# Valuing Intrinsic and Instrumental Preferences for Privacy

Tesary Lin*

October 12, 2020

[*Click here for the latest version*]

**Abstract**

I empirically separate two motives for consumers to protect privacy: an intrinsic motive, which is a "taste" for privacy; and an instrumental motive, which reflects the expected economic losses from revealing one's private information to the firm. While the intrinsic preference is a utility primitive, the instrumental preference arises endogenously from a firm's usage of consumer data. Combining a two-stage experiment and a structural model, I find that consumers' intrinsic preferences for privacy range from 0 to 5 dollars per demographic variable, exhibiting substantial heterogeneity across consumers and categories of personal data. This rich heterogeneity in intrinsic preferences dominates the magnitude of instrumental preferences in my experiment. Consumers self-select into sharing their personal data, driven by the combination of these two preference components. The resulting selection pattern deviates from the "nothing-to-hide" argument, a prediction given by models with pure instrumental preferences. I then evaluate two strategies that firms may adopt to correct for biases caused by this privacy-induced selection when collecting and analyzing consumer data. Both strategies can effectively alleviate bias when consumer data are used for inference.

***Keywords:*** privacy, revealed preference, value of data, experiment, selection, bias

*Boston University Questrom School of Business; tesary@bu.edu. I am grateful to Pradeep Chintagunta, Sanjog Misra, Brad Shapiro, and Oleg Urminsky for their advice and support. Thanks go to Eric Anderson, Michael Dinerstein, Jean-Pierre Dubé, Avi Goldfarb, Wes Hartmann, Yufeng Huang, Garrett Johnson, Nitin Mehta, Kanishka Misra, Sarah Moshary, Harikesh Nair, Sridhar Narayanan, Aniko Öry, Anita Rao, Robbie Sanders, Avner Shlain, Valeria Stourm, Abigail Sussman, Anna Tuchman, Dana Turjeman, Kosuke Uetake, Ken Wilbur, my two reviewers of the MSI Alden G. Clayton dissertation proposal, and participants at various seminars for their thoughtful comments. This study was approved by the Institutional Review Board at the University of Chicago (IRB18-0234).

# 1 Introduction

With the arrival of privacy regulations across the globe, companies increasingly need to seek consumers' consent before collecting and processing their personal data. Within the EU, General Data Protection Regulation (GDPR) mandates that firms deliver transparent information and seek opt-in consent before data processing;[1] as a result, European websites lost 12.5% of their recorded traffic due to increased consumer vigilance (Aridor et al. 2020). Outside Europe, the transparency and consent elements in GDPR have been adopted by privacy regulations in many other countries and local states, including California in the US, Brazil, Chile, and India.[2] They are also part of the core principles for the proposed US federal privacy legislation.[3]

When consent becomes a prerequisite for personal data processing, consumers' preferences for privacy determine what data and whose data firms are able to collect. Consumers exhibit heterogeneity in privacy choices when informed (Goldfarb & Tucker 2012*b*, Varian et al. 2005, Johnson et al. 2020). This heterogeneity poses a potential selection problem in the data shared by consumers. Selection in data and the resulting bias in data-driven insights have been in the spotlight, with examples spanning automatic resume sorting (Cowgill et al. 2020), medical research (Al-Shahi et al. 2005), and public opinion polling.[4]

To better understand consumers' privacy preferences and how they affect the selection pattern in voluntarily-shared data, I empirically distinguish between two preferences for protecting privacy. Privacy preferences can emerge because privacy itself is valued as an *intrinsic* right (Warren & Brandeis 1890). They can also arise as an *instrumental* value, the payoff of preventing their private "type" from being revealed through data (Stigler 1980, Posner 1981). Consumers can hold both types of privacy preferences. Intrinsically, people find it creepy to have smart thermostats tracking their activities at home, regardless of whether their behaviors are benign or objectionable (Pew Research Center 2015). Instrumentally, risky drivers may avoid installing telematics devices that allow an insurance firm to monitor their driving habits (Jin & Vasserman 2018, Soleymanian et al. 2019).

Although the conceptual distinction of intrinsic and instrumental preferences dates back to Becker (1980), it has drawn little empirical attention thus far. I argue that empirically separating these preference components is crucial for two reasons. First, it allows us to understand how consumers self-select into sharing their data, and how this selection affect a firm's inferences about consumers and business decisions. The reasoning "if you've got nothing to hide, you've got nothing to fear" is only valid when consumers harbor a purely instrumental preference for privacy. On the other hand, assuming consumers value privacy solely intrinsically can lead to the misleading

[1] https://gdpr-info.eu/recitals/no-39/; https://gdpr-info.eu/recitals/no-32/.

[2] https://piwik.pro/blog/privacy-laws-around-globe/.

[3] See Section "Individual consumer rights" in the *Privacy and Data Protection Framework* proposed by the US Senate: https://www.democrats.senate.gov/imo/media/doc/Final_CMTE%20Privacy%20Principles_11.14.19.pdf

[4] https://www.nytimes.com/2019/07/02/upshot/online-polls-analyzing-reliability.html

conclusion that people who value privacy more are no different from the rest of the population. Incorrect inferences about consumers who protect their privacy can lead to suboptimal decisions, such as under- or over-targeting, or putting these consumers under excessive scrutiny.[5]

Second, empirically separating these two preferences also enables us to evaluate firms' data collection and analysis strategies by understanding how privacy choices respond to these strategies. While the intrinsic preference is a utility primitive, the instrumental preference arises endogenously from how the firm uses consumer data. As such, instrumental preferences can shift with the purpose of data collection, the performance of a firm's model used for processing the data, and what other data the firm already obtains. Separating the instrumental preference from the intrinsic allows us to model these shifts accordingly. Accounting for the endogenous nature of instrumental preferences allows us to better calculate the equilibrium impact of privacy regulations and firms' data utilization strategies.

In this paper, I show how consumers' dual privacy preferences determine the selection pattern in shared data, then evaluate data utilization methods that account for the selection bias. I start by introducing a conceptual model that builds on Becker's (1980) seminal paper on two-dimensional privacy preferences. Using this model, I show the selection pattern in shared data is jointly determined by the heterogeneity of and correlation between the two preference components. In particular, the prediction offered by classical economic models of disclosure (i.e., high types self-select into sharing data, which is the reasoning behind the "nothing to hide" argument) can be rejected when the intrinsic preferences are relatively heterogeneous and negatively correlated with the instrumental preferences.

I design an experiment that measures revealed preferences for privacy in dollar terms and captures preference heterogeneity. Revealed preferences are solicited by requesting consumers to share sensitive data with a company. To capture the heterogeneity in privacy preferences and its impact on selection, I use a novel two-stage design, which sequentially records consumers' private types and their privacy choices. This design enables me to observe the contents of personal information, *even from consumers who choose not to share their personal data*. The experiment generates three layers of variation needed to identify my model: (a) the level of instrumental incentives that separates the two preference components, (b) the amount of compensation that enables me to calculate the dollar values of privacy preferences, (c) the default choice that permits comparison of privacy choices in opt-in versus opt-out policy regimes. The experiment also contains a conjoint survey, which allows me to counterfactually calculate the value of personal data in the context of price targeting.

