ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

HOW MUCH IS DATA PRIVACY WORTH? A PRELIMINARY INVESTIGATION

Angela G. Winegar
Cass R. Sunstein

Forthcoming in *Journal of Consumer Policy*

Discussion Paper No. 1017

09/2019

Harvard Law School
Cambridge, MA 02138

How Much Is Data Privacy Worth? A Preliminary Investigation

Angela G. Winegar* · Cass R. Sunstein**

ABSTRACT

Do consumers value data privacy? How much? In a survey of 2,416 Americans, we find that the median consumer is willing to pay just $5 per month to maintain data privacy (along specified dimensions), but would demand $80 to allow access to personal data. This is a "superendowment effect," much higher than the 1:2 ratio often found between willingness to pay and willingness to accept. In addition, people demand significantly more money to allow access to personal data when primed that such data includes health-related data than when primed that such data includes demographic data. We analyze reasons for these disparities and offer some notations on their implications for theory and practice. A general theme is that because of a lack of information and behavioral biases, both willingness to pay and willingness to accept measures are highly unreliable guides to the welfare effects of retaining or giving up data privacy. Gertrude Stein's comment about Oakland, California may hold for consumer valuations of data privacy: "There is no there there." For guidance, policymakers should give little or no attention to either of those conventional measures of economic value, at least when steps are not taken to overcome deficits in information and behavioral biases.

KEY WORDS: Data privacy, Endowment effect, Willingness to pay, Willingness to accept, Behavioral biases

* Master of Business Administration Candidate, Harvard Business School; Master of Public Policy Candidate, Harvard Kennedy School.

** Robert Walmsley University Professor, Harvard University.

I. WHY VALUE DATA PRIVACY?

Social media platforms extract a great deal of data from users, raising questions about possible invasion of privacy. Incidents of improper use of personal data have contributed to increasing calls for governments to protect such data. In Europe, for example, the General Data Protection Regulation (GDPR) was outlined in 2016 and implemented in 2018 to strengthen and to standardize data privacy law across the European Union, increase penalties for organizations that breach privacy regulation, and mandate consumer protection provisions like the Right to Be Forgotten (requiring organizations to delete any data collected upon request). In the United States, individual states are begun to move in similar directions, with Vermont's H.764, "An Act Relating to Data Brokers and Consumer Protection," enacted in May 2018, and California's "The California Consumer Privacy Act of 2018" (CCPA).

Laws of this kind are designed to increase people's control over access to and use of their personal data. But there remains an unanswered question: How much do consumers actually care about data privacy? It is standard to answer questions of that kind by asking about consumers' willingness to pay (WTP) or willingness to accept (WTA). The general theory is that WTP or WTA is the best available measure of how much consumers value goods, and hence about the welfare effects of providing them. The theory is most secure when consumers are asked how much they are willing to pay for a good with which they are familiar, such as food or clothing; it may also be secure when people are asked how much they are willing to pay to *use* a social media outlet with which they have experience (Allcott et al. 2019). It is far less secure when consumers are asked to pay for a good whose actual effects on their lives are unfamiliar, unclear, or ambiguous. In such cases, WTP or WTA may amount to a stab in the dark.

In real markets involving data privacy, both WTP and WTA are highly relevant (Acquisti et al. 2013). In some contexts, consumers are asked to pay a specified amount of money to give up their privacy. In other contexts, they are asked how much they would demand to do so. But for reasons that we will explore, the WTP/WTA criteria run into serious concerns in the context of data privacy.

Our principal goal here is to report on an experiment designed to cast light on these questions (for a valuable related study, see Acquisti et al. 2013). The primary finding is simple: The median participant is willing to pay relatively little ($5 per month) for privacy, but demands much more ($80 per month) to give up privacy. This is an unusually large disparity between WTP and WTA – a kind of superendowment effect. We speculate that one reason for the sheer magnitude of the endowment effect is that both WTP and WTA answers are largely *expressive*, and hence do not give a helpful account of the welfare effects of maintaining or relinquishing data privacy. We support that speculation by reference to findings about consumers' pervasive lack of information in the context of data privacy, and to findings of relevant behavioral biases.

Our most general claim is that while WTP and WTA are useful measures of welfare in numerous settings, they have limited value in the context of data privacy in light of limited information and behavioral biases. This claim has broad implications for use of WTP and WTA to measure the welfare effects of multiple goods, where consumers lack the information that would enable them to solve what is essentially a prediction problem.

II. EXPERIMENTAL DESIGN

To avoid possible distortions, we ran an exceedingly simple, single stated-preference survey by asking each respondent only one question up front without other context, then following up with other background and demographic questions. This survey was run on Amazon's Mechanical Turk. Respondents were American; they were compensated for their time.

