] See Joshua A.T. Fairfield & Christoph Engel, *Privacy as a Public Good*, 65 Duke L. J. 385 (2015).

[62]Gmail is another real-world illustration. In exchange for free email service, Gmail users agree to have their emails searched by machines (the searches are used by Google to target ads). But it isn't only Gmail users' emails that are searched—Google also searches email sent to Gmail users by others, even if the sender is not using Gmail. As a consequence, the privacy behaviors of an individual Gmail user affect not just that individual user's privacy, but also that of the people who correspond with her.

The Gmail example illustrates that privacy harms are often *nonexcludable*—that is, the harms from an individual's decision to sacrifice her privacy cannot be confined to the individual making the decision. And the same is true of decisions to protect privacy—the benefits are nonexcludable because they flow

We hypothesize that the heuristic would lead subjects to disregard the uncertainties of their own privacy risks in networks and the risks they may impose on others, as it reduces the disclosure decision to a judgement of trustworthiness of the person or service they are directly interacting with. We further hypothesize that when we block subjects from using the trust heuristic, they will be better able to consider the social dimension of their privacy decision, and, further, to better consider whether they are willing to limit their own disclosures to reduce the externalities they impose on others but also the externalities the network may impose on them. Thus they may begin to contribute to preserving privacy as a public good.

2. Design and Methods

To implement Experiment 3 we partnered with the company Social Links, which developed and owns a web application that allowed us to elicit all publicly accessible data linked to subjects' Facebook accounts.[63] The application is registered with Facebook and complies with Facebook's data protection rules. The information we are able to collect using Social Links includes all of an individual's Facebook postings, comments and likes that subjects have published and to which subjects have not applied a privacy setting that restricts access. We are also able to collect content that other users have posted about the subjects; for example, photos on which the subjects are tagged and that are not subject to a restrictive privacy setting. To collect the data the application accessed the Facebook profile of participants who consented to disclosure and was automatically disconnected once the information was received. This consent applies also to the comments and likes of others, if they are linked to the subjects' account (we explain this in detail below).

Since not all potential participants have an active Facebook account, we first prompted subjects to indicate which social media platforms they use and in particular whether they have a Facebook account. This information allowed us to consider only

---

not only to the person making the privacy decision, but to others whose privacy is preserved as a result of that decision. To avoid the negative externalities, an individual who doesn't use Gmail must stop corresponding with Gmail users. This possibility becomes more costly and less feasible as the number of Gmail users grows. Unless Gmail users decide to switch away from Gmail, the individual who doesn't use Gmail cannot effectively protect her privacy through her own decision-making. Her fate lies partially in the hands of others.

Similarly, Facebook users have only partial control over their privacy on Facebook. No individual Facebook user is able to understand fully, at the time they consent to Facebook analyzing their data and selling it, what data will be created on Facebook and associated with them. The data that is produced and associated with their profile depends not only on what *that individual* does on Facebook: It depends also on the actions of many who post their comments and likes on the individual's page, or who otherwise link content to the account of another Facebook user—for example, by tagging that user in a post or in a photograph. The content about any particular individual on Facebook is produced by a group of Facebook users in interaction. Once the individual is part of the network she basically loses the control over her privacy, and the more she interacts and the larger the group with which she interacts, the less control she retains over what is produced, and what is disclosed. The only way to avoid these privacy externalities is to stop using Facebook.

[63] See https://sociallinks.io/.

subjects for our experimental sample who could give us access to their Facebook accounts (22.5% of the participants did not have an FB account).

After learning whether subjects have a Facebook account, we asked them to indicate whether they would allow our web application to collect all publicly available Facebook data associated with their Facebook profile. We incentivized their choice, offering subjects a payment of €4.00 (~$4.50) if they agreed to disclose the data. We informed subjects that the information we would collect might include the groups they have joined on Facebook, the interests they indicated, the discipline they study, and the Facebook likes and comments that they have published as well as content that other users have posted and linked to their profile. We instructed the participants that we would use the information to better understand which of our future (social media) studies might be appropriate both for them and for their Facebook friends. We also advised subjects that we would reach out to their Facebook friends and invite them to register for our experiments—without informing their friends that their participation in our study was how learned about and were able to contact them.

Subjects made two decisions. First, they decided whether to permit us access to their account. We then asked them on a separate screen to provide us with their list of Facebook friends (to the extent it was publicly available on their Facebook profile). We incentivized this decision by informing participants that if they disclose only partial information (either their friends list or their Facebook data, but not both), we may reduce compensation accordingly.

We classify the information we collect into two categories. In the first are subjects' own activities: their own postings, comments, photos and likes, the groups they join, pages from organizations or notable people they liked, event pages they visited, etc. The second category is information produced as a result of the subject's connections to other Facebook users: that is, data from all activities other users have linked to a subject's profile like posts, likes or photos in which the subject is tagged, or the fact that the third person has become Facebook "friends" with the subject.

Both categories of information—activities and connections—pose social privacy risks. First, subjects have incomplete control over or understanding respecting what data they disclose to us. We collect the postings they have made anywhere on Facebook, subject to the privacy settings of multiple other users, which might also have been changed after the fact. To be aware in full of what posts they made in public or other users (later) linked publicly to their Facebook profile is practically impossible; those postings may be public or private, depending on the privacy settings of yet again multiple other users, and the subjects may not even know of the content being linked to their profile, because those posts can be spread over the whole network. The risk that information that subjects would have preferred to keep private is revealed obviously rises along with the number of contacts and interactions a subject has, as is

typical for networks. But it is difficult for any subject to have a complete and accurate picture of the true social risks of disclosure.

Subjects face similar difficulties in assessing the privacy risk their own decision to disclose imposes on others. Subjects who opt to disclose are authorizing us to collect other users' postings linked to the subject's Facebook profile. This is consistent with Facebook's own rules.[64] The more content subjects disclose the larger this externality becomes: the more data collectors learn about users in general, the better can they draw conclusions about the behavior of a particular individual. For example, Facebook has acquired a patent for inferring the creditworthiness of particular individuals. Each financial decision a user makes online and she allows Facebook to collect the data of is feed into the algorithm that estimates creditworthiness and that will ultimately decide whether a friend in the same social network are someone who has never even been o Facebook but shares similar characteristics is extended a loan.[65] The data has established a digital redline, the individual seeking the loan cannot influence with her personal behavior. These are the characteristic risks of a public good: the privacy decisions of one user affects not only her own privacy but potentially the privacy of the whole community.

We compare the two treatments *Positive Intent* and *Legal Requirement*, structured similarly to those treatments as they are administered in Experiments 1 and 2: First, in Experiment 3's version of both treatments, we inform subjects that if they wish to register to participate in future studies, they must choose how they want to collect earnings from those studies. We offer them anonymous payment (using a code as in Experiments 1 and 2) alongside three payment methods that require them first to indicate their full name and then confidential information: bank transfer, PayPal, and payment via mail to their home address.[66] Additionally, we inform subjects in both treatments that they can also register to participate in future studies without disclosing the requested Facebook data.