I then estimate a structural model to quantify the heterogeneity of each preference component, taking into account that instrumental preference is endogenous. Intrinsic preferences are highly

[5] As an example, users of the TOR browser used to receive more CAPTCHA challenges than other Internet users, sometimes as much as 10 CAPTCHAs per session. This practice has later been discontinued. See https://www.zdnet.com/article/cloudflare-ends-captcha-challenges-for-tor-users/.

heterogeneous across both consumers and categories of data. Without instrumental preferences, consumers' mean willingness to accept (WTA) ranges from \$0.14 to \$2.37 across categories of data requested. However, consumers at the 95% quantile value each personal variable from \$2.19 to \$5.08. To obtain a representative set of data, a firm will need to pay as high as \$29.72 per consumer per demographic profile. For instrumental preferences, I find that consumers' beliefs in the instrumental outcome are roughly consistent with the actual payoff scheme. Such belief consistency is conditional on the information about data usage being provided in a plain and transparent manner, as is required in the majority of new privacy regulations. This finding implies that taking the information environment as fixed, the magnitude and heterogeneity of instrumental preferences will match the magnitude of the actual payoff scheme, even when the latter changes endogenously.

Taken together, the coexistence of intrinsic and instrumental preferences paints a more nuanced picture of the selection pattern. Within the experiment, intrinsic preferences play a dominant role in determining the direction of consumer self-selection, even when intrinsic and instrumental preferences have the same magnitude on average. This pattern is caused by a higher degree of heterogeneity in intrinsic preferences among consumers. As a result, high-type consumers can self-select *out of* rather than *into* data sharing when the two preference components are negatively correlated.

In the counterfactual analysis, I evaluate two strategies to improve a firm's collection and analysis of consumer data by addressing the privacy-induced selection: sampling and tagging non-sharing consumers. I examine both strategies in the context of price targeting using consumer data. At the data collection stage, by sampling a random subset of consumers and giving them high compensation for sharing, the firm opts for preserving data representativeness at the expense of volume. Prioritizing representativeness makes the most sense when the firm does not know what type of consumers self-select out of sharing, as is often the case when consumers have dual privacy preferences. I find that this strategy is more useful when data exhibits information externality, in particular when data are used for training a model applied to all consumers. By contrast, this strategy does little to improve the efficiency of data collection when data are used for customer profiling, where information externality is absent.

At the data analysis stage, tagging non-sharing consumers means characterizing them using a privacy-choice indicator when building consumer behavior models. This adjustment allows the model to pick up the difference between the sharing and non-sharing consumer groups from the data itself, rather than relying on arbitrary assumptions. I find that this strategy is effective in de-biasing the targeting outcome caused by privacy-induced selection. On the other hand, the improvement in targeting precision can be slight when the difference within the non-sharing group is much larger than the between-group difference, which is more likely when intrinsic preferences are more heterogeneous.

I describe how the experiment can be replicated in the field, taking into account the institutional and informational constraints absent in my setting. By measuring consumers' intrinsic and instrumental preferences, firms can learn about the nature of selection in shared data, which allows them to design more efficient data acquisition strategies. The ability to repeatedly measure privacy preferences also means subsequent research can use the experiment to further explore how privacy preferences change across contexts.

This paper contributes to several strands of literature. First and foremost, my paper formalizes and extends Becker's (1980) dual-privacy-preference framework (also see Farrell 2012, Cooper 2017, Choi et al. 2019, and Tirole 2020 who adopt similar frameworks).Compared with these papers, mine documents how the coexistence of intrinsic and instrumental preferences determines the selection pattern in data shared by consumers, and how it subsequently affects the quality of data-driven decisions. In doing so, my paper builds the link between consumers' privacy preferences and the quality of consumer data as firms' input.

Second, my paper builds on existing work that measures revealed privacy preference, including Goldfarb & Tucker (2012*b*), Athey et al. (2017), Kummer & Schulte (2019), as well as Acquisti et al. (2013) and Tang (2019) who provide dollar-value measures. Compared with these papers, mine separates intrinsic and instrumental preferences. Given the endogenous nature of instrumental preference, separating these two components is useful for characterizing equilibrium privacy choices and market outcomes when evaluating new privacy regulations or firms' data utilization strategies.

My paper also contributes to the literature on context-dependent privacy preferences by highlighting how instrumental preference responds to changes in (perceived) economic consequences of sharing, such as entities that have data access (Martin & Nissenbaum 2016) and information that changes consumer belief on data usage (John et al. 2010, Athey et al. 2017, Miller & Tucker 2017). As such, it complements the previous literature (Egelman et al. 2009, Acquisti et al. 2012, 2013, Adjerid et al. 2019, Lee 2019), which emphasizes psychological factors that generate context dependence.

Lastly, by discussing how consumers' privacy choices affect firms' inferences and resultant profits in the new policy regime, my paper adds to the research on how privacy regulations influence firms' managerial outcomes, including the effectiveness of advertising (Goldfarb & Tucker 2011, Tucker 2014), funds raised (Burtch et al. 2015), innovation activities (Goldfarb & Tucker 2012*a*, Adjerid et al. 2015), and profits (Johnson 2013, Goldberg et al. 2019, Aridor et al. 2020, Johnson et al. 2020, Batikas et al. 2020). Instead of examining the holistic impact of a particular regulation, my paper focuses on one mechanism: how consumers' self-selection into data sharing affects the quality of firms' data-driven decisions. In doing so, I am able to evaluate strategies that allow firms to address the impacts of selection.

## 2 The Conceptual Framework

In this section, I use a stylized model to clarify the distinction between intrinsic and instrumental preferences for privacy. It describes how the instrumental preference emerges endogenously from the way consumer data is used by the firm, how consumers self-select into sharing their data, and how this selection pattern differs from predictions generated by models that assume monolithic privacy preference.

To illustrate the key elements, for now I assume consumers have rational expectation, and use the same notation to represent the actual payoff and the payoff perceived by consumers. I also assume the firm does not have other information about consumers before requesting their data. In Section 2.4, I discuss how consumers' privacy preferences may change when these two assumptions are relaxed, and show the core prediction of the model remains valid. Both assumptions are relaxed in the empirical analysis.

### 2.1 Setup

Consider a firm that sells a product to many consumers. Consumers have different *types*, which calls for customized offers; denote consumer $i$'s type as $d_i$. The firm requests personal data from consumers in order to know their types. At a later stage, the firm gives customized offer $T(.)$ to consumer $i$, which maximizes the firm's expected profits conditional on *the firm's understanding of i's type*. For example, $T(.)$ can be price discount while $d$ is price sensitivity; or $T(.)$ can be the annual limit in an insurance contract while $d$ is risk type. To encourage data sharing, the firm may incentivize consumers using compensation, denoted as $P$. Examples of compensation include perks offered to consumers who sign up for a loyalty program, or gift cards for sharing email. While $T(.)$ may be a customized price for products, $P$ is the price for data common across consumers.