Each respondent was presented one of eight questions at random. We asked eight questions because we were interested in the effects of specifying what "personal data" means, and in particular in seeing whether WTP or WTA answers would vary with the specification. Question 1 and Question 5 were baseline questions; they had no such specification and differed only in asking about WTP or WTA. The remaining questions varied in the (very brief) specification of the meaning of "personal data," with the hypothesis that the variations might prime certain concerns. The questions took the following form:

1. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. For what amount (in US dollars) per month would you be willing to allow all these entities to access your personal data?*

2. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. For what amount (in US dollars) per month would you be willing to allow all these entities to access your personal data (name, age, gender, profession, household income, address, picture)?*

3. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. For what amount (in US dollars) per month would you be willing to allow all these entities to access your personal data (age, gender, political affiliation, religion, sexual orientation)?*

4. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. For what amount (in US dollars) per month would you be willing to allow all these entities to access your personal data (age, gender, personality traits, physical and mental health)?*

5. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. What would you be willing to pay per month (in US dollars) to delete all of your personal data from all parties that hold it?*

6. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. What would you be willing to pay per month (in US dollars) to delete all of your personal data (name, age, gender, profession, household income, address, picture) from all parties that hold it?*

7. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. What would you be willing to pay per month (in US dollars) to*

delete all of your personal data (age, gender, political affiliation, religion, sexual orientation) from all parties that hold it?

8. *It is known that most online platforms (e.g., Facebook, Google, other digital marketers) collect user personal data. What would you be willing to pay per month (in US dollars) to delete all of your personal data (age, gender, personality traits, physical and mental health) from all parties that hold it?*

We then asked these questions:

9. *Do you feel you have an understanding of how much of your personal data is collected and used by online entities (e.g., Facebook, Google, other digital marketers)?*

With the response options of "great understanding", "decent understanding", "some understanding", "limited understanding", and "no understanding."

10. *How do you feel about digital advertisers collecting your personal data online?*

With the response options "very positive", "somewhat positive", "I have no feelings about this", "somewhat concerned" and "very concerned".

We also asked respondents for general demographic data (age, gender, political affiliation, household income).

III. RESULTS

Of 2,440 respondents (with the first question being equally randomly assigned across respondents), twenty-four were removed for not completing the survey, leaving 2,416 responses. After random assignment, the final response numbers were 301, 296, 299, 301, 299, 311, 311 and 298 for questions 1 through 8 respectively.

Of the 2,416 respondents who completed the survey, 63 responded with exceptionally high (greater than $25,000 per month) valuations of data privacy (unstandardized summary statistics may be found in Table 1). Twenty-eight stated that their willingness to accept was one million dollars or more per month, and three stated that their willingness to pay was one million dollars or more per month. For purposes of analysis, we took steps to standardize responses. To determine a threshold at which to cut off responses, we took the 99th percentile of income in 2017, which IPUMS reported as roughly $300,000 per individual (IPUMS-USA). This equates to roughly $25,000 per month, and since it seems unlikely that participants would actually be willing and able to pay this amount (only 40 respondents of the 2,416 reported household income over $200,000 per year), we converted any amount of willingness to pay greater than $25,000 to $25,000. In order to compare across both willingness to pay and willingness to accept, we also standardized any willingness to accept responses larger than $25,000 at $25,000.[1]

The updated summary statistics with the outliers reassigned can be found in Table 2.

A. Sample Demographics and General Concern About Data Privacy

Demographic data of respondents are summarized in Tables 3, 4, and 5. It is worth noting the disproportionately high participation of millennials (age 22-37), who account for roughly 60% of respondents.

Respondent's self-reported knowledge of data collected on them online, and their concern about the collection of their personal data online, are summarized in Table 6. Over 70 percent of respondents reported feeling "somewhat concerned" or "very concerned" about the personal data that are collected on them online. Even among respondents who reported limited (36) or no understanding (391) of data that is being collected on them, the overwhelming majority reported

[1] One participant stated a willingness to pay to prevent access to personal data at $-10, perhaps because he or she expected adverse effects if their personal data were no longer collected by social networking and data brokerage companies. This was also re-assigned a value of zero.

feeling "very concerned" or "somewhat concerned" at their private data being collected. Notably, the proportion of respondents that reported feeling "very positive" about their personal data being collected by online advertisers is disproportionately high among respondents who also reported a "great understanding" of what data is being collected: Only 31 of the 2,107 respondents with a less than "great understanding" reported feeling very positive about their personal data collection, while 76 of the 309 respondents with a "great understanding" reported feeling very positive about their personal data collection.

This raises the possibility that for some segment of the population, having personal data collected and the associated results (receiving ultra-targeted ads, experiencing greater ease of purchasing products online, and so forth) provide a positive value, potentially greater than the privacy concerns associated with sharing their data. To be sure, this may be an issue with self-reporting; people who feel positively about data collection might also desire to appear to be more knowledgeable on the topic. But it may also indicate that educating consumers about what data is collected and why, potentially through a "right to know" policy and information campaigns about data collection, may incline citizens feel more positively about their personal data privacy.

B. Summary Statistics

Full summary statistics of responses can be seen in Tables 1 and 2 in the Appendix. Median and mean numbers for each category can be seen below, with the standardized (max at $25,000, min at $0) mean in parentheses.

General: WTP $5 ($290.9) / WTA $80 ($1606)
Demographics: WTP $5 ($150.8) / WTA $77.5 ($1843.5)
Identity: WTP $5 ($118.8) / WTA $50 ($1065)
Health: WTP $5 ($146.5) / WTA $100 ($2921)

WTA is obviously much higher than WTP for both medians and means. It is noteworthy that 14 percent of respondents were not willing to pay anything for data privacy. Note that the mean numbers can easily be seen to understate the disparity between WTA and WTP. As noted, 57 participants required at least $25,000 a month, or greater than $300,000 per year, to sell their personal data, with some entering as many numbers as the response would allow. At least in the context of a survey, these respondents appear to think that no amount of compensation is sufficient to justify relinquishing their personal privacy.

