Measuring Lay Reactions to Personal Data Markets

Aileen Nielsen
aileen.nielsen@gess.ethz.ch
ETH Zurich, Center for Law and Economics
Zurich, Switzerland

ABSTRACT

The recording, aggregation, and exchange of personal data is necessary to the development of socially-relevant machine learning applications. However, anecdotal and survey evidence show that ordinary people feel discontent and even anger regarding data collection practices that are currently typical and legal. This suggests that personal data markets in their current form do not adhere to the norms applied by ordinary people. The present study experimentally probes whether market transactions in a typical online scenario are accepted when evaluated by lay people. The results show that a high percentage of study participants refused to participate in a data pricing exercise, even in a commercial context where market rules would typically be expected to apply. For those participants who did price the data, the median price was an order of magnitude higher than the market price. These results call into question the notice and consent market paradigm that is used by technology firms and government regulators when evaluating data flows. The results also point to a conceptual mismatch between cultural and legal expectations regarding the use of personal data.

CCS CONCEPTS
- Security and privacy → Economics of security and privacy;
  Social aspects of security and privacy: Information accountability and usage control;
- Human-centered computing → Laboratory experiments.

KEYWORDS
privacy law, personal data markets, experimental law and economics

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies be not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AIES ‘21, May 19–21, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8473-5/21/05...$15.00
https://doi.org/10.1145/3461702.3462582

1 INTRODUCTION

The collection and sale of personal data is an important and common activity, both in the technology sector and also in more traditional domains. But, the digital collection and sale of personal data emerged as a market activity through technological fiat rather than through cultural evolution or consensus-building [27]. Thus technology firms have evolved market infrastructures for personal data, but cultural norms may not have incorporated or accepted such market-like categories and logic.

Rather, concomitant with the emergence of markets for personal data has been a sense of apathy or defeat by users of common digital technologies [25]. Consumers have developed a sense that resistance to personal data collection and sales is futile, and they have good reason to feel this way. “Do not track” and other privacy requests by consumers have been routinely disregarded [10], [9], [27]. Tech companies and other entities who collect data and build machine learning models have likewise routinely violated their own privacy policies [6]. Perceived and actual transgressions of privacy norms have also occurred in the case of data collection for academic research [4], [18]. The resulting widespread discontent has led to consistent calls for legal reform [14], as much behavior found offensive by ordinary people is nonetheless entirely legal.2

Yet, U.S. federal lawmaking has been dormant with respect to the proliferating sales of personal data and the routine commercial surveillance that characterize current lived digital experiences. In particular, U.S. privacy law remains grounded in a market metaphor accepted from the technical fait accompli of technology firms, without any indication that data subjects - that is ordinary people - agree with the use of commercial markets and market transactions to distribute personal data.3

The presumption of market appropriateness implicit in laws that facilitate personal data markets is strongly undermined by the privacy attitudes of lay people. Anecdotal data and recent experimental results alike (described in the next section) suggest that monetary indicators and market trading for privacy do not provide reliable indicators of the value of privacy to ordinary people. The present study extends this recent line of work by exploring the question of whether lay responses to market transactions for personal data are consistent with expected market behaviors. The study explores the possibility that the existence and legality of personal data markets may not coincide with acceptance by lay people of market transactions or market-like behaviors.

1 [27] has depicted the emergence of personal data markets and other elements of surveillance capitalism as akin to the European colonial invasions of the Americas, in which distinct cultural entities, Native Americans versus Europeans, had drastically different conceptions of what their exchanges meant as well as widely disparate levels of technical capacity to shape reality to their desired outcomes.

2 Note that not all behaviors described in this paragraph are legal under current U.S. law. The main point, however, is that even the legal behaviors anger ordinary people.

3 Nations in the European Union are considered to be more protective of personal data privacy, but their data protection laws also recognize and facilitate personal data markets premised on notice and consent. Thus the results of this paper apply more widely than the U.S., but an international analysis of privacy law is beyond the scope of this work.
This experimental study presents evidence that data privacy is mismatched with market mechanisms, and that market allocation of personal data fails to respect the cultural and contextual expectations of ordinary people [16]. Specifically, this study probes the degree to which ordinary people are willing to participate in a market-like exercise of pricing personal data when they have the option not to do so. The results show that many ordinary people reject market-like rules for the allocation of personal data despite this being the dominant paradigm in privacy law and commercial practice.