Consumer $i$ owns personal data that can reveal their type. Assume a one-to-one mapping exists between the content of personal data and a consumer's type.[6] We can always define consumer types such that the offer $T(d)$ is monotonic in $d$. For example, suppose $d$ is age and the middle-aged group has the lowest price sensitivity, followed by the older and then the youngest. Then we can label the middle-aged group as $d = 1$, the older group as $d = 2$, and the youngest group as $d = 3$. I define consumer types such that $T(.)$ is increasing in $d$, and refer to consumers with higher $d$ (who can obtain higher payoffs upon revealing their type) as the high type.

Consumers decide whether to share their data with the firm. $s_i \in \{0, 1\}$ indicates whether $i$ shares $d_i$: $s_i = 1$ means the data are shared. For consumers who share no information, the firm

[6] One way to understand the one-to-one mapping assumption is the following. In cases where the data indicates a range that the consumer's valuation lies in, we can define this range as his type. In cases where two levels of variable $d$ correspond to the same product valuation, we can code the two levels as having the same value.

forms beliefs about their types and chooses the amount of offer accordingly: $T(s=0) = T(F_d(d|s=0))$, where $F_d(d|s=0)$ is the distribution of consumer type conditional on the consumer choosing not to share his data. For consumers who share data, the offer is conditional on the content of the data, written as $T(d_i)$.

### 2.2 Privacy Preferences

A consumer has an *intrinsic preference* for privacy $c_i$, which is a taste for protecting his data *regardless of* the economic consequences induced by revealing his type. He also has an *instrumental preference* for privacy, the *expected* economic gain from not revealing his type:

$$\Delta T(d_i) \equiv T(F_d(d|s=0)) - T(d_i).$$

For example, suppose $T(.)$ is price discount and $d$ is age. If younger consumers have higher price sensitivity, the firm will give them higher discounts upon learning their age. Anticipating this outcome, younger consumers will have lower instrumental preferences.

The key distinction between intrinsic and instrumental preferences is whether they are *induced by* the consequences of revealing one's private information to the firm. The intrinsic preference is a utility primitive: It represents a cultural intuition not directly connected to the intended usage of data, and persists regardless of the consumer's "type" relevant to this market.[7] By contrast, the instrumental preference is endogenously driven by how the firm uses data to deliver targeted payoff; thus, it changes with the payoff function $T(.)$ as well as his type in this particular market $d_i$. The intrinsic preference can also be correlated with a consumer's type. However, his instrumental preference for privacy changes with the (perceived) usage of data $T(.)$, such as the purpose of data collection and the technology used for processing data; his intrinsic preference does not.

Instrumental preference and the utility from compensation are also distinct constructs, even though both are derived from payoffs. The instrumental motive reflects the value of private information. It is a function of the hidden type that *the firm cares about*, that is, information about the consumer that can help the firm refine the optimal offer. On the other hand, compensation is the same across different types of consumers, because it is the "price for data" offered before the firm learns about consumers via the data.

### 2.3 Who Chooses Not to Share Personal Data?

A consumer shares data iff the privacy cost is offset by the compensation that the firm provides:

$$s_i = 1 \text{ iff } -c_i - \Delta T(d_i) + P > 0. \tag{1}$$

[7]Intrinsic preference may take the form of a pervasive and nebulous concern about data misuse, such as concern for identity theft. However, as long as such concern is not tied to the intended data usage, it is considered intrinsic.

The firm wants to learn about consumers who choose not to share data in order to give them the offer amount that maximizes profits. A model that assumes privacy preferences are purely instrumental will generate the following prediction: Only low types choose to withhold their data in equilibrium, because these are the consumers who incur a larger loss upon sharing data (Grossman & Hart 1980, Milgrom 1981). This reasoning is the underpinning of the "nothing to hide" statement. Alternatively, a theory that assumes privacy preferences are pure intrinsic may fail to capture the nuance of consumers' self-selection into sharing.

The dual-preference framework paints a more nuanced picture of how consumers self-select into sharing personal data. The intrinsic preferences for privacy are likely to be heterogeneous among consumers. This heterogeneity should change the firm's inference $F_d(d|s = 0)$ because nondisclosure no longer signals low-type customers. The degree to which privacy choice reveals information about a consumer's type depends on both the relative heterogeneity of the intrinsic preference and its correlation with the instrumental preference. This is formally characterized by the proposition below (see proof in Appendix A).

**Proposition 1.** *Denote the standard deviation of intrinsic and instrumental preferences as $\sigma_c$ and $\sigma_t$, respectively, and their correlation coefficient as $\rho$. The following conclusions hold:*

*(A) In data shared with the firm, sample selection goes in the same direction as predicted by a model with pure instrumental preference iff $\rho + \frac{\sigma_t}{\sigma_c} > 0$.*

*(B) Privacy choice is more indicative of a consumer's type $d_i$ when $\frac{\sigma_t}{\sigma_c}$ is higher.*

To illustrate this proposition, suppose older consumers (who would have obtained better discounts upon sharing their age) care more about privacy intrinsically, and that the intrinsic preference is highly heterogeneous across age cohorts compared with the instrumental. Then the intrinsic preference will play a dominant role in privacy decisions: On average, consumers who choose not to share their data are more senior and should receive more generous discounts. This pattern forms a stark contrast to the case with a pure instrumental preference for privacy.

In sum, the dual presence of intrinsic and instrumental privacy preferences has two main implications. First, although the intrinsic preference is a utility primitive, the instrumental preference is endogenously determined by the market environment. This fact explains why preferences for privacy vary across the contexts of data used, who gets access to the data, and what data are requested. Second, when the intrinsic preference for privacy is heterogeneous, privacy choice no longer unambiguously signals a specific type of customer. The more heterogeneous the intrinsic preference is relative to the instrumental, the less we can assume a consumer's type based on his privacy decisions. Accounting for this fact is essential for analyses based on voluntarily contributed personal data.

### 2.4 Extending the Stylized Model for Empirical Analysis

Consumer behaviors and business data collection practices in the real world may deviate from the assumptions in the stylized model in two aspects. Below, I discuss how these deviations can change the magnitude of privacy preference components and the resulting selection pattern, and show how my empirical analysis accounts for them. I also discuss a type of instrumental preference not featured in the stylized model, and show that it plays the same role as the other preference components in affecting the selection pattern.

**Consumers' imperfect expectation.** The instrumental preference comes from the difference between two belief components: $T(d_i)$, indicating how much consumers' payoff depends on their private information revealed through shared data; and $T(F_d(d|s = 0))$, indicating what the firm infers about consumers who decline to share and the associated payoff. In reality, consumers may or may not correctly infer either of them. Their expectation on $T(d_i)$ determines the degree of heterogeneity in instrumental preferences, and in turn, the degree of self-selection induced by it. Their expectation on $T(F_d(d|s = 0))$ changes the level of instrumental preference: If they believe the firm holds the "nothing to hide" belief, their expected payoff from not sharing becomes lower; hence, the instrumental preference for withholding data also becomes lower. However, a pure level shift of instrumental preference does not change the degree of selection induced, provided that the level shift is the same across consumers.