We also found differences in WTA numbers (but not WTP), depending on the description of "personal data." Without any specification, participants required $80 per month in order for advertisers to use their data. When personal data are described as including "name, age, gender, profession, household income, address, and picture," respondents valued it similarly – with a median around $77.5 per month. When personal data are described as including "age, gender, political affiliation, religion, and sexual orientation," respondents actually valued their data significantly lower, with a median around $50. But when personal data are described as including "age, gender, personality traits, and physical and mental health," respondents' valuations jumped to $100 per month. The differences in means among the three groups are statistically different in an ANOVA test at the .001 level (see Table 7).

Finally, it is important to emphasize the extraordinarily high variance in responses. That WTP ranged from $-10 (see footnote above) to *billions* of dollars per month, and that WTA ranged from $0 to more than *trillions* of dollars per month, suggest another challenge to both theorists and policymakers in using these traditional measures to value data privacy. At least where there is not a great deal of variance, a median value of a good may sometimes be appropriately applied to an entire population in order to maximize welfare. But the high variance in responses indicates that for data privacy, that approach would likely harm many people, since any median is so vastly higher and lower that the personal valuation of numerous respondents. It is usual, and right, to think that with high levels of variance, a degree of personalization and targeting would be appropriate if feasible (Allcott and Sunstein 2015). But apart from feasibility, there is a further problem with that approach here. The extraordinary variance might be seen as evidence not of diverse valuations of data privacy, but that people's answers reflect something other than the anticipated welfare effects of data privacy – a point to which we will return.

IV. PUZZLES AND EXPLANATIONS

These findings leave two evident puzzles. The first involves the very large difference between WTP and WTA (for a similar finding in a field experiment, see Acquisti et al. 2013). The second involves the effects of the description of "personal data."

A. A Superendowment Effect

In many important contexts, a difference between WTP and WTA has been observed (Kahneman et al. 1990), with some attributing the resulting "endowment effect" to the evolutionary process (Huck et al. 2005). The WTA:WTP ratio for many goods (like mugs and event tickets) is typically on the order of 2:1. Some studies (Cummings et al. 1986) observe much larger ratios for environmental goods, such as protection of endangered species, sometimes on the order of 10:1. The magnitude of this disparity remains to be explained, but a plausible account is that in the environmental context, a high figure for WTA reflects a kind of *moral outrage*. For an environmental good (clean air, safe drinking water, an endangered species), the WTA question undoubtedly triggers moral concerns. Consumers and others do not like to think that they are responsible for producing some kind of environmental wrong, loss, or harm in return for a specified amount of money, at least if that amount is not very high.

The 16:1 ratio found here is extremely high – among the very highest in the existing literature on the endowment effect. It may fairly be described as a *superendowment effect*. It is noteworthy that in the context of *use* of social media, such as Facebook and Twitter, ratios similar to that found here have also been found (Sunstein 2019).

Whether the disparity involves use of social media or data privacy, it is a genuine puzzle, and the explanation may not be the same in the two contexts. For data privacy, begin with the WTP number. Because the monetized value of data privacy is not self-evident, it is difficult to say, in the abstract, whether the $5 monthly amount is low or high. It is possible, of course, that many people do not care at all about data privacy, and that many people care only a little. Recall that about 14 percent of respondents registered a WTP of zero. But because the WTA answers are so much larger, it seems too simple to say that data privacy does not matter to people.

If we stipulate that the WTP number is relatively low, we might think that for many respondents, it is best taken as *expressive*, a kind of protest answer, rather than as a reflection of a considered judgment about the anticipated welfare effects of having or not having privacy. People might think that they already have the right to delete personal data; the idea of paying for something that they already "own" may lead consumers to say $0 or to offer low numbers. Loss aversion undoubtedly plays some kind of role. If people are asked to pay more than the reference point (in this case $0), they will rebel (Bewley 1999). They might well think that the change is unfair and hence the protest. We suspect that the numbers – so low in comparison to those for WTA – are best explained, at least in part, by reference to this factor.

Turn now to the WTA numbers, and let us stipulate that they are relatively high. (For a certain number of users, it is not necessary to stipulate; the numbers are self-evidently high.) Those answers might also be expressive. People might be effectively saying: "I greatly value my privacy, and you are going to have to pay me a great deal to give it up." When people demand high amounts to allow "access to" their personal data, or when they say something like, "no amount is high enough," they are expressing moral outrage, rather than making a judgment about the welfare effects of allowing access. Shane Frederick has put this point crisply, suggesting, "sellers use high values to signal that their dignity is not for sale, and buyers use low values to signal their refusal to accept the implication that they are entitled to only intermediate levels of privacy" (quoted in Acquisti et al. 2013, p. 255). We will return to this point shortly.

There is a separate point, and it involves opportunity costs. The WTP question puts opportunity costs on the cognitive table, at least for many people much of the time: When people are asked how much they are willing to pay for some good or service, they are often going to think what else they could do with that money. The WTA question is different. When people say that they would demand a very high amount of money to give up some good that they own (e.g., coffee mugs, lottery tickets), they might not be focused on other potential uses of that money (Frederick et al. 2009). For that reason, there is additional reason to doubt whether a very high median, in response to WTA questions, is sufficiently informative about the welfare effects of data privacy.