2 RELATED WORK

It has long been recognized that stated and revealed preference are at odds when it comes to valuing privacy, a phenomenon known as the "privacy paradox". Specifically, people state that privacy is very important to them but often fail to behave accordingly either in laboratory or field experiments [3]. Such results have been obtained in many and varied circumstances, including when privacy is made highly salient.

The privacy paradox did not emerge with the most recent advances in digital technologies but has been documented for decades [1]. Nonetheless it remains important to legal and ethical considerations of socially relevant AI technologies, many of which rely on the availability of personal data, that is, data about people.

The question of how to value privacy, and whether markets are an appropriate mechanism to do so, has a long intellectual history. [11] questioned the assumption that a market-like mechanism populated by rational actors would result in desirable social outcomes. More recently, [22] has shown that the ability to make rational data control or sharing decisions is important to welfare outcomes in privacy markets. Thus the question of whether people can behave as rational economic agents with respect to their privacy is crucial to the question of whether markets are an appropriate venue for privacy. Yet, as [12] noted in a working paper, existing theoretical or behavioral scholarship on privacy leaves undiscussed the existence and implications of rational privacy preferences or choices.

[12] were the first to address explicitly the existence of rational privacy preferences. They did so with an experiment in which personal data points (body fat percentage and IQ) were measured. Research subjects were then presented with a variety of privacy options in which they traded off which of the two collected kinds of information might be shared to various audience sizes. [12] thus developed multiple data privacy trade-off points for an individual and so could assess whether there was a consistent pattern of tradeoff decisions.

[12] found that a majority (63 %) of participants did evince consistent privacy preferences when trading off various privacy scenarios against each other. But, a substantial minority (37 %) deviated from rational orderings. Use of monetary metrics for privacy trade-offs decreased the rate of logically-consistent choices as compared to when participants made direct trade-offs of privacy options.

3The reduced adherence to rational orderings with the use of monetary metrics is consistent with findings from the taboo trade-offs literature [23]. The taboo trade-offs literature shows that individuals demonstrate exaggerated bounded rationality effects when making culturally taboo trade-offs.

4Analogously, it is well known that people, even experts, are rather poor at Bayesian thinking when not working through the formulas formally. Yet, this does not negate the fact that many statistics students are able to come to correct results with Bayesian methodologies on statistics exams. Thus it is a matter of what people tend to do versus what they can do given strong incentives.
The study had a 2 x 3 between-subjects factorial design, varying the relational framing of the data as represented by the website where the data was collected (a college alumni messaging platform or a messaging platform to buy and sell auto parts) and the form of information provided about data market price (no information, explicit market information, or veiled information).

The relational framing of the data was varied because past results indicate that a market framing should make a subsequent sale of an object acceptable, as indicated by all participants agreeing to provide a price.

The pricing information was varied for two reasons. First, the no information and explicit information treatments provided a way to control for background belief or assumption by participants regarding the likely market price. Second, the veiled information treatment provided a way to gauge participant interest in knowing the price by having the option to know or not to know the price.

The first four questions mirrored those [15] used to create a composite distress measure, with four Likert-scale judgments about the acceptability of the transaction. In this way, participants’ overall affective reaction to the transaction was measured.

Next, two questions were presented regarding the data sale in quantitative market terms. First, participants were asked how much John should charge per data point. Second, participants were asked what percentage of the proceeds John should share with the website users.

5 PROCEDURE

Participants were randomly assigned to one of the six conditions reflecting the factorial design and read the appropriate vignette. Participants then responded to four Likert scale questions (on a scale of 7) regarding the judged appropriateness of John selling the data. Participants answered a question regarding how much John should ask per data point if he sells the data. Potential responses were provided in multiple choice format and ranged in successive orders of magnitude from $0.001 to $1.00 per data point.9 Participants also had the option to refuse to provide a price and indicate that the sale should not take place.