We can think of expectation about $T(d_i)$ as the *first-order belief*, and expectation about $T(F_d(d|s = 0))$ as the *higher-order belief*, which requires expectation about the firm's as well as other consumers' belief and the latter's privacy preference distribution. In the empirical analysis, I directly measure consumers' first-order and higher-order beliefs. Since my goal is to understand the selection pattern, I focus on examining whether consumers have the correct first-order belief.

**Firm's existing knowledge about consumers.** The availability of existing information about individual consumers can decrease the marginal value of additional personal data. Once consumers anticipate this decrease, their instrumental preference will shrink, meaning both the level and heterogeneity will decrease. Nevertheless, as long as the requested data still reveal additional private information, the instrumental preference will persist. The firm's existing knowledge about consumers does not change their intrinsic preference. Therefore, when the firm has pre-existing information, intrinsic preference is more likely to play a major role in shaping the selection pattern in shared data.

**Data that improve the horizontal match value.** To study selection, my paper focuses specifically on the instrumental outcome (and the associated preference) that distinguishes high- versus low-type consumers when using the data to target consumers. There are cases where the level of instrumental outcome does not vary with the vertical type, such as targeted advertising that matches consumers and products. This form of instrumental preference will affect the selection and firm inference in a way that resembles either the intrinsic preference or the compensation, depending on whether the match values that consumers experience are observable. Such instrumental preference is still endogenous to the firm's data usage strategy.

# 3 The Experiment

The goal of the experiment is twofold. It needs to provide variation to empirically separate intrinsic and instrumental preferences, as well as variation that allows me to quantify them in dollar terms. It also needs to record both privacy choices and the contents of private data, so that I can characterize privacy preference heterogeneity and the selection pattern in shared data. In this section, I explain the empirical challenges for achieving these goals and how my design addresses the challenges. I also describe how the choice environment is set up to match key elements in the new privacy regulatory regime.

## 3.1 Empirical Challenges and Solutions

Empirically separating intrinsic and instrumental preferences is difficult. First, the economic incentive is usually fixed in observational settings, making it infeasible to separate instrumental preferences from the intrinsic. Second, in most observational settings, the request for personal data is bundled with product provision. As a result, the preferences for privacy are confounded with the preferences for products concurrently offered. For example, consumers may keep using Gmail even after learning that Google analyzes all their email texts, due to either their low preferences for privacy or their high valuation of Gmail service. Lastly, both consumers' privacy choices and their private types need to be observed to identify my model, yet privacy choices are precisely the decision concerning whether to reveal these private types. If this challenge is not accounted for, the collected data will exhibit self-selection as long as variation in privacy decisions exists.

My experiment includes three main features to circumvent these challenges. First, instrumental incentives are turned on or off across treatments. I can thereby measure the intrinsic preferences directly when the instrumental incentives are off, and use the difference between treatments to measure the instrumental preferences. Second, I exclude the confound from product preference by using monetary incentives (which have known values) to compensate for data sharing. Furthermore, the amount of compensation to encourage data sharing varies across treatments, allowing me to measure the dollar values of privacy preferences. To overcome the last challenge, I adopt a two-stage design, where the first stage collects participants' private information, and the second stage solicits revealed preferences for privacy.

### 3.2 Examination of the New Policy Regime

Measuring privacy preferences in a relevant choice environment is important given their context-dependent nature. To this end, my experiment specifies a choice environment that features key elements common in recent privacy regulations and principles, described below.

**Transparency of data usage.** Recent privacy regulations and principles require firms to deliver plain and accessible information about data collection and its purpose. For example, both GDPR and CCPA require data controllers and processors to use clear and plain language to describe the purpose of data processing and consumer rights.[8] To match this element, my experiment explains clearly the usage and flow of the data, and explicitly notifies participants about their options related to data sharing.

**Consumer control and consent.** This is represented by various rights clauses in major regulations, such as *the right to know*, *the right to deletion*, *the right of access*, *and the right of data portability*.[9] One key component in these rights clauses is the *explicit consent* requirement, which is implemented differently across regulations in terms of the default action. For example, EU laws such as GDPR and ePrivacy Regulation require opt-in consent, while the US adopts a mixed approach, with opt-in consent required for sensitive data and for data sales.[10] Regardless of the regulatory requirement, requests effectively operate in an opt-in condition for data that are not generated or tracked by default, such as survey responses or test results. My experiment includes both opt-in and opt-out conditions, but the empirical analysis focuses on the opt-in condition given its more prominent empirical relevance. I compare privacy choices in different consent regimes in Appendix G.1.

### 3.3 Experiment Design

The experiment uses a survey as an instrument to solicit revealed preference. This is achieved by including personal questions with varying degrees of sensitivity: A participant's decision to share the response to a question indicates his level of privacy cost associated with this personal variable. This technique has been deployed by Acquisti et al. (2012) and Goldfarb & Tucker (2012*b*). Research shows that in the domain of privacy preferences, attitude- and behavior-based measures often disagree (Harper & Singleton 2001, Spiekermann et al. 2001). I focus on revealed preference because it is not only incentive compatible, but also more relevant than attitudes for managerial decisions and policy analysis. In addition, I avoid using a Becker-DeGroot-Marschak mechanism, which is shown to produce results closer to stated attitude than revealed preference when used for measuring privacy preferences (Benndorf & Normann 2018).

[8] https://gdpr-info.eu/recitals/no-39/; https://oag.ca.gov/privacy/ccpa.

[9] See https://www.bakerlaw.com/webfiles/Privacy/2018/Articles/CCPA-GDPR-Chart.pdf.

[10] Examples include Illinois' *Biometric Information Privacy Act*, the *Fair Credit Reporting Act* on medical data, *CCPA* on the sale of personal information and on offering financial incentives to encourage sharing, and the *Washington Privacy Act* for biometrics, geolocation, and other sensitive data.

The experiment consists of two stages. In stage one, participants see themselves participating in a market research survey sent by the University of Chicago. The survey includes conjoint questions about smartwatch attributes and about participants' *intent to purchase* a digital device in the near future. They are followed by demographic questions, including *gender, age, education level, income, relationship status, whether they have children, zip code*, and *ethnicity*. Each personal question in the first stage includes a "prefer not to say" option; people who find the question too sensitive are thus allowed not to respond rather than being forced to fabricate a response. Up to the end of the first stage, consumers are unaware that they will later be requested to share personal data with the firm, and thus not actively considering privacy.

Stage one serves two roles. The first is to record private information from consumers, including those choosing not to share data in the subsequent stage. This full information allows me to measure heterogeneity in privacy preferences and characterize how the interplay between intrinsic and instrumental motives determines selection in shared data. Second, the conjoint questions provide inputs for calculating the value of data to firms in a pricing context, which becomes the basis for evaluating the data collection and analysis strategies in my counterfactual analysis.