The *Legal Requirement* treatment differs from *Positive Intent* in that the subjects are told, as in Experiments 1 and 2, that the University's rules require experimenters to enable anonymous participation in any research study, such that students must be allowed to register for experiments without providing bank information or their Facebook data.

---

[64] https://www.facebook.com/privacy/policy/?subpage=1.subpage.2-FriendsFollowersAndOther

[65] Joshua A. T. Fairfield & Christoph Engel, Privacy as a Public Good, 65 Duke Law Journal 385-457 (2015) page 390.

[66] At each stage of the experiment, when we (1) ask subjects to list the social media platforms they use, when we (2) prompt them to provide access to their public Facebook data and when they are requested to (3) disclose their friends list, subjects are offered to participate in our studies while remaining fully anonymous, "you can also register for the experiments without providing this information."

We compare across the two treatments how many of the subjects (1) disclose their Facebook data and (2) disclose their friends list. For subjects who choose to disclose data, we (3) compare the amount of data participants reveal.

Experiment 3, is a partial field study: while subjects are aware that their privacy choices are subject to our research, they perform an everyday decision they are accustomed to and that decision could affect their real privacy, as they disclose all their publicly available Facebook data to us.[67] This is the same decision they would make in the real world if, for example, an app on Facebook asks for access to their Facebook data in exchange for allowing them to use the app.

3. Hypothesis and Results

We assume that subjects in the *Positive Intent* treatment will employ the trust heuristic and frame their privacy decision as a bilateral interaction with the experimenters. Basing their decision on the heuristic we expect them to trust the experimenters and to disregard the privacy risks that the decision's social dimension may impose on them as well as on others.

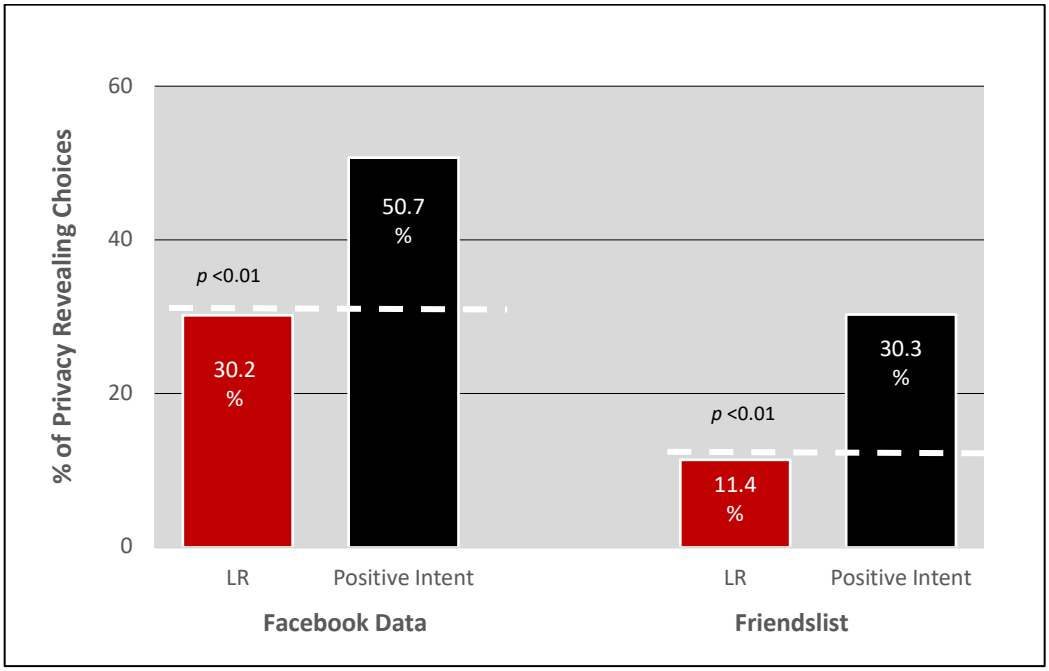
By contrast, we expect that subjects in the *Legal Requirement* treatment who cannot apply the trust heuristic, because experimenters are bound by law and cannot decide freely whether to offer privacy protection or not, will be more likely to consider the public good nature of their privacy decision, and the uncertainty of the privacy risk they would take on and impose on others and as a consequence will be more hesitant to disclose their Facebook profile information and friends list. This leads to our first hypothesis for Experiment 3**:** Subjects in *Legal Requirement* are less likely to give us access to their public Facebook data relative to subjects in the *Positive Intent* treatment.

The results shown in Figure 3 support our hypothesis. In the *Positive Intent* treatment where subjects can ascribe trustworthiness to a counterparty who has offered them an opportunity to participate in the studies without having to reveal their identity, financial information, or Facebook data—i.e., an offer more beneficent than what the subjects perceive the counterparty's action space might otherwise permit—subjects are significantly more likely to allow us access to their Facebook data relative to subjects in the *Legal Requirement* treatment whose use of the heuristic is blocked ($p < 0.01$ Fisher Exact). Apparently, subjects in the *Positive Intent* treatment who we expect to employ the heuristic may fail to consider the social risks of disclosure when such risks are present while when we block the trust heuristic, subjects appear to change

[67] We did not provide subjects with a sample report from Social Links that would demonstrate the capability of their software, as this would have given subjects a state of precise information that is not typical for privacy choices—in most cases data protection protocols remain an abstract description and do not demonstrate vividly the scope of data that is disclosed and how it is going to be used. This uncertainty and vagueness is part of the experiment's ecological validity. After the experiment was completed, all Facebook data was anonymized and will be stored only as evidence for the sincerity of the study.

their decision-making strategy: They may take the social risks of their privacy decision into account.[68]