The pricing treatment was applied in the pricing question. There were three pricing treatment options. Coupled with the pricing question was either no information about the typical market sale, explicit information (“Such information usually sells for around $0.01 per user, that is, about a cent per user.”), or the option to click a toggle to show the information about typical market price (revealing the same text as in the explicit information treatment). The pricing treatment applied only after the distress indicators regarding the transaction were already collected, and so analysis for pricing includes pricing treatments but analysis for distress does not.

Following the pricing question, participants answered a question regarding what portion of the proceeds John should pass back to

---

7[15] used four kinds of relational framing drawn from [8], one of which was a market framing. Here the framing is simplified to a market framing and a non-market framing, with the latter corresponding to equality matching in [15].

8The full vignette and experimental screen flow is available in the supplementary materials, available at https://osf.io/de5hc/.

9The market price of $0.01 that was chosen for the data on a per-data-point basis was set based on media coverage of market value for similar data, likely representing an upper bound rather than actual price for such data. The decision to round up (as compared to published market prices of $0.006 or $0.007) was taken to simplify the numerical data for participants and avoid confusion regarding the number of zeros in the price.
his customers, with multiple choice options ranging from 0% up to 95%. As with the pricing question, participants could also refuse to choose a percentage and instead indicate that a sale should not take place.

Finally, participants answered two comprehension checks regarding the manipulations as well as a series of demographic questions. Demographic data are reported in the online supplementary materials at https://osf.io/de5hc/.

Following preregistration of the experiment (https://osf.io/de5hc/), data was collected in January 2021 on a sample of 346 U.S. adults via the Prolific online platform. Of those participants, 295 passed the two comprehension checks and were included in the data analysis. All data selection and analyses were pre-registered unless identified as post hoc.

6 RESULTS

6.1 Distress regarding the transaction

A distress indicator was calculated as the arithmetic mean of the four Likert scale outrage indicators. There was no difference between the relational framings as measured by the mean distress about the proposed data sale.

These results were consistent with a pre-registered equivalence hypothesis with a large effect size (Cohen’s $d = 1.0$, $t(293) = -7.622$, $p < .0001$, 90% confidence interval (-0.22, 0.73)). The results suggest that distress regarding the proposed data sale does not result from the relational framing of the data but rather from the nature of the data itself.

6.2 Interest in seeing the price

More than 85% of participants who were offered the option to see the market price information opted to do so (the veiled treatment condition). Yet a large portion (40%) of the participants who opted to see the price went on to refuse to provide a price for the data. These results suggest that refusal to price personal data need not be related to ignorance of or disinterest in the market price.

6.3 Refusal to price

More than 40% of participants refused to price regardless of treatment condition ($p < .01$). These results diverge significantly from the expected pattern for a typical market good, for which the refusal-to-price rate would presumptively approach 0%.

These results were consistent with a pre-registered equivalence hypothesis with a large effect size ($Cohen's \ d = 1.0$, $t(293) = -7.622$, $p < .0001$, 90% confidence interval (-0.22, 0.73)). The results suggest that distress regarding the proposed data sale does not result from the relational framing of the data but rather from the nature of the data itself.

Figure 2: Percent refusing to provide a price. Error bars are +/- two standard errors.

While Figure 2 suggests a possible difference in the refusal to price rates according to relational framing, this is not borne out by post hoc statistical testing. A larger sample size might render detectable some effect of relational framing. Importantly, however, any such effect is clearly smaller than the effect of personal data itself as a good that participants chose not to price at high rates. Also, the results in Figure 2 suggest that any such effect would go in the expected direction that a market framing would make the sale acceptable to a larger percentage of people as compared to the non-market relational framing.