Stage two solicits privacy choices. After finishing the survey, participants navigate to a new screen. Here, they are requested to share survey responses with a third party, which is a smartwatch manufacturer that wants to use the data to inform its product-design decision. Participants can choose whether to share each personal variable separately via check boxes.[11] Data sharing is encouraged by compensation in the form of a gift-card lottery. Participants are not aware of the possibility of sharing data with the third-party until they answer all questions in stage one. Once they reach the second stage, the "return" button is disabled, preventing them from deliberately changing previous responses to facilitate sharing. These two features, along with the "prefer not to say" option in Stage one, are included to ensure responses in the first stage are truthful.

Stage two is also where all treatments take place. Figure 1 displays the three layers of treatments: the incentive scheme, the amount of compensation, and the sharing default. These treatments are orthogonal to each other.[12] The first treatment layer varies the incentive scheme:

- Treatment 1 (compensation): The amount of compensation increases proportionally to the amount of data shared and is common across all participants. In particular, sharing one additional personal variable increases the probability of winning the gift card by one percentage point. In other words, the price for data is the same regardless of what the firm learns about the consumer.

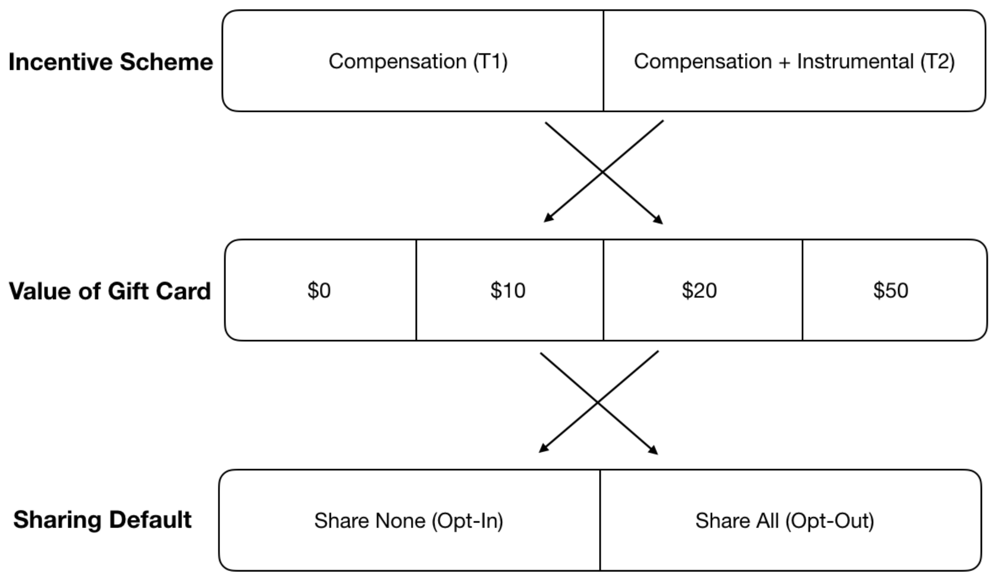
- Treatment 2 (compensation + instrumental incentive): A baseline level of compensation exists and takes the same form as in Treatment 1. The baseline amount is then adjusted based on whether *the company perceives* the participant to be a potential customer *from analyzing the data*

[11]Only informative responses (i.e. other than "prefer not to say") in Stage one are allowed to be shared in Stage two.

[12]One exception is that by design, participants who receive zero compensation do not receive any instrumental incentives.

*it obtains*. Likely customers receive higher compensation than the baseline, whereas unlikely customers get a cut in the compensated amount. Participants are told the company's target customers are high-income people who intend to buy a digital product, and therefore, they will receive more *if the shared data indicate* they fit this profile.

Figure 1: Treatment Design




*Note:* The three layers of treatments are orthogonal to each other. Treatments are assigned with equal probability in each layer.

Appendix B displays the information shown in each treatment. Overall, the incentive scheme is presented in a transparent and clear manner. The incentive scheme is displayed in two parts. The main page explains who collects the data and for what purpose, and how a participant's payoff will qualitatively depend on the data shared. The detailed screen shows quantitatively how the payment is calculated, and is displayed when a participant clicks the "see details" link. This design is similar to the format of most post-GDPR website banners.

In sum, *privacy choices in Treatment 1 alone identify intrinsic privacy preferences*. Here, the stated purpose of data collection does not imply continuous tracking or any other future interactions with consumers. Moreover, participants did not know about this company prior to entering the experiment; thus they are unlikely to anticipate the instrumental consequences of sharing data from interacting with the firm in the future. By contrast, choices in Treatment 2 are motivated by both intrinsic and instrumental preferences. The instrumental preferences are induced by an incentive scheme that depends on a participant's income and product-purchase intent. These two characteristics constitute a consumer's "type" in this experiment. Therefore, the *differential responses between Treatments 1 and 2 identify instrumental preferences for privacy*.

The other treatments are designed as follows. The second treatment layer changes the value of the gift card (essentially cash) across participants, creating variations for measuring the dollar values of privacy preferences. The third layer varies default choice, which is set to either sharing all data (opt-out) or sharing none (opt-in). Within each layer, treatments are assigned with equal probability.

To measure if participants understand and trust the validity of incentive treatments, I send follow-up questions to participants after they make the data-sharing choices. These questions include the perceived purpose of the study, what determines the amount of expected compensation, the reasons they choose (not) to share the survey responses, and if they prefer a sure reward with the same expected value as the gift-card lottery.

### 3.4 Discussion

Using a controlled field experiment allows me to design a control group that measures intrinsic preference in a relatively clean manner. In a real business setting where consumers already know the firm, they are likely to fixate their expectations on how the firm usually uses consumer data, and thus always have some instrumental preference. In Section 7, I show how firms can run a different version of my experiment in the field to decompose the two preference components, by leveraging an assumption on consumers' belief stability.

The experiment intentionally uses type-dependent monetary compensation instead of personalized product prices to induce the instrumental incentive. Although the latter is more natural, it may not induce variations of instrumental preference in my setup. Given that participants have never interacted with the featured company (it is fictitious), they may not plan to engage in future transactions with this company. In this case, the firm's pricing practices will not matter to them.

Using a lottery instead of sure rewards for compensation may bias preference measurement if participants predominantly have the same direction of risk preference. If participants are risk averse, their perceived gain from the gift-card lottery will be lower than its objective expected value, and the estimated dollar value of privacy preferences will be an upper bound of their true valuation; the opposite holds if participants are risk-seeking. In the follow-up survey question, 35% of the participants prefer the lottery, while the rest prefer the sure reward. This pattern suggests that consumers' risk preference distribution is more balanced and thus less likely to systematically bias the measurement result.