B. Which Personal Data?

In the abstract, the term "personal data" might be vague or ambiguous; people might not have a sense of its concrete meaning, or if they do, it is because specific illustrations readily come to mind. We expect that if people think that "personal data" means "first letter of last name," or "nation," or even "city," they would not be especially concerned. But if it means "whether they have committed a crime" under existing law, or "whether they are engaging in a secret romance," or "social security number, birthday, and mother's maiden name," or "location within a meter for the past five years," they might much want to keep the relevant data private. We intended our brief specifications of the meaning of "personal data" to be tests of whether WTP or WTA would increase with varying "primes."

For this reason, it should not be especially surprising to find that a specific reference to "personality traits and physical and mental health" inflates WTA. For many users, that information is distinctly sensitive. People may well consider personal and mental health to be involve quintessentially private facts, or less attractive traits, and therefore demand more in order to allow that information to be made public (Huberman et al. 2005). People may be uncomfortable about their weight, depression, anxiety, or other factors, and therefore willing to pay more to prevent such information from being disclosed. Note in this regard that Huberman, Adar, and Fine ran a second-price auction to determine how individuals value their height and weight data. They found that the less desirable a personal trait, the greater the price a person demands for releasing the information (Huberman et al. 2005) – a finding that may explain ours here.

We might also speculate that some people may have been influenced (for example, via the Health Insurance Portability and Accountability Act of 1996, known as "HIPAA") to believe that their health data should be private, and while demographic data may contribute to better advertisements (which have some value to them), they may see no need for advertisers to have their health data. The latter conclusion would be in line with previous findings (Acquisti et al. 2013), which outlined positive- and negative-reinforcement loops with data privacy: As privacy becomes better protected, it becomes more highly valued by individuals, and as it becomes more highly valued, it becomes better protected, and vice versa. In other words, consumer valuation of privacy is endogenous to the existing legal regime. This is a particularly relevant finding for policymakers to consider, as it suggests that any enacted data privacy policy might influence consumers' preferences and valuations of their personal data.

It is less straightforward to explain why "age, gender, political affiliation, religion, and sexual orientation" produce a *lower* WTA figure than the unspecified data category, or than "name, age, gender, profession, household income, address, and picture." Apparently people believe that household income, address, and picture are relatively sensitive – more so than standard demographic questions. Perhaps this too is a result of priming effects: Age, gender, political affiliation, religion, and sexual orientation are all descriptors that people can put on their public Facebook profile, while household income may be taken to be sensitive, and is more socially taboo to share.

But a puzzle remains. The median WTP was unaffected by the various specifications of personal data. We are not sure how to explain that finding. But note that people were not willing to pay much to ensure data privacy, perhaps for the expressive reason we have outlined. In these circumstances, it may not be so surprising that the (relatively minor?) signals given by the specification would not inflate or deflate WTP.

V. INFORMATION AND BIASES

If standard economic valuations are reliable, WTP would seem to provide some clues to the monetary value of data privacy – unless it reflects moral outrage or protest answers rather than projected welfare effects. If we do not trust WTP, WTA might seem better (Allcott et al. 2019) – unless it reflects moral outrage or protest answers rather than projected welfare effects.

But the significant disparity between them, combined with the best explanation for the numbers found here, raises serious questions about both WTP and WTA in this setting. To produce reliable measures of welfare effects, consumers would have to have significant information about the likely consequences (good and bad) of having or not having data privacy, and the probability of that each of these consequences would come to fruition. And even if they had that information, they would have to process it in a relatively unbiased way.

A. Tradeoffs and Information

It is true that on standard economic assumptions, personal information will be revealed by each party during a transaction at an optimal level, depending on the benefits and costs of disclosing it (Stigler 1980). This would imply that individuals are rational in their disclosure of personal data, and in this specific context, that users make reasonable tradeoffs in considering whether to disclose personal data to Facebook and other online data collectors. If social media providers and others do not clearly signal how they store, protect, and use personal data, or do not offer sufficiently clear and salient signals, perhaps the best response is to require them to do so, and then to allow consumers to make their choices.

The superendowment effect found here, alongside other research on data privacy (Acquisti et al. 2013), raises serious doubts about whether users are now making reasonable tradeoffs when exchanging personal data for free platform use; whether clearer disclosure could enable them to do so; and whether use of WTP or WTA figures would enable policymakers to make such tradeoffs. Once again: To make the relevant tradeoffs, users would need, at a minimum, a sense of what personal data is collected, how it might be used and stored, and the costs and benefits of its collection and use (including probabilities – low, we might hope – of terrible misuse, however that is measured). Obtaining and assessing that information is challenging. And even if it were possible to overcome that obstacle, consumer preferences might be endogenous to the method of elicitation, raising serious questions about whether they are stable and also about their normative standing (Acquisti et al. 2013).

It now seems evident that consumers lack a clear knowledge of how their data is collected, secured, and used online. Only 47 percent of our survey respondents felt they had a "great" or "decent" understanding of what personal data is collected online. On the basis of a review of existing research, Acquisti et al. concluded that in online settings, consumers are now unable to make informed decisions about the privacy of their personal data (Acquisti et al. 2016). Consumers typically have highly imperfect information about whether their data is collected, which data is collected, and how their data is used by online advertisers. Acquisti et al. also noted a serious incongruity. On the one hand, consumers say that they greatly value data privacy. On the other hand, consumers are quite willing to give up data privacy in exchange for ease of internet use, and in some studies, they demand very little in order to do that (Spiekermann et al. 2001). The divergence between statements of value and actual behavior, together with imperfect information and the wide variation in monetary valuation depending on seemingly irrelevant contextual features (Acquisti et al. 2013), make it exceedingly difficult to place any kind of monetary value on data privacy.