---

10 Full summary statistics reported in the supplementary materials at https://osf.io/de5hc/.
11 For example, in [15]'s results, there was a 0% refusal to price in the market framing condition.
12 There is not a statistically significant difference between the relational framing for either the no information condition ($W = 1130$, $p = .56$) or for the explicit information condition ($W = 991$, $p = .08$).
6.4 Distribution of provided prices

Of the prices provided, the median price was significantly above the market price ($0.01) in three of the four examined cases ($p < .05^{13}$), as shown in Figure 3.^{14}

![Figure 3: Distribution of prices, including refusal to price because data should not be sold.](image)

The full distribution of pricing is plotted in Figure 3. There may be differences in distribution between treatments, as influenced by price information and relational framing, but the current sample size combined with the high rate of refusal to price suggests that applying post hoc testing on the distribution is unlikely to detect any possible effect. This is left to future work with a larger sample.

6.5 Distribution of recommended revenue-sharing percentages

After selecting a price for the data, participants were given the option to recommend the percentage of the proceeds the website owner should share with the data subjects. This scenario might have been expected to make the data sale more acceptable, but the rate of refusing to price was the same as before participants were made aware of the option to share proceeds. A post hoc analysis indicated that the rate of refusing the sale even with the possibility of sharing was statistically equivalent to the rate of refusing the sale when no such possibility was mentioned (TOST analysis, Cohen’s $d = 1.0$, $p < .001$ for all cases^{15}).

The distribution of recommended sharing percentages is shown in Figure 4. Of those who chose a sharing percentage, very few chose 0% sharing, which indicates widespread support for the notion of some redistribution of the economic value of personal data. However, as with the distribution of prices, no statistical analyses were undertaken due to the sample size.

13Full summary statistics reported in the supplementary materials at https://osf.io/de5hc/.
14Again, these results are distinct from [15]’s results, in which prices were at or slightly above market price for all treatments.
15Full summary statistics reported in the supplementary materials at https://osf.io/de5hc/.

![Figure 4: Distribution of sharing percentages, including refusal to select an option because data should not be sold.](image)

7 DISCUSSION

If lay people support the current legal and industry allocation of personal data via sales in commercial markets, then judged distress and the rate of refusals to price in a typical and legal data sale should be low in all cases. This is not reflected in the results.

An alternative theory is that perhaps lay people do not agree that all legal data sales are acceptable but rather apply the same cultural rules that [15] highlighted, notably that some relational framings make personal data sales acceptable while others do not. If personal data are like market goods, relational framing should have substantial effects on judgments of a sale, such that a pure market framing makes a transaction acceptable to participants, as concretely manifested via low or zero rates of refusing to price. This is also not reflected in the results.

Refusal to price is not due to uncertainty about market price, which is controlled for in the explicit information condition. Refusal to price is similarly not due to disinterest in the price information, which is controlled for in the veiled information condition. People can recognize that personal data has a market value and be interested to know the price, but still condemn the idea of such a market and also refuse to participate when they have the option of refusal. The experimental results support various proposals that would move the emphasis on privacy law away from a market-premised, contractual model of privacy based on notice and consent [20] [5] [19].

The results also speak to existing scholarship on data labor and personal data as property. Some scholars and politicians alike have proposed that popular discontent with the big data society and commercial data collection might, at least partly, be addressed by redistribution of the economic gains of data collection through the recognition and financial compensation of data ownership or data labor [7] [17]. The results of this study, specifically that the option to share economic proceeds does not reduce rejection of data sales, suggest that data property and revenue sharing options would
have to be carefully designed to be consistent with the conceptual categories ordinary people ascribe to their data. Transferring or sharing the ownership of personal data and derived economic income streams may not be enough to address the concerns of ordinary people about personal data markets if such suggested reforms are perceived to merely extend or reinforce personal data markets.

In this experiment, market appropriateness was discussed as a binary indicator, but it is more likely on a continuous spectrum, with people relatively more or less accepting of market allocation rather than strictly accepting or not. The results here simply establish an important divergence between cultural norms on the one hand and the current form of typical and legal data transfers, as operationalized by a discussion of market appropriateness. There is more to investigation, and future research can better delineate acceptable modes and spheres of exchange for personal data on a spectrum that may better elucidate how ordinary people make trade-offs among culturally incommensurate goods, particularly in the domain of technology interacting with core personal values, such as those implicated by privacy.