Consistent with the conceptual framework, the experiment focuses on the case in which consumers cannot send fake information to the firm. Cases abound where consumers' personal data are truthfully recorded as long as they choose to share, such as location tracking, browsing history tracking, and genetic testing. In some cases, fabricating information is technically possible but involves a high cost, and is usually adopted by only the most tech-savvy consumers. One can extend this framework by introducing heterogeneous costs of data fabrication as the third dimension in consumers' preferences. The measurement results in this paper serve as a useful building block for such extensions.

# 4 Data and Descriptive Evidence

In what follows, I describe the data source and sample characteristics, and then present model-free patterns of intrinsic and instrumental preferences. The main analysis focuses on privacy choices in the opt-in regime. Data show how consumers purposefully share some data while protecting others, how the instrumental incentive changes the composition of consumers that share data, and how this compositional shift changes the quality of data shared.

## 4.1 Data Source and Cleaning

Participants of the experiment come from Qualtrics Panels. To the extent that Qualtrics panel members may be more willing to share personal information without anticipating any instrumental consequences, my measurement result provides a lower bound for the population-level intrinsic preferences. Nevertheless, existing work finds the Qualtrics panel is more representative than alternative online panels (Heen et al. 2014, Boas et al. 2018). To further reduce possible discrepancies, I apply stratified sampling so that the demographics of participants entering the survey resemble the distribution given by the 2018 US Census. Qualtrics provides three demographic variables on the back end, including income, age, and ethnicity. I use these data to validate the truthfulness of responses in the first stage. Not all demographic variables I intend to collect are available through Qualtrics. Therefore, having the first stage is still necessary.

A total of 4,142 participants enter the survey; 3,406 of them proceed to the data-sharing-request stage. For people who leave the survey upon seeing the request, I code their choices as sharing nothing, regardless of the default condition. Figure C.1 shows the participant attrition throughout the experiment. Among the 18.4% of participants who leave the survey before seeing the treatment, 91% exit before or during the conjoint survey. This pattern indicates that attrition is mainly caused by a lack of interest in the conjoint questions rather than a reluctance to share personal data in the first stage. To prevent treatment contamination, I deduplicate the respondents by IP address. I also exclude respondents whose time spent on the survey, or time spent in responding to the data-sharing request, is at the lowest decile. The cleaned data include 2,583 participants.

## 4.2 Sample Characteristics

Attrition and sample cleaning can change the characteristics of the final sample. Table 1 summarizes the demographics of survey participants in the cleaned sample, and compares them with the 2018

Current Population Survey (CPS) whenever similar statistics are available. Some discrepancies come from differences in counting. For example, the mean age provided by CPS includes juniors (ages 15–18), whereas my sample contains only adults; "black" in my sample includes mixed-race groups, while CPS's definition excludes it. Another difference comes from the fact that some participants choose not to share all demographics during the first stage. As a result, the percentages of different income levels do not sum up to 1, whereas in the census, the disclosure is complete. Compared with the population, participants who finish the survey tend to be female, less educated, and have lower incomes.

Table 1: Demographics of Experiment Participants (Cleaned Sample)

| | Variables | Experiment Sample | 2018 Census |
|---|---|---|---|
| | Female | 65.31% | 50.80% |
| | Married | 47.39% | 51.16% |
| | Have young kids | 24.78% | – |
| | Mean age | 47.60 (16.89) | 45.9 (–) |
| Education | High school degree or less | 47.00% | 39.93% |
| | College degree | 40.65% | 48.67% |
| | Master's degree or higher | 11.39% | 11.40% |
| Race | White | 71.27% | 76.60% |
| | Black | 15.37% | 13.40% |
| Annual Household Income | $25,000 or less | 21.99% | 20.23% |
| | $25,000 to $50,000 | 29.54% | 21.55% |
| | $50,000 to $100,000 | 30.12% | 28.97% |
| | $100,000 or more | 13.55% | 29.25% |
| No. Observations | | 2,583 | – |

*Source of the census data*: U.S. Census Bureau, Current Population Survey, 2018 Annual Social and Economic Supplement. "–" indicates that no corresponding statistics are available.
*Note:* For discrete variables, values in the survey are collapsed into larger groups to facilitate the exhibition. Numbers corresponding to the same category may not sum to 1, given that smaller groups are left out and that some participants choose not to respond in the first stage. For continuous variables, mean values are reported with standard deviation in parenthesis.

*Purchase intent* is one of the consumer types in the instrumental-incentive treatment. It is calculated based on participants' responses to two questions in the first stage: (A) "How likely will you buy a new smartwatch within the next 3 months?" (B) "How likely will you buy any other digital devices within the next 3 months?" Each question uses a 5-point Likert scale. Different answers are then given different scores. For example, "extremely likely" is scored 2, while "extremely unlikely" is scored -2. Purchase intent is then constructed by summing up these two scores; a higher value indicates higher purchase intent. Across participants, the mean purchase-intent score is -0.17, with a standard deviation of 1.72.

### 4.3 Intrinsic Preferences

Table 2 shows how the frequency of sharing varies with compensation and the category of personal data in Treatment 1 (intrinsic preference only). Consumers do not want to share personal data when not compensated: In the first column, the frequencies of data being shared are all at or below 50%, which is the indifference benchmark.

Table 2: Frequency of Data Sharing with Intrinsic Utility

| | Category of Data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Compensation | Gender | Age | Education | Income | Relationship | Kids | Zip | Race | Purchase Intent |
| = 0 | 0.50 | 0.47 | 0.43 | 0.36 | 0.46 | 0.29 | 0.41 | 0.42 | 0.43 |
| > 0 | 0.70 | 0.68 | 0.62 | 0.56 | 0.66 | 0.53 | 0.63 | 0.63 | 0.54 |

*Note:* "Relationship" corresponds to their responses about marital status. "Kids" corresponds to responses to the number of children they have. Among the compensated groups, the value of gift card is $33 on average, with a 1% increase in the possibility of winning for each variable shared.

Compensation is effective in shifting privacy decisions. An average price of 33 cents per variable increases the probability of sharing by about 20% across variables. However, this average response among participants masks preference heterogeneity, which is crucial for understanding the impact of privacy decisions on the selection in shared data. I revisit preference heterogeneity in the estimation results section.