Figure 3: Subjects Disclosing Facebook Data by Treatments



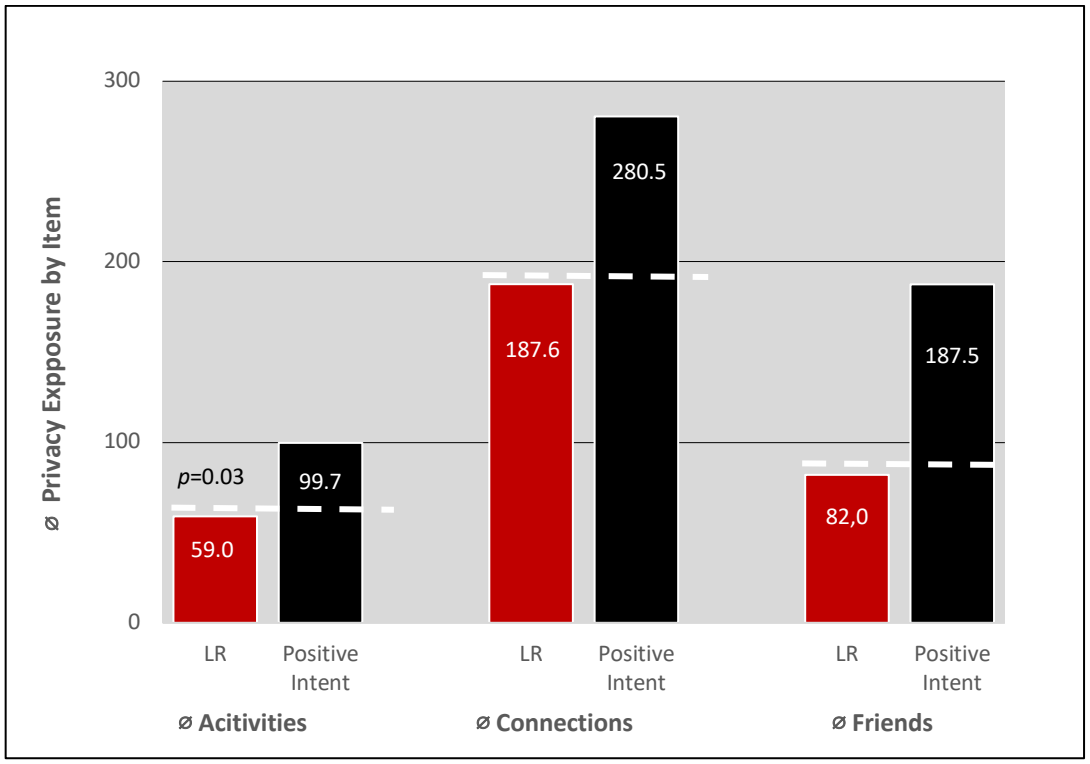
To understand whether subjects indeed consider the social risks of their privacy choices, once they are blocked from using the trust heuristic, we analyze the relative privacy risk of the content that subjects reveal to us. Based on the Social Links reports, we are able to calculate "privacy exposure" scores for each individual subject—i.e., a measure of how much Facebook data an individual subject is disclosing to us that allows us to compare how the privacy risk posed by disclosure varies across treatments. We calculate three privacy exposure scores for each subject: *own activity*, *number of connections* and *number of friends*. Calculating these scores gives us a way to assess whether subjects in *Legal Requirement* are indeed considering the social risks of privacy. If they are, then subjects in *Legal Requirement* with lower privacy exposure scores should be more likely to disclose Facebook data to us relative to subjects with higher privacy exposure scores, while we expect subjects in *Positive Intent* to be relatively insensitive to their privacy exposure scores when making their disclosure decision. This is the second hypothesis for Experiment 3.

The data reported in Figure 4 support our hypothesis. On the left side of the figure, we see the participant's own activities: i.e., the subject's own postings, likes etc.

[68] Recall that as in Experiments 1 and 2 the payment options in the *Positive Intent* and *Legal Requirement* conditions are the same.

Subjects do not have full control of this part of their data. Take for example a comment a subject makes on another Facebook user's post. Recall that the author of the original post may change his or her privacy setting, thereby opening comments or likes which had previously been private. So, this score is capturing some part of the risk associated with the public goods character of privacy.

Figure 4: Privacy Exposure by Choices and Treatments



In the middle of the x-axis we see the "connections" score, which refers to comments or likes *other* Facebook users have linked to the participants' postings or photos on which users tagged the participant. Obviously, subjects have much less influence on the data that is produced by others (of course the more comments they post themselves, the more responses their own activity may trigger). Typically, they will also have far less knowledge of what others post, as parts of that data may not appear on their personal page. Thus, the connection score focuses on the social risk that the subjects impose on themselves when they reveal their Facebook data to us.

Obviously, when we aggregate this data, we cannot determine the relative weight of the privacy concern a particular posting or photo carries for a subject; that is private knowledge. We therefore assign all postings, comments and likes equally the same

value and aggregate all values into one score, one separate score each for the subjects' own activities and their connections.[69]

The results show indeed that subjects who permit us to assess their data in the *Legal Requirement* treatment disclose significantly fewer items on both dimensions: Subjects in *Legal Requirement* revealed fewer own-activities recorded on Facebook than subjects in *Positive Intent* (p=0.026 t-test), and also disclosed fewer connections to other Facebook users than subjects in *Positive Intent* (p=0.036 t-test). Note that if subjects in *Legal Requirement* were simply more cautious overall about permitting us access to their Facebook data than subjects in *Positive Intent*, but were not actually considering the social dimensions of their privacy decision, then we should see fewer participants reveal their data in *Legal Requirement*, but the subjects who decide to disclose data should disclose about the *same amount* of data in both treatments if their decision was blind to the content and the social risk of what they disclose. When we compare the distribution of data points in both treatments, we see in the distribution of the activity and connection scores in *Legal Requirement* compared to the scores of the disclosing subjects in *Positive Intent* almost a total cut-off of disclosures of large quantities of data. This suggests that in particular those subjects in *Legal Requirement* who would have revealed a large quantity of data decided not to permit us access—and that suggests that subjects in *Legal Requirement* were making their privacy choice in light of their increasing exposure to social privacy risks as indicated by their scores for both own activities and connections.

There is one remaining uncertainty. When we look at the two scores, we see about the same reduction in comparison to *Positive Intent* on both dimensions (own activities and connections). So, potentially subjects did not consider the externalities and social risks of their privacy decision but considered only their own activities and personal risks in their disclosure decision—not wanting to reveal the postings they had made and could remember.

To address this uncertainty, we presented subjects with a second decision that does not mix own-activity and social connections, but focuses *solely* on social privacy risks. Subjects were asked to reveal their friends list to us. We informed participants that we would employ the list to make connection between people whose data we already had collected, even when they do not explicitly interact on Facebook liking each other's content or making comments. Learning about their friends list would allow us to form a richer baseline of, how shared interests, activities or social contexts may affect privacy behavior. The privacy risk associated with disclosing their friends list is typical for the collective risks in a social network: not one but many subjects must permit us to access their data to create a baseline. Only if the data pool is large enough could

[69] Since we perform this aggregation in both treatments we compare, subjects who may have only few posts they strongly care about and others who may have many they care less about, should be distributed evenly across both treatments and not distort our findings.

conclusions about subjects' preferences, political convictions and traits be derived. And if the pool is formed, inferences could also be made about those subjects who do not consent and disclose their data. Thus, the more people give access, the larger are the overall externalities.

Importantly, as we only collect public data subjects who give us access to their friends list, do not disclose any personal information they have not already revealed: their friends list is posted in one spot on their personal Facebook page for anyone easily to access along with the number of their friends and everyone who they are friends with. The upshot is by giving us access, they primarily cause a social privacy risk, and not a personal one. As is characteristic for a public good, their personal risk is not primarily driven by their own actions, but by the choices of others who may also reveal their friends list to us. Subjects have to trade off taking the money for personal benefit versus contributing to the public good; here, the privacy of the collective.

Of course, plausible alternative explanations exist that answer why subjects may refuse to disclose the list of their friends. On one hand, subjects may perceive it as inappropriate to involve their friends in the study without having their consent and in exchange for money. On the other hand, subjects may also reveal the list because they want to support research; they may also think that the study is interesting and they want their friends to have the same experience or be able to earn the 4€. However, all these alternative motivations should, if they exist, be present in both the *Positive Intent* and *Legal Requirement* treatments. Of course, their impact may change with the treatment, for example subjects may be more willing to involve their friends even without having their consent in *Positive Intent* because they trust the counterparty. This is, however, exactly the treatment effect we want to measure.