8 LIMITATIONS
There are limitations intrinsic to the methodology of the study. Those who participate in online forums are more technologically knowledgeable than the general population and so may not be fully representative of the reactions of a more heterogeneous population [2]. In this experiment, however, any potential sampling concerns tend to undercut the possibility that the experimental results occur due to the sampling venue or methodology. Quite the opposite, the participants in this study earned money by providing some form of information about themselves. Those who do not participate in online survey markets are likely even more resistant to data markets than the instant data pool.

Also, this study does not support the broad conclusions that all people reject the notion of data markets. Indeed, some participants did agree to price the data, did then price at market rates, and did suggest low percentages for sharing the proceeds. It may be that the population has heterogeneous attitudes about data sales, and that there is a distinct population of lay people fully on board with personal data markets even when others are not. Such heterogeneity should be examined in future studies. This possibility does not negate the results or conclusion of this study but rather provides experimental evidence to support concerns raised by e.g. [22] regarding the distributive effects of data markets.

Another possible critique of the study is that it does not replicate the model vignette of [15] despite taking inspiration from that study. Thus a conclusion that personal data are not like pens is not directly supported by the experimental results. But, the results of this study stand on their own. The main result of this study is that high rates of participants refuse to price data in a typical and legal market transaction. The results do not need comparison with [15].

Another limitation relates to the relevance of the study, namely whether the lack of cultural acceptance of a market for personal data should matter from a legal perspective. Many economically significant activities are also culturally significant, even sacred, such as getting married or mourning a death. Why should the market framing of personal data be problematic if other significant events can be handled with contracts and monetary pricing, such as contracts with wedding planners or standard pricing at funeral homes? But this oversimplifies the matter. Law does recognize cultural categories as relevant to legal ones. Consider that the law for gifts is not the same as that for market transactions, and that such legal designations can impact individuals’ willingness to participate in different forms of exchange [24]. Culture and law alike do consistently recognize that norms and expectations for market transactions are not the same as those in other spheres of exchange [13], and doing so in the case of personal data should not be considered particularly unusual.

9 CONCLUSION
This study suggests that the current legal regime and market realities governing personal data may not provide conceptual consistency with the intuitions or desires of ordinary people. Lay people refuse data sales at high rates, even when there is an opportunity to share the economic gains of such sales with data subjects. Likewise, lay people refuse data sales at high rates even after voluntarily learning about the market rates for data. Lay people are thus aware of data markets as a concept and can be interested in them, all while rejecting personal data sales. The results have direct implications for privacy law and for responsible AI. Privacy law regimes premised on market metaphors will likely continue to fail in meeting the expectations and desires of ordinary people, and so legal reforms should look beyond contractual models when seeking to update privacy laws. Likewise, the responsible AI community should look beyond notice and consent paradigms to account for cultural norms when considering fair data collection and exchange practices.

ACKNOWLEDGMENTS
Funding was provided by ETH Zurich’s Center for Law and Economics. For helpful comments and insights on earlier drafts, the author thanks Svetlana Yakovleva and Alexander Stremitzer, as well as participants of the Information Law Institute Privacy Research Group at NYU Law School, the ICML 2020 Workshop on Economics of Privacy and Data Labor, and the Privacy Law Scholars’ Conference 2020.

REFERENCES
[1] 2001. E-Privacy in Second Generation E-Commerce: Privacy Preferences versus Actual Behavior. Proceedings of the Third ACM Conference on Electronic Commerce, edited by Michael P. Wellman and Yoav Shoham. New York: Association for Computing Machinery.
[2] 2019. How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples. 2019 IEEE Symposium on Security and Privacy.
[3] A. Acquisti, L. Brandimarte, and G. Loewenstein. 2015. Privacy and Human Behavior in the Age of Information. Science 347 (2015), 509–515.
[4] K Crawford and T Paglen. [n.d.]. Excavating AI: The Politics of Training Sets for Machine Learning. Retrieved January 15, 2021 from https://excavating.ai
[5] KE Davis and F Marotta-Wurgler. 2019. Contracting for Personal Data. New York University Law Review 94, 4 (2019), 663.
[6] L. Fair. [n.d.]. FTC’s 5 billion Facebook settlement: Record-breaking and history-making. Retrieved May 1, 2020 from https://www.ftc.gov/news-events/blogs/business-blog/2019/07/ftcs-5-billion-facebook-settlement-record-breaking-history