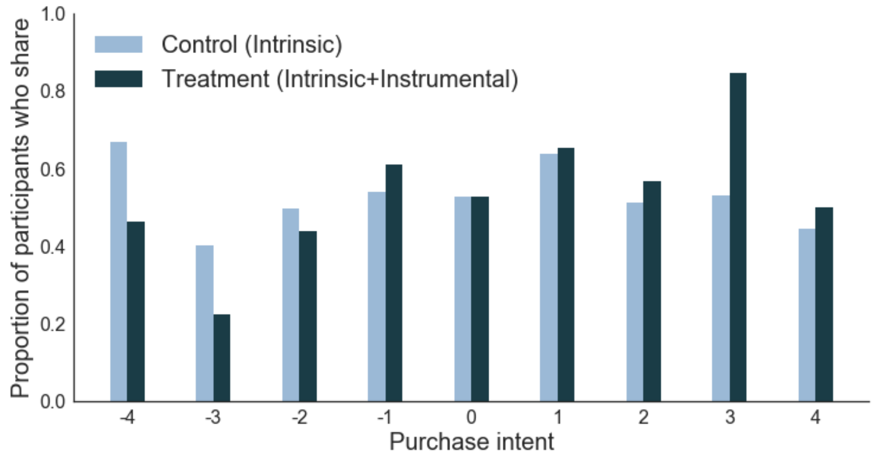
Different data are valued quite differently, and the sensitivity ranking across personal variables remains largely unperturbed regardless of whether data sharing is compensated. Data about household income and about their children are valued the most, whereas gender is viewed as the least sensitive. Overall, the table shows that participants make attentive trade-offs in the experiment, and that different data are valued differently by consumers.

### 4.4 Instrumental Preferences

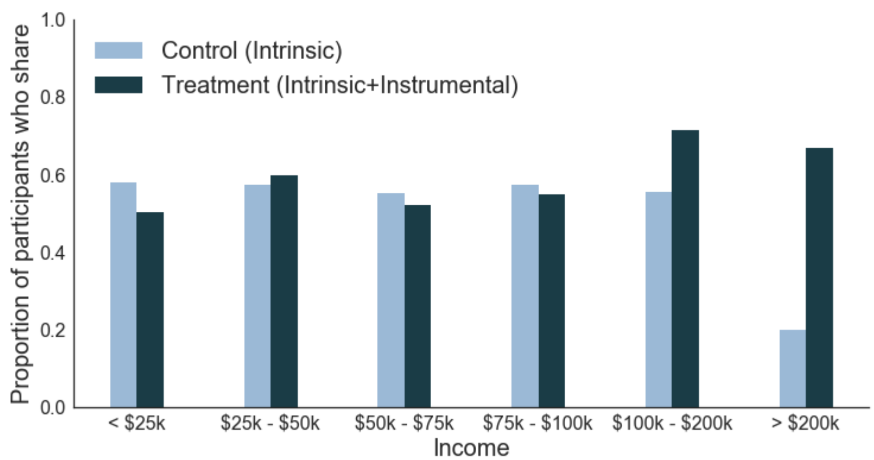
Treatment 2 introduces the instrumental incentive: Participants benefit more if they are perceived as wealthy or intend to buy digital products in the short term (hereafter *high types*). Figure 2 shows how instrumental incentives influence privacy choices and how this influence is moderated by intrinsic motives. Panel (a) plots the proportion of participants choosing to share their purchase intent data across purchase intent cohorts for each incentive treatment. High-type consumers are more willing to share personal data in Treatment 2 than in Treatment 1, whereas the reverse pattern occurs for low-type consumers. This pattern indicates that participants are attentive to the instrumental incentives.

Figure 2: Frequency of Data Sharing across Incentive Treatments

(a) Purchase-Intent Sharing




(b) Income Sharing




*Note:* Frequency is calculated as the proportion of participants who share their income data within each income cohort (not across). The sum of bar heights can be greater than 1.

Panel (b) shows the same plots for the income sharing decision. Here, the behavioral differences between the treatment and control groups are overall insignificant. This lack of behavioral difference may be caused by a greater heterogeneity in intrinsic preferences, which makes the utility variation caused by instrumental preference zoom smaller when translated to choice variation. Interestingly, wealthier participants have stronger intrinsic preferences for privacy than their low-income counterparts, which is opposite to the direction that instrumental preferences indicate.

## 4.5 Dual Privacy Preferences and the Selection in Shared Data

To further examine how the two privacy preferences affect the distribution of data shared, I compare the mean purchase intent and income between the *shared* (data reflecting Stage two sharing decisions) and the *true* data (all data collected in Stage one), separately for each treatment group. Table 3 displays the t-test statistics for this comparison. With purchase intent, the existence of instrumental incentive makes the shared data feature more high-types than the true data has (see column 2 of Panel (a)); the difference between the shared and true data is marginally significant at the 0.06 level. This selection pattern is consistent with prediction offered by the classical economic model, due to the fact that with purchase intent sharing, intrinsic preferences are largely homogeneous among different types.

By contrast, Panel (b) shows the instrumental preference does not cause a significant selection among the shared data. This is because the intrinsic preference for sharing income data is both more heterogeneous and negatively correlated with the instrumental incentive: Wealthier participants have stronger intrinsic preferences for privacy than their low-income counterparts. Taking the messages together, the joint distribution of the two preference components determines the final selection pattern in the shared data.

Table 3: t-Test for Equal Means ($H_1$: E[D | shared] - E[D | true] $\neq$ 0)

(a) Purchase Intent

| | Control | Treatment |
|---|---|---|
| t-statistic | 0.190 | 1.847 |
| p-value | 0.849 | 0.065 |

(b) Income

| | Control | Treatment |
|---|---|---|
| t-statistic | −0.969 | 1.053 |
| p-value | 0.333 | 0.293 |

*Note:* Control = Intrinsic Utility; Treatment = Intrinsic + Instrumental Utility. Shared data are constructed based on consumers' decisions in the second stage as to whether to share their data with the firm; true data refers to all data collected from the first stage.

## 5 The Structural Model

The structural model serves three purposes. First, it estimates the dollar value of privacy preferences. The dollar value facilitates the translation of consumers' privacy preferences to the firm's costs of buying consumer data. Second, it clarifies how instrumental incentives shift privacy choices by changing consumers' beliefs about payoffs. While the instrumental incentive is endogenous, consumers' ability to anticipate the economic consequences of revealing private information is the primitive for a given information environment. Lastly, it allows me to simulate privacy choices and evaluate the information value of shared data in counterfactual regimes where the firm's data utilization strategy becomes endogenous.

## 5.1 Setup

Consumer $i$ is endowed with a vector of personal data $D_i = [d_{i1}, d_{i2}, \dots d_{iK}]$; $d_{i1}$ is income, and $d_{i2}$ is purchase intent. His sharing decision is characterized by a vector with equal length $S_i$: Each entry is an indicator of whether the associated personal variable is shared. For example, $S_i = [0, 0, 1]$ means $i$ shares $d_{i3}$ but not $d_{i1}$ or $d_{i2}$. Sharing decision $S_i$ brings an intrinsic privacy cost, a type-induced payoff from sharing (if the consumer is in the instrumental treatment), baseline compensation, and a random utility shock:

$$U(S_i; C_i, D_i) = \sum_k - \underbrace{c_k(X) \cdot s_{ik}}_{\text{intrinsic preference}} + 1_{instr} \cdot 1_{k\in\{1,2\}} \cdot \underbrace{\beta \cdot p_i \cdot w_k \cdot \widehat{E}[d_{ik}|S_i, D_i]}_{\text{type-induced payoff}} + \underbrace{\beta \cdot p_i \cdot s_{ik}}_{\text{util from compensation}} + \epsilon_{ik}. \tag{2}$$

$C_i = [c_1, c_2, \dots c_K]$ is the intrinsic preference for privacy; each $c_k$ can be expanded as a function of observables $X$. $1_{instr}$ is the instrumental-treatment indicator. $1_{k\in\{1,2\}}$ selects the data-sharing decisions that are subject to the influence of instrumental incentives. $\beta$ is the marginal utility of monetary rewards. $p_i$ is the value of gift card multiplied by 1%. $w_k$ is the consumer's expected increase in the percentage winning probability for an adjacent, higher type; this is their *first-order belief*. Meanwhile, $\widehat{E}[.]$ is their *higher-order belief*: the consumer's expectation of the firm's expectation about his type. The baseline compensation is proportional to the amount of data shared, represented by $p_i \cdot s_{ik}$. Lastly, $\epsilon_{ik}$ is the random utility shock associated with choice $S$; $\epsilon_{i1}, \epsilon_{i2}, \dots \epsilon_{iK} \overset{iid}{\sim} TIEV$.