The first hypothesis, as in Experiments 1 and 2, suggests that subjects in *Legal Requirement* will be less likely to disclose their friends list than the participants in the *Positive Intent* treatment. The data support this hypothesis (see Figure 3). As expected, we find that the rate of disclosure is significantly lower in *Legal Requirement* with 11.4% compared to the *Positive Intent* treatment with 30.3% ($p < 0.01$ t-test), suggesting that subjects in *Positive Intent* rely on the trust heuristic while the use of the trust heuristic in *Legal Requirement* is blocked, making subjects more cautious. Generally, in both treatments fewer subjects disclose their friends list, compared to the first decision whether or not to reveal their Facebook data. The lower rates of disclosures may suggest that subjects are reluctant to involve their friends in the study. However, as expected this hesitation appears to be equally present in both treatments (comparing decisions 1 vs. 2 in *Positive Intent* -20.4%; $p < 0.01$ (Fisher) and *Legal Requirement* -18.8%; $p < 0.01$ (Fisher).

Finally, we want to confirm whether subjects indeed considered in their decision the risk their behavior would impose on the social risk of privacy. The risk corresponds with the amount of data they would feed in the data collection—the longer their list

of friends that would be contacted and potentially participate in the study, the larger the data pool would grow. We expect that subjects in the *Legal Requirement* treatment who decide to pass on their friends list will reveal a relatively smaller number of friends than participants in the *Positive Intent* treatment.

The results reported in Figure 4 support this hypothesis: the 23 participants in *Legal Requirement* who disclose their friends list, reveal an average of 82.03 friends, while, by contrast, the 61 participants in *Positive Intent* who permit us access to their list, disclose with an average of 187.62, which is significantly more contacts ($p = 0.01$ t-test).

These results show that our mild intervention—blocking the use of the trust heuristic—is effective in making subjects considerate of social privacy risks that do not directly threaten the subject's personal privacy interest: the subjects in the *Legal Requirement* treatment appear to consider their privacy decision as a contribution to privacy as a public good and more often refuse to take the personal benefit of disclosure, unlike subjects in the *Positive Intent* treatment, who are significantly less likely to consider social privacy risks in their decision.

While our data suggests a strategy for making privacy decision more considerate of social risks, it also conveys another important message for privacy regulation: to the extent that individuals make privacy decisions that seem to disregard the externalities that their choices impose on others, these decisions for many subjects do not suggest that they are self-interested and unconcerned about the welfare of others. Rather, the cognitive tools that individuals employ to make privacy decisions have a tremendous influence on how their perception of privacy protection is framed. As long as the heuristic indicates that the requesting party can be assumed to be trustworthy, social privacy risks are likely to be less salient: a trustworthy, is assumed will protect the privacy of those revealing their data to her.

Our data suggests that if regulation can disrupt the framework that causes individuals to use the trust heuristic individuals may, be more considerate of privacy externalities and willing to contribute to the public good of privacy.[70] As a consequence, in considering how to generate a regulatory approach that empowers individuals to make more socially-conscious privacy choices our data is fundamental, We will discuss the wider implications of our results for privacy policy in the next section.

* * *

Summarizing the results of our three experiments, we were able to establish the following:

[70] Joshua A.T. Fairfield and Christoph Engel PRIVACY AS A PUBLIC GOOD, Duke Law Journal Vol. 65, No. 3 (December 2015), pp. 385-457.

Experiment 1:

(1) Subjects tend to make privacy choices by relying on a trust heuristic that suggests whether to treat their direct counterparty to whom they would reveal their data to as trustworthy or not.
(2) We can stop subjects from employing the heuristic by blocking the information the heuristic processes: whether the counterparty offers them privacy protection voluntarily.

Experiment 2:

(3) Subjects using the heuristic tend not to consider otherwise credible information—whether positive or negative signals —regarding the safety of a potential disclosure. The heuristic appears to crowd out consideration of such information.
(4) By blocking the heuristic, we can create conditions where subjects become more likely to heed the otherwise credible information that the heuristic crowds out.

Experiment 3:

(5) Subjects using the heuristic tend to frame privacy decisions as bilateral trustworthiness assessments, and therefore tend to ignore the social/public goods aspects of their privacy decisions.
(6) Subjects blocked from using the heuristic are more likely to forgo personal benefits and contribute to preserving privacy as a public goods in a networked setting.

NOTE: THE FINAL SECTION IS STILL IN RELATIVELY ROUGH FORM AND INCOMPLETE

IV. Discussion of Results and Policy Design

A. The Validity of our Results

We will first discuss factors that relate to both the external and internal validity of our results. By "external validity" we mean the extent to which our findings can be extrapolated from the context of our study, and applied to real-world privacy decisions.[71] By "internal validity" we mean the degree of confidence we can have in the cause-and-effect relationship that our studies purport to establish between use of

[71] Michael Findley, Kyosuke Kikuta Michael Denly, . External Validity, Annual Review of Political Science 24(1):365-393, 2021.

the trust heuristic and privacy decision-making.[72] Our studies are robust, we believe, with respect to both external and internal validity.

External Validity. When people are aware that they are observed and subject to research, their behavior can change. This is referred to as experimenter demand effects. Participants try to figure out the objective of the research and what choices might be expected of them. Some may try to help experimenters reach the assumed scientific goal or attempt to conform their behavior to what they perceive is socially desirable.[73] In this study we have used a partial field design that implements several elements of field work to avoid such experimenter demand effects.[74] Our subjects are not aware that a study is under way when we begin to observe their behavior.[75] When they decide how they want to be paid, they assume that the experiment is over and that their choices are not the subject of a study. Their decisions are comparable to a choice of payment they make at a cashier in a supermarket or when they register and pay for a sports program at their university.

Our study also takes advantage of a second element characteristic of field experiments: participants make decisions that directly affect their situation in the real world. Instead of completing an abstract laboratory task with induced payments, they disclose a relevant part of their private data—in the first two experiments their financial contact information and in the third experiment their Facebook data. The direct similarity of the experimental and real-works tasks supports the external validity of the studies.

Finally, our subjects are not assigned experimental roles they are not (yet) trained for or accustomed to. We do not, for example, put students in the role of managers or entrepreneurs to investigate how they respond to particular stimuli. Instead, we ask subjects who hold financial accounts and are familiar with using them to make payment choices and consider revealing their account information, and we ask subjects who use Facebook to consider disclosing their data and friends-list. Subjects are presented with decisions that they have already confronted in their ordinary experience before participating in our study.[76]

---

[72] Schram, Arthur ,The tension between internal and external validity in economic experiments". Journal of Economic Methodology. 12 (2): 225–237, 2005.

[73] Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. Psychological Bulletin, 77(4), 273-295.

[74] See the categories of field experiments by John List & Glen Harrison, Field Experiments, 42 J. OF ECON. LIT. 1009-1055 (2004).