16 Similar arguments can be made regarding the study’s relevance to the Responsible AI movement.
[7] C Fisher. [n.d.]. Presidential hopeful Andrew Yang wants your data to be your property. Retrieved May 1, 2020 from https://www.engadget.com/2019-11-14-andrew-yang-2020-tech-policy.html
[8] AP Fiske. 1991. Structures of social life: The four elementary forms of social relations. Free Press.
[9] G Fleishman. [n.d.]. How the tragic death of Do Not Track ruined the web for everyone. Retrieved May 1, 2020 from https://www.fastcompany.com/90380668/how-the-tragic-death-of-do-not-track-ruined-the-web-for-everyone
[10] K Hill. [n.d.]. ‘Do Not Track’, The Privacy Tool Used by Millions of People, Doesn’t do Anything. Retrieved May 1, 2020 from https://gizmodo.com/do-not-track-the-privacy-tool-used-by-millions-of-people-182868324
[11] J Hirshleifer. 1980. Privacy: Its Origin, Function, and Future. The Journal of Legal Studies 9, 4 (1980), 649.
[12] Y. Lee and R Weber. [n.d.]. Revealed Privacy Preferences: Are Privacy Choices Rational. Retrieved May 1, 2020 from https://drive.google.com/file/d/1FWt9gAaOInU6YIY69_6BRBBQw80xa/view
[13] M Mauss. 1925. The Gift: The Form and Reason for Exchange in Archaic Societies.
[14] D McCabe. [n.d.]. Congress and Trump Agreed They Want a National Privacy Law. It Is Nowhere in Sight. Retrieved January 10, 2021 from https://www.nytimes.com/2019/10/01/technology/national-privacy-law.html
[15] A.P. McGraw and P.E. Tetlock. 2005. Relational Framing, and the Acceptability of Exchanges. Journal of Consumer Psychology 15, 1 (2005), 2–15.
[16] H Nissenbaum. 2010. Privacy in Context. Standard Law Books.
[17] R.A. Posner and E.G. Weyl. 2018. Radical Markets: Uprooting Capitalism and Democracy for a Just Society. Princeton University Press
[18] N Rea. [n.d.]. ow ImageNet Roulette, an Art Project That Went Viral by Exposing Facial Recognition’s Biases, Is Changing People’s Minds About AI. Retrieved January 5, 2021 from https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305
[19] N Richards and W. Hartzog. 2005. Taking Trust Seriously in Privacy Law. Stanford Technology Law Review 19, 1 (2005), 431–472.
[20] L.J. Strahilevitz. 2010. Reunifying Privacy Law. California Law Review 98, 6 (2010), 2007–2048.
[21] D Svirsky. 2019. Why do people avoid information about privacy? Journal of Law & Innovation 2, 23 (2019), 23–46.
[22] C. R. Taylor. 2004. Consumer Privacy and the Market for Customer Information. The RAND Journal of Economics 35, 4 (2004), 631 – 650.
[23] P.E. Tetlock, O. Kristel, B. Elson, and J. Lerner. 2000. The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals. Journal of Personality and Social Psychology 78, 5 (2000), 853–870.
[24] R. Titmuss. 1997. The Gift Relationship: From Human Blood to Social Policy. The New Press.
[25] J Turow, M Hennessey, and N Draper. [n.d.]. The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation. A Report from the Annenberg School for Communication, University of Pennsylvania. Retrieved May 1, 2020 from https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf
[26] A.G. Winegar and C.R. Sunstein. 2019. How Much Is Data Privacy Worth? A Preliminary Investigation. Journal of Consumer Policy 42 (2019), 425–440.
[27] Shoshana Zuboff. 2019. The Age of Surveillance Capitalism. Public Affairs.