In support of that conclusion, an instructive study asked respondents to measure the degree to which a described scenario either met the respondent's privacy expectations or conformed to a privacy notice (Martin 2015). The study found that when participants were shown a privacy notice, they often perceived the notice as offering greater protections than it actually did. Notably, and in what is apparently a form of motivated reasoning, respondents projected their own preferred privacy protections onto the notice.

In an especially important study, with findings broadly compatible with those here, Acquisti et al. ran a field experiment in which subjects were asked to choose between gift cards that varied in both value and with respect to their privacy features (Acquisti et al. 2013). As here, the authors found a significant difference between WTP and WTA. If subjects believed that their privacy would be protected by default, they were five times more likely to reject cash offers for their personal data than if they believed their privacy was not protected by default. They also found evidence of ordering effects (Schwarz 1999), meaning that the order in which they presented offers for data privacy mattered. In short, consumers' valuations of personal data are not stable but malleable, and they can be affected seemingly irrelevant factors (Acquisti et al. 2013).

B. Behavioral Biases

A great deal of research also finds that consumers do not place much value on data privacy (Spiekermann et al. 2001; see Acquisti et al. 2013, for an overview). Our finding that the median WTP for data privacy is just $5 would seem to support this conclusion. Does that tell policymakers much? Perhaps not, for a number of behavioral biases may influence respondents' valuations of data privacy. Motivated reasoning is one possibility. Acquisti et al. (2016), found not only that consumers lack an accurate picture of how their data is collected and secured online, but also that their perception of data security is subject to their own preferences and hopes of what data security looks like (see also Martin 2015).

Unrealistic optimism (Sharot 2012) and present bias (Wang and Sloan 2018) may also be at work. For example, consumers may undervalue data privacy on the optimistic assumption that their personal data will not be misused. That assumption might, of course, turn out to be correct, but if consumers are prone to unrealistic optimism (Sharot 2012), then they will be inclined to accept it even if it is wrong. With respect to choices about whether to give up data privacy, consumers might undervalue the long-term risks and emphasize the short-term gain (including access to social media platforms and websites). If so, their judgments will again be distorted (Wang and Sloan 2018).

Although all of these factors may lead to undervaluation, it would be reckless to conclude that consumers generally care too little about data privacy. Our own findings of a high WTA, and an extraordinarily high WTA for some people, suggest that overvaluation is a possibility. The availability heuristic (Tversky and Kahneman 1973) may lead people to think that the risks of a breach are much higher than they are, especially in the aftermath of a well-publicized breach. If people are subject to "probability neglect" (Sunstein 2002), the intense emotions associated with a breach may lead people to focus on worst-case outcomes, rather than the probability that they will occur. Because some people place a high value on the very idea of data privacy, they may not be enthusiastic about making tradeoffs, even when that it is the rational thing to do. Recall Frederick's suggestion that "sellers use high values to signal that their dignity is not for sale" (quoted in Acquisti et al. 2013, p. 255).

C. The Possibility of International Differences

To elaborate on a point made earlier: Acquisti et al. (2013) noted that the more something is made private, the more highly it is valued, and the less private the data, the lower it is valued. This point is connected with the possibility of large differences across regions, nations, and cultures. The survey reported here involved Americans, and it is possible that such differences, including priming via policy, may yield very different results in Europe, where (we hypothesize) data privacy may be more likely to considered a right, or in China, where personal data is constantly readily available to government officials, companies, and others.

To know whether this is so, surveys would be quite valuable. But we offer a cautionary note about a reasonable reaction to our findings, which is that the valuation of data privacy is predictably low among Americans, and that it would unquestionably be higher in Europe. The cautionary note is that WTA is not low at all, and for many Americans, it is extraordinarily high. The real news is not that the values are low, but that WTA is so much higher than WTP. It is possible, of course, that both WTA and WTP would be higher in Europe; that would be illuminating to learn. But to date, we are unaware of any study that finds a high WTP for data privacy in Europe (Acquisti, et al. 2013, contains an overview; Spiekermann et al. 2001). We also speculate that individuals worldwide are also likely to lack relevant information and to be subject to the behavioral biases outlined here.

D. Welcome to Oakland?

Our primary goal here is not to offer a policy recommendation on how to value data privacy, but to report our findings and to suggest that they raise serious doubts about both WTP and WTA in this setting. It would be reasonable to conclude that with respect to privacy, it is essential for policymakers to attend not to WTP or WTA, but to what consumers actually gain and what they (might) lose by allowing access to personal data. Consumer welfare deserves priority, at least if welfare is properly understood. A form of welfare analysis, focused on actual or anticipated gains and losses, would be preferable to reliance on WTP or WTA.