[75] The third experiment does not share this element of field work. Here subjects are aware that the Facebook data they reveal to us will be analyzed and used for research. However, we made sure to avoid demand effects nevertheless: We used a between subject design; that is, subjects received only the instructions for the treatment they participated in. So, they can hardly figure out what manipulation we implemented and what research goal we have.

[76] Glenn W. Harrison and John A. List (Journal of Economic Literature, vol. 42, no. 4, 2004, pp. 1009-1055) develop criteria for field experiments referring to the elements our experiments meet: that subjects are not aware of being studied, that the task subjects are presented with is real and consequential

It is possible that the external validity of our findings might nevertheless be limited by the characteristics of our participants—current and recently graduated university students—but we believe that this population is more likely to strengthen the studies' external validity rather than weaken it. The online behavior of academically trained subjects is likely to differ from individuals without that training. Indeed, our subjects can be expected to have experience in deciding whether to share their data; they are typically active on social media platforms and will have made decisions regarding their privacy in a number of online interactions, including, for many subjects, online banking. If anything, one would expect our study population to be *more* aware of privacy risks and better prepared to protect themselves and others relative to individuals with less experience. That our relatively sophisticated subjects nonetheless appear to be making privacy decisions according to the trust heuristic may not limit but instead reinforce the policy significance of our results, underscoring the importance of regulatory change.[77]

Another possible external validity limitation arises from information uncertainty. Even though we instruct subjects clearly, they may not fully understand (or not pay attention to) what content they are revealing when they agree to disclose banking information, or publicly available data linked to their Facebook accounts. With respect to the Facebook information, for example, a subject may not recognize that the data revealed includes all comments or photographs that others have associated with the subject's profile for the full time-span that the subject has operated a Facebook account. It is possible that had they fully considered these threats to their personal privacy and the privacy of others, subjects may not have chosen to disclose their data. This failure to fully consider risks, however, is not a limitation to our experiment but rather an element of its ecological validity.[78] Such potential underestimations or

("nature of commodity and stakes") and that the task is theirs to solve in the real world ("match of task and subject pool").

[77] Relatedly, because we are academic researchers (and because one of us is associated with subjects' home university), participants may trust us more relative to a for-profit firm which might benefit from analyzing and selling their data. That difference in baseline trust may exaggerate subjects' tendency to make privacy choices by relying on a heuristic based on a single cue regarding trustworthiness while ignoring other information. However, we observe and are only interested in *relative* treatment effects. That is, we compare the change we observe in our subjects' choices when they can rely on the trust heuristic for making privacy decisions versus their choices when we block their use of the heuristic, nudging them to consider the more relevant cues we provided them. Even if subjects trust us more than they trust a profit-oriented firm, and are therefore more likely to disclose their data to us in general, this higher level of trust would affect both treatments equally. Therefore this higher base-level of trust cannot cause subjects to more readily disclose their data in the *Positive Intent* treatment where they can rely on the heuristic compared with *Legal Requirement* where they cannot use the heuristic. We addressed this concern in part in one of the two information conditions of Experiment 2, where we concealed our identity and gave subjects no other information than a Gmail address of our assistant about who they were giving their banking information to. We found that their behavior did not significantly change compared to the full information condition in which the researchers' biography and affiliations were revealed.

[78] Herbert Simon coined the term, which means that heuristics are not per se rational or irrational. Instead, whether they lead to good outcomes. depends on their match to the environment in which the

misconceptions of what data the requester would in fact be accessing, if participants gave permission, are typical for privacy decisions on media platforms and services. Indeed, the failure by potential disclosers to fully inform themselves regarding the extent of requested disclosure, or the implications of that disclosure, is one of the overarching weaknesses of the notice and choice approach to privacy.[79] It is no surprise that people in the real world, as well as in our experiments, may forget what data they have stored or posted when they decide to disclose. And it is also no surprise that both people in the real world and subjects in our experiments may fail to fully consider the ramifications that disclosure of particular data may have for them and for others.

Internal Validity. With respect to our study's internal validity, we account in our study's design for the possibility that the legal policy requiring the provision of anonymous payment, which we informed subjects about in *Legal Requirement*, may have suggested to subjects they should be cautious in disclosing their data. We addressed this potential confound in Experiment 1 in the *Expressive Signal* treatment, where we informed participants that the university had enacted the data policy, but that it was not yet effective at the time the study was conducted and that we would nevertheless voluntarily offer participants the same protection. In contrast to the *Legal Requirement* treatment, the manipulation allows subjects to use the trust heuristic. Our results did not reveal signs of an expressive effect of the enacted legal provision; subjects were significantly more likely in *Expressive Signal* than they were in *Legal Requirement* to permit us access to their data—which shows that subjects in *Expressive Signal* and *Legal Requirement* are not responding to any normative message that the law is sending but rather that subjects in *Expressive Signal* are using the heuristic, while the participants in *Legal Requirement* are blocked from using it.[80] Since the results in *Expressive Signal* were clearly significant we did not repeat this treatment in Experiments 2 and 3.

It is also possible that some participants in our study had little data to disclose, and therefore had little at stake in their privacy decisions. For example, some subjects might be willing to disclose their data and earn the €4.00 because they use Facebook only rarely and therefore are aware that they have little data to reveal. Similarly, with respect to the choice of payment methods in our experiments, some subjects may have no bank account or PayPal address and may therefore choose to pick up their earnings anonymously in cash even if they would have preferred to receive the money conveniently by providing name and bank details or their PayPal address.

---

heuristic used. Hertwig, R., Leuker, C., Pachur, T., Spiliopoulos, L. and Pleskac, T.J. (2022), Studies in Ecological Rationality. Top. Cogn. Sci., 14: 467-491.

[79] Yannis Bakos, Florencia Marotta-Wurgler and David R. Trossen. Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts. The Journal of Legal Studies, Vol. 43, No. 1 (January 2014), pp. 1-3.

[80] CROSS REFERENCE to discussion above of this point.

While this may be true for some of our participants, it does not challenge the internal validity of our experiments. We assign our subjects randomly to the treatments we test. Therefore, we should see in each treatment approximately the same number of subjects who have no bank account or PayPal address, or subjects who seldom use Facebook or have stopped using it. Because these subjects are randomly distributed among treatments, they will not affect the effects of the treatments on privacy behavior we measure.

B. **Designing a Behaviorally-Aware Privacy Policy**

1. The Behavioral Challenges of Privacy Decision-Making

Note, finally, that our results do not contradict other psychological mechanisms that have been reported to contribute to people revealing data more readily than is consistent with their stated privacy preferences. Indeed, our results may add to the understanding of the relevance of these other mechanisms and may also intertwine with some of them in driving privacy decisions in the real world. We will briefly review these other potential mechanisms and their interaction with the trust heuristic.