Of course it is true that any such analysis is difficult to perform without an understanding of consumer preferences and values. A welfare analysis might be accompanied by consideration of the preferences of informed (and behaviorally unbiased) consumers, through reliable elicitation procedures (Allcott and Sunstein 2015). Consideration of those preferences would be exceedingly valuable in view of the fact that some consumers undoubtedly do have views about the appropriate tradeoffs between privacy and other values; they might care greatly about privacy, or a little, or not at all. Their preferences should be counted, at least if they are informed and free from behavioral biases. Note again the highly diverse responses in our data, raising the possibility that targeted or personalized data privacy policies would be welfare-maximizing (ibid.).

At the same time, design of reliable elicitation procedures would present special challenges in light of the fact that consumer preferences seem endogenous to context and design as well as to existing policy (Acquisti et al. 2013). As Gertrude Stein observed about Oakland, California: "There is no there there" (Stein 1937, p. 289). With respect to consumer preferences about data privacy, the problem may go deeper than a lack of information and behavioral biases. For some or many consumers, there may be no there there.[2]

VI. CONCLUSION

[2] It is also important to note that if and to the extent that some social media providers – including Facebook – have monopoly power, they exercise it not by charging users (access is free), but by extracting more data than they would in a competitive market. That issue deserves far more attention.

In the context of data privacy, we have found a superendowment effect, with a median WTP of $5 and a median WTA of $80. It is tempting to suggest that future work should investigate the reasons for this disparity and lean in the direction of one or another number (Allcott et al. 2019) – or perhaps start with the thought that the two figures suggest lower and upper bounds.

We would resist that conclusion. Both WTP and WTA are best taken as predictions of the welfare effects of goods (including intangibles). When they are useful, it is because those predictions tell us something important about those welfare effects. In the context of familiar goods, it is usually safe to assume that the predictions are reliable. In the context of data privacy, that assumption is hazardous. Because of a lack of information and behavioral biases, both WTP and WTA measures are unlikely to be reliable guides to the welfare effects of retaining or giving up data privacy. It is reasonable to speculate that these conclusions could be extended beyond data privacy to other goods and services for which consumers have limited information and are prone to behavioral biases, or for which preferences are constructed, rather than found, by elicitation procedures.

REFERENCES.

A Privacy-Focused Vision for Social Networking | Facebook Newsroom. (n.d.). Retrieved from: https://newsroom.fb.com/news/2019/03/vision-for-social-networking/ (accessed 28 April 2019).

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, *42*(2), 249–274.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442–492.

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2018). The welfare effects of social media. *SSRN Electronic Journal*. doi:10.2139/ssrn.3308640

Allcott, H., & Sunstein, C. R. (2015). Regulating internalities. *Journal of Policy Analysis and Management*, *34*(3), 698–705.

Americans and Cybersecurity | Pew Research Center. (2017, January 26). Retrieved from: https://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/ (accessed 28 April 2019).

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

Bewley, T. F. (1999). *Why wages don't fall during a recession*. Cambridge, Mass.: Harvard Univ. Press.

Boyce, R. R., Brown, T. C., McClelland, G. H., Peterson, G. L., & Schulze, W. D. (1992). An experimental examination of intrinsic values as a source of the WTA-WTP disparity. *The American Economic Review*, *82*(5), 1366–1373.

Brynjolfsson, E., Collis, A., & Eggers, F. (2019). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, *116*(15), 7250–7255.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). How Cambridge Analytica turned Facebook 'likes' into a lucrative political tool. *The Guardian*. Retrieved from: https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm (accessed 13 April 2019).

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), 205316801562207.

Cummings, R. G., Brookshire, D. S., Schulze, W. D., Bishop, R. C., & Arrow, K. J. (1986).

Valuing environmental goods: an assessment of the contingent valuation method. Totowa, N.J: Rowman & Allanheld.

Dietz, S., & Venmans, F. (2019). The endowment effect, discounting and the environment. *Journal of Environmental Economics and Management*.

Frederick, S., Novemsky, N., Wang, J., Dhar, R., & Nowlis, S. (2009). Opportunity cost neglect. *Journal of Consumer Research*, *36*(4), 553–561.

Goldfarb, A., & Tucker, C. (2011). Implications of "online display advertising: targeting and obtrusiveness". *Marketing Science*, *30*(3), 413–415.

Gross, R., Acquisti, A., & Heinz, H. J. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society - WPES '05* (p. 71). Presented at the the 2005 ACM workshop, Alexandria, VA, USA: ACM Press.

Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *IEEE Security and Privacy Magazine*, *3*(5), 22–25.

Huck, S., Kirchsteiger, G., & Oechssler, J. (2005). Learning to like what you have – explaining the endowment effect. *The Economic Journal*, *115*(505), 689–702.

IA Privacy Principles For A Modern National Regulatory Framework. (n.d.). *Internet Association*. Retrieved from: https://internetassociation.org/files/ia_privacy-principles-for-a-modern-national-regulatory-framework_full-doc/ (accessed 28 April 2019).

IPUMS-USA: descr: INCTOT. (n.d.). Retrieved from: https://usa.ipums.org/usa-action/variables/INCTOT (accessed 18 April 2019).

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase Theorem. *Journal of Political Economy*, *98*(6), 1325–1348.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263.

Keeping Advertising Safe and Civil. (n.d.). *Facebook Business*. Retrieved from: https://www.facebook.com/business/news/keeping-advertising-safe-and-civil (accessed 28 April 2019).

Kroes, E. P., & Sheldon, R. J. (1988). Stated preference methods: an introduction. *Journal of Transport Economics and Policy*, *22*(1), 11–25.