Cognitive Errors. Some authors have shown that privacy decisions suffer from a form of *cognitive error*; specifically, that individuals considering disclosing data often confuse the control they have over the disclosure of their information with the control they have over how that information is used *after* it has been disclosed and how that use may affect their privacy.[81] The cognitive error is likely interacting with the use of the trust heuristic both mechanisms complementing each other, because individuals employing the heuristic focus on the intent of the person to whom they are considering disclosing information, they may be more vulnerable to cognitive error—that is, more likely to focus on the first form of control, the decision whether to disclose information (as opposed to what might happen with the information once disclosed), because it is the task that the heuristic processes. Once the heuristic is applied and the counterparty classed as trustworthy, the trust judgment may tend to supplant any concern with control over what happens with personal data after disclosure, as is suggested by the findings in our Experiment 3, where subjects who are able to employ the heuristic are less sensitive than subjects who are blocked from employing it to risks that their disclosure decision poses both to their own privacy and to the privacy of others.

Other authors have suggested that people may not acquire information about means to protect their privacy even if the acquisition is costless, because they want to avoid privacy questions entirely.[82] Our studies do not conflict with this finding; rather, they provide a possible explanation for it: When individuals rely on heuristic decision-

---

[81] Laura Brandimarte, Alessandro Acqusti and George Loewenstein Social Psychological and Personality Science 4(3) 340-347.

[82] Svirsky, Dan (2021) "Why Do People Avoid Information About Privacy?," Journal of Law & Innovation: Vol. 2 : Iss. 1 , Article 2.

making they tend to ignore information the heuristic does not process, so they will not acquire information that otherwise might seem relevant to their decision even if acquiring it is costless. However, they are not categorically averse to acquiring information: our study (particularly Experiment 2) demonstrates that if the trust heuristic is blocked, individuals are willing to apprehend and consider more information relevant both to their own privacy risks and the risks faced by others.

Another cognitive bias that appears to affect privacy decision-making is "over-choice."[83] Users can be discouraged by the number of privacy choices on offer: for example, decisions about managing cookies (up to 200 choices for a typical smartphone user[84]), about geo-tracking, about behavioral tracing, and many other choices relevant to privacy. Again, our findings do not conflict with the over-choice theory, instead the data may support the external validity our results and may help explain why so many people in our study apply the trust heuristic when making privacy decisions. Privacy over-choice may cause a feeling of being overwhelmed, which may in turn contribute to individuals' readiness to apply the heuristic— by doing so they economize on the cost of making so many decisions. In Part IV.B, below, we discuss a strategy of re-framing privacy decisions around an ex ante, standardized "Master Privacy Template" that, we believe, will discourage use of the heuristic, in part by reducing the pressure of privacy decision over-choice.

Hyperbolic Discounting. Disclosure often carries with it certain immediate benefits—convenience, access, or social engagement, to name just a few. But the risks of disclosure are usually only felt much later. As such, our tendency to overvalue current rewards while inadequately discounting the cost of future risks makes us more willing to share now. For example, Jentzsch et al. [21] found that people preferred barely less expensive movie tickets even though the cheaper ticket required more extensive personal information. Yet, consumer choices changed when tickets were offered at the same price—the privacy protective movie company won more customers. The authors concluded that consumers were heavily discounting the risks associated with disclosing personal information, even far below small differences in price. Other studies have shown that consumers make disclosure decisions without fully appreciating time inconsistent preferences. Wang et al. [45] found that users of social networks may gain some immediate pleasure from posting a salacious selfie, but often end up regretting it later and wish they had never posted the picture in the first place.

Dark Patterns. Many privacy scholars have also noted that online actors requesting disclosure frame and present their requests in ways designed to encourage disclosure, often without the disclosing party understanding that her decision has been manipulated. Again, our findings may help explain how these framing strategies are

[83] Neil Richards, Woodrow Hartzog The Pathologies of Digital Consent, 96 Wash. U. L. Rev. 1461 (2018-2019).

[84] Olmstead K, Atkinson M: Apps Permissions in the Google Play Store. Pew Research Center; 2015.

processed by individuals making privacy decisions. For example, firms may frame their privacy policy as "strict" (i.e., protective of the user's personal data) in comparison to competitors.[85] This framing directly feeds the trust heuristic by making salient that the platform could have opted for lower protection, but unlike competitors, did not. And then there is framing "by design," including the use of strategies sometimes referred to as "dark patterns."[86] An example is Facebook, which offers myriads of privacy options that allow users to manage in great detail and for every piece of information the user posts the extent of disclosure in relation to their contacts and other users within the network. The elaborated structure of these privacy options directly feeds the heuristic: it helps to construct a perceived action space in which Facebook seems to care about its users' privacy, and the user experiences this seemingly beneficent disposition anew with every posting she makes on the platform. By contrast, the privacy policy that governs the *relationship between Facebook and its users* is presented only once when users sign up and consent to Facebook's very broad authority over the data to be created by the user, and by others interacting with the user's account, in the many years of usage to come.[87] [MORE TO COME]

Discarding Information. Relatedly, the use of the trust heuristic may help explain why consumers do not appear to use the information about privacy protection that parties seeking disclosure are obliged by the notice and choice doctrine to provide. Our second study shows that individuals rely so strongly on the trust heuristic's judgments, that many parties considering disclosure discard relevant information.[88] And that result has an important knock-on effect: If parties considering disclosure do not respond to otherwise valid signals of credibility that the heuristic does not process (as is the case with our subjects in Experiment 2), then parties seeking disclosure lack an incentive to provide detailed and credible information about safety technologies or even to invest in developing and providing such technologies.

Most likely a complex bundle of entwined psychological mechanisms causes the gap between privacy preferences and privacy behavior that we have described above.[89]

---

[85] Adjerid I, Acquisti A, Brandimarte L, Loewenstein G: Sleights of privacy: framing, disclosures, and the limits of transparency. Proceedings of the Symposium on Usable Privacy and Security (SOUPS'13) ACM 2013:1-11.

[86] Harry Brignull who has coined the term, defines dark patterns as the deceptive design of a user interface that is supposed to lure users into choices they do not intend to make (Harry Brignull Dark Patterns. 2018. Available online: https://darkpatterns.org/). An example is "Privacy Zuckering" – named after Mark Zuckerberg –that aim at making users share more information than they wish to for example by purposefully complicating privacy notices or by making it difficult to reject data sharing. California has regulated these practices in the California Consumer Privacy Act.

[87] cite

[88] CITE Florencia and others.

[89] Hann, Il-Horn, Kai-Lung Hui, Sang-Yong Tom Lee, and Ivan P. L. Png. 2007.Overcoming Information Privacy Concerns: An Information Processing Theory Approach. Journal of Management Information Systems 24:13–42; Hui, Kai-Lung, H-H. Teo, and Sang-Yong Lee. 2007. The Value of Privacy Assurance: An Exploratory Field Experiment. MIS Quarterly 31:19–33. Still, that subjects are willing to accept a low payment of 4€ to disclose their Facebook data may seem surprising. But as we

Many of these other effects, as we have detailed, may encourage use of the trust heuristic, or even may be caused by it. As a consequence, it is possible that many of these effects will abate if the heuristic's use is blocked by the regulatory changes we are suggesting in the next section—changes which, together, are meant to empower individuals to make better, more reflective privacy decisions themselves.

2. Instruments for a Behaviorally Aware Privacy Policy

Our findings, and the insights we can derive from them, have significant implications for the design of privacy law and policy. First, our findings suggest that the gap between people's reported preference for privacy protections and their actual behavior may not be the result of individuals' loose or unstable privacy preferences, but rather the particular cognitive processes people use when engaging in privacy decisions. This is a significant finding, one which should shape our thinking about how to respond to the failures of notice and choice privacy regulation. If the variance in privacy behavior were due to unstable preferences, any policy intervention with a hope of success would have to be paternalistic: that is, privacy regulation would have to be designed largely to take privacy choices away from consumers by setting out a series of mandates defining what data may be elicited in different contexts, and the uses to which that data may be put. That type of regulation deprives people of choice. The design of privacy mandates not only requires a great deal of knowledge about the risks and benefits of disclosure across a wide range of different scenarios.[90] While our findings are not (and should not be taken to be) anything like a cure-all, they do suggest that to the extent that individuals' privacy choices are driven by use of the trust heuristic, that use can be blocked, and privacy decision-making improved. That is, behaviorally-informed regulation of the decision-making context can help consumers make better-grounded choices and as we will argue, even within the notice and choice framework.

Second, our results suggest that behaviorally-informed privacy regulation is needed, that responds to how individuals actually make privacy decisions.

Blocking the heuristic. One particular regulatory goal is obvious in light of our findings: Policies addressing the trust heuristic should start with a rule requiring that when law obliges an actor to provide substantive privacy protection, that actor *must also disclose that the protection is required by law*. Our results suggest that obliging parties

---

have seen above, multiple studies report similarly small valuations for sensitive data. Jentzsch N, Preibusch S, Harasser A: Study on Monetising Privacy: An Economic Model for Pricing Personal Information. European Union Agency for Network and Inf. Sec. (ENISA); 2012. Hyperbolic discounting may explain these particularly low prices as users receive the gratification for consenting immediately while potential privacy risks lie further in the future. Note that some disclosure decisions in our study appear also to be rationally motivated. Some of the participants indicate that they have never actively used their Facebook accounts. They let us access their profile without in fact revealing any personal data.

[90] Cite risks for privacy mandates

requesting private information to disclose that the privacy protection they provide is legally mandated may help to prevent consumers from classifying the requesting party as trustworthy solely on the grounds of the heuristic. Moreover, in our experimental setting the disclosure of the legal obligation restores the consumer's responsiveness to more credible cues; in particular it makes them considerate of the social risks of privacy. As a consequence, if disclosure of the source of the protection functions in the real world anything like it does in our experiments, it should block a substantial share of individuals from employing the trust heuristic, increase the likelihood of a group of consumers to respond to otherwise credible positive or negative cues about the risk of disclosure, and, in turn, increase the pressure on companies to provide credible cues in order to convince customers of their trustworthiness.

A rule requiring disclosure of the source of privacy protections when a protection is legally mandated will be especially important if, in response to the failures of the notice and choice paradigm, privacy regulation moves toward requiring the provision of certain protections, but still permitting consumers to waive those protections. Wider use of mandates without requiring disclosure that the law is responsible for the protections is likely only to feed the heuristic, by encouraging potential disclosers to frame the provision of particular protections as the requesting party's act of beneficence, rather than compliance with the law. The result might be that mandated privacy protections intended to reduce disclosure of sensitive information might end up increasing it as consumers waive protections in response to the input they get from a heuristic that tells them to trust the requesting party whose beneficence they perceive—incorrectly—to be the source of the protection they are "offered."

That said, for the moment, the U.S. remains firmly enmeshed in the notice and choice paradigm, and so understanding the immediate policy implications of our work requires us to consider how the heuristic operates, and how it can be blunted, within the framework of notice and choice. Of course, few people read privacy policies to inform themselves about the details of the disclosure that is requested from them. Given this propensity to ignore available information, which the heuristic reinforces, people must be presented very saliently with the information about the source of privacy protections—otherwise individuals can be expected to continue to employ the trust heuristic in their privacy decisions. This salient presentation is of course difficult as long as the presentation of privacy information is left in the control of the self-interested parties who seek disclosure. They will always have a strong motivation to frame the presentation of that information such that consumers are induced to use the heuristic, and to disclose at a rate higher than if they did not.

As a consequence, a policy to block use of the trust heuristic must rely on something more fundamental than disclosure. One way to go deeper would be to interfere with the tendency of potential disclosers to frame their decision as occurring within a bilateral interaction where the requesting party is presenting privacy choices

and the potential discloser is responding according to how she perceives the offer relative to her perceptions of the requesting party's action space. It is this characteristic—the imagined similarity of online privacy decisions to personal, face-to-face interactions between individuals—that drives individuals to apply the heuristic in the first place.

Imposing a Master Privacy Template. We propose a way to extract privacy decisions from consumers' tendency toward bilateral framing by mandating through law the use of an *ex-ante and standardized* machine-readable "master" privacy template. This could be accomplished on a webpage run by the federal government—perhaps it could be "privacy.gov" which would provide the tools for private use and the interface for any company platform or website that seeks to extract data. [MORE TO COME]

A master privacy template would define and present a standardized menu of the categories of data that may collected, as well as the purposes for which the collected data may be used. The user would select the categories of data she is prepared to disclose, and the purposes for which she is willing to have them used. Once those choices are made, any party seeking disclosure would be obliged to communicate—automatically, according to standardized protocols—with the database containing the individual's previously-expressed preferences. The data is disclosed without further inquiry only if there is a match—that is, the disclosure-seeking party does not seek disclosure of data the user is unwilling to reveal, or put that data to uses that the user has not approved. If there is no match the user must be presented each time with any request tailored to the type of data, or the type of use, that the user has not previously approved. Disclosure is then permitted only if the user indicates specific agreement.

Centralized Control and Enforcement. An immediate question is who would be responsible for creating and administering the master privacy template. Responsibility for design could be placed with an agency such as the Federal Trade Commission, currently the chief federal agency on privacy policy and enforcement.[91] The design will depend on experts surveying and collecting into a manageable number of categories the types of data that requesters seek to collect, and the uses that they wish to make of those different types of data.[92] There is an opportunity, in the design of the selections in the template, to discourage types of collection and use that the agency believes pose especially high risks—for example, by establishing defaults that disallow those types of collection and use, and requiring the potential discloser to alter those

[91] See Federal Trade Commission, Protecting Consumer Privacy and Security, https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security.

he template may also include different types of data requesters and categories of retention policies. About the tradeoff between a detailed taxonomy and an overhead that causes errors in implementation and usage see Ben-shahar O, Chilton A. Simplification of Privacy Disclosures: An experimental Test. J Legal Stud. 2016;45: 1–17; Lorrie F. Cranor & Joel R. Reidenberg, Can User Agents Accurately Represent Privacy Notices?, 30 TELECOMM. POLICY RESEARCH CONFERENCE.

defaults in order to permit the particular collection and/or use. Another possibility following the principles of privacy by design to switch the default on data disclosure and tracking entirely, that is if the user does not take action no data will be shared o even sold.[93] Rather than individuals having to opt out permissions to track and share their data, they would have to opt in. Switching the default would apparently make users much more cautious in revealing their data.