Martin, K. (2015). Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing*, *34*(2), 210–227.

O'Donoghue, T., & Rabin, M. (2015). Present bias: lessons learned and to be learned. *American Economic Review*, *105*(5), 273–279.

Rethinking "trust" in a new era of data privacy. (2018, October 10). *GlobalWebIndex Blog*. Retrieved from: https://blog.globalwebindex.com/chart-of-the-week/trust-data-privacy/ (accessed 28 April 2019).

Richterich, A. (2018). *The big data agenda: data ethics and critical data studies*. London: University of Westminster Press.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105.

Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a special price: the true value of free products. *Marketing Science*, *26*(6), 742–757.

Sharot, T. (2012). *The optimism bias: a tour of the irrationally positive brain*. New York: Vintage Books.

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM conference on Electronic Commerce - EC '01* (pp. 38–47). Presented at the the 3rd ACM conference, Tampa, Florida, USA: ACM Press.

Stein, G. (1971). *Everybody's autobiography*. New York: Random House.

Stigler, G. J. (1980). An introduction to privacy in economics and politics. *The Journal of Legal Studies*, *9*(4), 623–644.

Sunstein, C. R. (2002). Probability neglect: emotions, worst cases, and law. *The Yale Law Journal*, *112*(1), 61.

Sunstein, C. R. (2013). Deciding by default. *University of Pennsylvania Law Review*, *162*(1), 1–57.

Sunstein, C. R. (2019). Valuing Facebook. *Behavioural Public Policy*, 1–12.

Swire, P. P., & Litan, R. E. (1998). *None of your business: world data flows, electronic commerce, and the European privacy directive*. Washington, D.C: Brookings Institution Press.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: an experimental study. *Information Systems Research*, *22*(2), 254–268.

Turel, Serenko, & Giles. (2011). Integrating technology addiction and use: an Eepirical investigation of online auction users. *MIS Quarterly*, *35*(4), 1043.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Wang, Y., & Sloan, F. A. (2018). Present bias and health. *Journal of Risk and Uncertainty*, *57*(2), 177–198.

Wardman, M. (1988). A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy*, *22*(1), 71–91.

LEGISLATION.

California, United States
The California Consumer Privacy Act of 2018, 2018. Assembly Bill No. 375.

European Union
Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

APPENDIX.

Table 1 – summary of responses (unstandardized)

	Min.	1%	5%	25%	Median	Mean	75%	95%	99%	Max
Q1. WTA (general)	0	0	0	10	80	1 666 573	500	10 000	100 000	500 000 000
Q2. WTA (demographics)	0	0	0	10	77.5	3.378e+69	500	49 250	1 000 000	1e+72
Q3. WTA (identity)	0	0	0	10	50	3.344e+39	225	5 000	2 980 000	1e+42
Q4. WTA (health)	0	0	0	20	100	3.334e+314	500	1 000 000	60 000 000 000	9.999e+317
Q5. WTP (general)	-10	0	0	2	5	3 652	15	100	3344	1 000 000
Q6. WTP (demographics)	0	0	0	2	5	231.2	20	100	950	50 000
Q7. WTP (identity)	0	0	0	1	5	3254	15	100	1000	1 000 000
Q8. WTP (health)	0	0	0	1	5	3 356 000	20	100	555.74	1 000 000 000

Table 2 – summary of responses (standardized at max = $25,000, min = $0)

	Min.	1%	5%	25%	Median	Mean	75%	95%	99%	Max
Q1. WTA (general)	0	0	0	10	80	1 606	500	10 000	25 000	25 000
Q2. WTA (demographics)	0	0	0	10	77.5	1 843.5	500	25 000	25 000	25 000
Q3. WTA (identity)	0	0	0	10	50	1 065	225	5 000	25 000	25 000
Q4. WTA (health)	0	0	0	29	100	2 921	500	25 000	25 000	25 000
Q5. WTP (general)	0	0	0	2	5	290.9	15	100	3 244	25 000
Q6. WTP (demographics)	0	0	0	2	5	150.8	20	100	950	25 000
Q7. WTP (identity)	0	0	0	1	5	118.8	15	100	1 000	25 000
Q8. WTP (health)	0	0	0	1	5	146.5	20	100	555.74	25 000

Table 3 – Demographics (Gender, Age)

	Gender		**Age**			
	Female	**Male**	**18-21**	**22-37**	**38-53**	**54+**
Q1. WTA (general)	158	143	14	189	69	29
Q2. WTA (demographics)	151	145	10	167	76	43
Q3. WTA (identity)	168	131	23	168	72	36
Q4. WTA (health)	158	143	21	173	72	35
Q5. WTP (general)	154	145	14	164	77	44
Q6. WTP (demographics)	155	156	21	180	69	41
Q7. WTP (identity)	179	132	21	197	70	23
Q8. WTP (health)	146	152	12	199	55	32
Total	**1269**	**1147**	**136**	**1437**	**560**	**283**

Table 4 – Demographics (Politics)

	Politics

	Democrat	Republican	Independent	Other / don't wish to say
Q1. WTA (general)	116	65	108	11
Q2. WTA (demographics)	118	76	90	12
Q3. WTA (identity)	133	77	75	13
Q4. WTA (health)	119	74	99	9
Q5. WTP (general)	115	96	74	14
Q6. WTP (demographics)	123	73	104	11
Q7. WTP (identity)	123	82	92	14
Q8. WTP (health)	111	77	93	17
Total	**958**	**620**	**735**	**101**