Enforcement. [TO COME]

Ex ante Design. The key element of this regulatory approach is that consumers would be required to access and complete the online master privacy template *before* giving consent to any release of personal information to a particular requester. That is, the requester would be legally barred from soliciting the disclosure of data unless it has obtained consent through a standardized template that the potential discloser has already used to establish his or her privacy preferences.

This ex ante structure is vital. Under current arrangements, privacy decisions are made piecemeal for each counterparty requesting disclosure, and without a baseline that the potential discloser has established in advance setting out the types of data, and the types of uses, that the potential discloser is willing to assent to. That construct feeds the heuristic, because it leaves to the party seeking disclosure the design of the consent form as well as the provision of information specifying the data sought, the uses to which it will be put, and the protections against misuse offered or required. But also as we have seen about it feeds also hyperbolic discounting as users are pressured to invest a lot of time in their privacy protection, the illusion of control error is enforced by the many choices they have to make, and so on…

*Standardization and Informational Nudges***.** It is also important that the terms of the master privacy template are standardized and for all parties seeking disclosure.[94] Under current arrangements, each party seeking disclosure can design their own policies and consent form within the limited constraints of existing privacy law. To make an informed privacy decision, users must review each one of them, which of course very few users do. And because website owners have broad authority over terms and how those terms are presented, they can overwhelm users, present them with too many choices, manipulate defaults (such as by highlighting buttons for acceptance or requiring "opt out" choices), or complicate opt-out. Some of these strategies, as we have described, appear to be useful in exploiting, and perhaps even intended to exploit, the trust heuristic. Piecemeal requests also force the consumer to make a new decision each time data is requested, which is time consuming and creates the perception of over-choice that encourages the use of the heuristic in decision-making: indeed,

[93] Willis, Lauren E. "Why Not Privacy by Default?" Berkeley Technology Law Journal, vol. 29, no. 1, 2014, pp. 61–133.

[94] Lorrie Faith Cranor Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice,J. Telecommun. High Technol. Law 2012,10, 273-308.;

heuristic decision-making in general is often a response to a perceived need to economize on the cost of decision-making, in part by speeding it up. But if the privacy decision is, at least initially, removed from a particular transaction and made in advance without the involvement of a particular counterparty, that structure would be far less congenial to the heuristic's use, because there is no counterparty and no imagined action space that can be processed to assess trustworthiness in the way that the heuristic performs that task.

Deviations from the Privacy Template. The consumer's default—her pre-existing preferences as expressed in her selections in the master privacy template—are likely to be sticky.[95] this stickiness can be exploited for designing a process for additional disclosure requests that would require any requester who wishes to ask the consumer to disclose data or permit uses in excess of what the consumer has approved on the master privacy template to do so through the master privacy template itself. The requester would be required to transmit requests for additional disclosure to the website hosting and administering the template, and to channel the consumer's response to those additional requests through the template as well—ideally, without naming the requester. That interference with bilateral framing of the transaction would again limit the ability of the requester to take advantage of the trust heuristic, also the dark patterns and hyperbolic discounting.[96] In addition, when a decision is re-framed so that the consumer's ex ante selections from a standardized template becomes a strong default, any website that asks for additional disclosures or permitted uses will be imposing transaction costs on the consumer, and will have to consider the value of the additional disclosure or use against the risk that consumers may refuse or even prefer to deal with a competitor (at least where the service has competitors).

Costs and Benefits of Data Provision. Finally, we should make clear that we are not advancing a suggestion regarding an optimal level for either privacy protection or the disclosure of confidential information. Costs and benefits of disclosure (that is, on the total social benefits of disclosure versus the risks of disclosure) are likely to vary substantially across different settings, making general assertions about an optimal balance suspect.

Further, we would caution that in privacy policy discussions people may tend to focus on the risks of disclosure, but even with respect to information that, if disclosed,

---

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Chapter 11, p. 177, Yale University Press

[96] Error resiliency privacy nudges can assist consumers, as decisions on privacy often favor risky and not thought through decisions, without taking possible long-term consequences into account. This is based on so-called hyperbolic discounting, in which the immediate benefit is overestimated, and costs incurred later are underestimated by individuals. To counteract this, a time delay can be used as a privacy nudge. In this way, the individual should be persuaded to act less impulsively and to rethink the message and possible negative consequences.

presents substantial individual and social risk, the benefits of disclosure must also be understood and considered. For example, the disclosure of information about a person's finances and creditworthiness can produce substantial individual and social risk, but can also produce substantial benefits: –disclosure is necessary to produce credit ratings and scores, and these services are themselves public goods that foster social trust and cooperation.[97] Health data is also characterizes for the two sides of data sharing: While it can have life changing impact from job search to health insurance, when it is not protected, disclosure of health records also enable research on many life-threatening diseases that often require to observe patients for decades like in studies of strokes, heart attacks or other organ damages.[98]

Although the optimal levels of privacy protection and of disclosure may vary, there is no reason to believe use of the trust heuristic will conform individual privacy decisions to the balance of benefits and costs across settings. Changing the framework from bilateral to social and blocking the heuristic should make the potential discloser more cautious. Although we do not test this directly, we expect that once freed from the heuristic, people will be better positioned to consider the technological aspects of the disclosure—e.g., how widely can the information be disseminated once disclosed, and how likely is further dissemination (how secure is the platform, has it been breached in the past, how likely is a future breach?). [MORE TO COME]

V. Conclusions.

The policy intervention we suggest leaves the actual privacy decision with the individuals considering disclosure. Optional privacy protection is preferable in general to the mandate of fixed procedures, as it respects individuals' actual privacy preferences and allows them to make a personal cost-benefit analysis by weighing the privacy risk against the costs more rigorous privacy procedures may impose on them. Importantly, however, a privacy policy based in personal choice is only useful when customers are at least reasonably likely to consume the information they need to reliably assess the objective privacy risk they face.

The intervention is likely to improve the situation for at least two reasons. First, because it removes the initial privacy decision from the bilateral framework which the heuristic makes decision-makers envision their privacy-relationships. And second because it turns the default around for any request for disclosure beyond what the consumer has approved via her choices in the master privacy template. Under current law, if users don't read a specific requestor's privacy terms and consent to disclosure—

---

[97] Simeon Djankova , Caralee McLiesh a , Andrei Shleifer Private credit in 129 countries Journal of Financial Economics 84 (2007) 299–329

[98] Heidi Beate Bentzen, Rosa Castro, Robin Fears, George Griffin, Volker ter Meulen & Giske Ursin Remove obstacles to sharing health data with researchers outside of the European Union Nature Medicine volume 27, pages 1329–1333 (2021)

in many instances likely employing the heuristic—they may reveal everything the requestor has asked for simply by using the website, typically without knowing what they are disclosing or the purposes for which it is disclosed. On the other hand, once a consumer has made her selections in the master privacy template, unless the provider of the service makes requests that do not exceed the types of data and the uses of that data that the consumer has previously approved, the disclosure is rejected.