Table 5 – Demographics (Income)

	Income							
	Less than \$20,000	\$20,000 to \$34,999	\$35,000 to \$49,999	\$50,000 to \$74,999	\$75,000 to \$99,999	\$100,000 to \$149,999	\$150,000 to \$199,999	\$200,000 or more
Q1. WTA (general)	37	73	41	56	35	38	15	6
Q2. WTA (demographics)	41	58	42	80	39	26	8	1
Q3. WTA (identity)	37	58	62	60	40	28	7	7
Q4. WTA (health)	34	61	54	70	45	27	7	3
Q5. WTP (general)	39	59	52	65	52	22	5	5
Q6. WTP (demographics)	40	52	57	67	54	24	9	8
Q7. WTP (identity)	40	49	41	78	53	31	12	6
Q8. WTP (health)	38	63	48	68	24	38	15	4
Total	**306**	**473**	**397**	**544**	**342**	**234**	**78**	**40**

Table 6 – Concern about collection of data vs. understanding of what is collected (% of respondents)

		Feelings about data collection					
		Very Concerned	Somewhat concerned	No feelings	Somewhat positive	Very positive	Total
Understanding of data collected	**Great understanding**	3.8%	2.4%	1.4%	2.0%	3.1%	**12.8%**
	Decent understanding	9.8%	14.7%	4.7%	4.1%	.7%	**34.1%**
	Some understanding	8.0%	18.6%	5.5%	3.0%	.3%	**35.5%**
	Limited understanding	5.9%	8.3%	1.6%	.2%	.2%	**16.2%**
	No understanding	.8%	.5%	.1%	.0%	.0%	**1.5%**
	Total	**28.3%**	**44.5%**	**13.4%**	**9.4%**	**4.4%**	**100%**

Table 7 – Summary of ANOVA test (WTA, standardized)

```
            Df    Sum Sq   Mean Sq F value   Pr(>F)
Question     3 5.471e+08 182353937   5.854 0.000574 ***
Residuals 1193 3.716e+10  31152226
```

Table 8 – WTP Linear Regression

The omitted categories are women, ages 21 and under who identify politically as democrat and have a household income of $100,000 to $149,999.

```
Residuals:
    Min      1Q  Median      3Q     Max 
 -824.2  -293.9  -132.3     5.7 24626.5 

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)  
(Intercept)                        283.047    284.893   0.994   0.3207  
Age22-37                          -210.777    231.224  -0.912   0.3622  
Age38-53                          -421.120    247.595  -1.701   0.0892 .
Age54+                            -474.561    269.597  -1.760   0.0786 .
GenderMale                         -67.264    104.597  -0.643   0.5203  
PoliticsIndependent                 63.448    126.251   0.503   0.6154  
PoliticsOther / don't wish to say -176.884    257.247  -0.688   0.4918  
PoliticsRepublican                 134.453    130.789   1.028   0.3042  
Income$150,000 to $199,999           8.888    327.969   0.027   0.9784  
Income$20,000 to $34,999             2.938    208.841   0.014   0.9888  
Income$200,000 or more             -51.288    413.248  -0.124   0.9012  
Income$35,000 to $49,999            70.007    213.109   0.329   0.7426  
Income$50,000 to $74,999           368.536    200.793   1.835   0.0667 .
Income$75,000 to $99,999            31.295    215.109   0.145   0.8844  
IncomeLess than $20,000            421.708    224.751   1.876   0.0609 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1802 on 1203 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.01493,	Adjusted R-squared:  0.003469 
F-statistic: 1.303 on 14 and 1203 DF,  p-value: 0.1983
```

<u>Table 9 – WTA Linear Regression</u>

The omitted categories are women, ages 21 and under who identify politically as democrat and have a household income of $100,000 to $149,999.

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4778.2 -1996.1 -1375.7  -886.1 24176.2 

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)                        3174.65     871.53   3.643 0.000282 ***
Age22-37                          -1075.59     718.87  -1.496 0.134861    
Age38-53                          -1466.23     763.57  -1.920 0.055068 .  
Age54+                              -45.84     833.63  -0.055 0.956155    
GenderMale                          173.23     331.02   0.523 0.600842    
PoliticsIndependent                 157.53     387.69   0.406 0.684576    
PoliticsOther / don't wish to say  1431.89     885.12   1.618 0.105988    
PoliticsRepublican                 -284.50     417.86  -0.681 0.496091    
Income$150,000 to $199,999         1630.28    1060.58   1.537 0.124523    
Income$20,000 to $34,999           -785.44     626.22  -1.254 0.209996    
Income$200,000 or more            -1493.80    1456.18  -1.026 0.305179    
```

```
Income$35,000 to $49,999          -884.63     652.46  -1.356 0.175409
Income$50,000 to $74,999          -547.42     620.26  -0.883 0.377657
Income$75,000 to $99,999          -701.57     680.82  -1.030 0.302992
IncomeLess than $20,000            465.28     694.54   0.670 0.503045
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5595 on 1179 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.02119, Adjusted R-squared:  0.009566
F-statistic: 1.823 on 14 and 1179 DF,  p-value: 0.03101
```