Belief about a consumer's type depends on not only the contents of shared data, but also potentially the *sharing decision* itself: $\widehat{E}[d_{ik}|s_{ik} = 1, D_i] = d_{ik}$, $\widehat{E}[d_{ik}|s_{ik} = 0, D_i] = \widetilde{d_k}(p_i)$. I let $\widetilde{d_k}(p_i) = \delta_{k0} + \delta_{k1} \cdot p_i$ to allow for different levels of rationality.[13] If both the firm and consumers are rational, they will expect that consumers who withhold data are more likely to consist of low types as the instrumental incentives increase, reflected by a positive $\delta_{k1}$. If agents form naive beliefs instead, $\delta_{k1}$ is zero.

The belief parameters $\{w_k, \delta_{k0}, \delta_{k1}\}$ represent the extent to which consumers understand the actual instrumental payoff, which reflects their attentiveness and degree of sophistication. Directly estimating them is useful, as consumers may fail to correctly anticipate actual data usages and sharing practices (Stutzman et al. 2013, Ben-Shahar & Chilton 2016, Athey et al. 2017). The experiment allows me to directly estimate consumer beliefs while fixing the instrumental payoff scheme. Later in the counterfactual when I endogenize the firm's data collection and analysis strategy, the belief estimates allow me to back out the magnitude of instrumental preference by combining it with the shifting payoff scheme.

[13] A fully parameterized $\widetilde{d_k}(p_i)$ will require strong assumptions on consumer expectation. These additional assumptions include consumer belief on privacy choices among other consumers and how they are correlated with consumer types, as well as their belief on the firm's reasoning. Imposing these assumptions does not make sense, given the premise of this paper is that even a sophisticated firm may not have the correct belief about non-sharing consumers.

Correctly estimating heterogeneity in intrinsic versus instrumental preferences is key to understanding how consumers self-select into sharing. I characterize heterogeneity by allowing preference parameters to be functions of observables $X$, including demographics, time entering the experiment, time spent on each question, browser used, and device specifications. In models that allow for heterogeneity in intrinsic preferences, $c_k(X) = c_{k0} + c_{kx} \cdot X$. $\delta_{k0}(X)$, $\delta_{k1}(X)$, and $\beta(X)$ are specified similarly, except that variables in $\delta_k$'s do not include income or purchase intent so that the model can remain identified. There is also a "built-in" heterogeneity in instrumental preference, coming from the fact that instrumental incentives vary with consumer types.

Apart from privacy preferences, psychological factors can also affect data sharing choices. First is the default frame. The literature has proposed different mechanisms underlying the stickiness to default, which implies different ways that the default frame and utility parameters interact with each other (Bernheim et al. 2015, Goswami & Urminsky 2016, Goldin & Reck 2018). To be agnostic about the mechanism, I estimate models separately for each default frame. The estimated parameters represent behavioral preferences under each frame, which are the relevant objects for analyzing firm-side implications of privacy choices. Section 6 focuses on the opt-in regime given the current regulatory focus; a comparison between behaviors in the two regimes can be found in Section G.1. The model also includes a behavioral response term, $m \cdot (p_i \geq 0) \cdot s_i$, to account for a combination of the mere-incentive effect and potential anchoring effects at the start of the survey. The estimation result and interpretation for this term can be found in Section G.2.

With the specification above, the log-likelihood can be written as the sum of log logit probabilities:

$$LL = \sum_{i=1}^{N} \sum_{k=1}^{K} s_{ik} \cdot (\Delta u_{ik}) - \ln(\exp(\Delta u_{ik}) + 1),$$

where $\Delta u_{ik}$ is the difference in mean utilities between sharing and not sharing data $k$, experienced by consumer $i$ (heterogeneity functions are omitted for the clarity of exposition):

$$\Delta u_{ik} = -\underbrace{c_k}_{\text{intrinsic preference}} - 1_{i,inst} \cdot 1_{k \in \{1,2\}} \cdot \underbrace{\beta \cdot p_i \cdot w_k \cdot \left[\delta_{k0} + \delta_{k1} \cdot p_i - d_{ik}\right]}_{\text{instrumental preference}} + \underbrace{\beta \cdot p_i}_{\text{util from compensation}} + m \cdot (p_i \geq 0). \tag{3}$$

## 5.2 Identification

Coefficients to be estimated include $c_k$, $w_k$, $\delta_{k0}$, $\delta_{k1}$ for $k \in \{1, 2\}$, $\beta$, and $m$. Parameters in $c_k$ are identified as the utility intercept of the participants who enter the intrinsic treatment; since treatment is randomly assigned, these coefficients are the intrinsic preferences shared by all participants. Belief parameters are identified from the instrumental treatment. $w_k$ is identified from how *different* types react differently to instrumental incentives. $\delta_{k0}$ and $\delta_{k1}$ are identified from responses to the instrumental incentives that are *common* across types. In particular, the identification of $\delta_{k1}$

comes from the interaction between the instrumental treatment and the amount of compensation. Parameter $\beta$ is identified through the variation in gift-card values. Given that $\beta \cdot p_i$ is linear, and that multiple gift-card values exist across treatments, $m$ is identified from the different responses to zero and non-zero incentives.

The key parameters in this model consist of the following: intrinsic preference, $c_k$; first-order belief about the instrumental consequence, $w_k$; and the sensitivity to income, $\beta$. Identification of these primitives allows me to construct consumers' privacy choices under different counterfactual scenarios. In particular, measuring the first-order belief is important, because it is this belief component that generates the adverse selection pattern created by instrumental incentives. To see this, note that $w_k$ scales the type-dependent payoff when a consumer chooses to share his data $\widehat{E}[d_{ik}|s_{ik} = 1, D_i]$. In comparison, the higher-order belief $\widehat{E}[d_{ik}|s_{ik} = 0]$ does not affect the selection pattern, given that it is not a function of the consumer's private information. Other parameters in the model are auxiliary: They provide flexibility and absorb confounding factors that may otherwise affect the key parameter estimates. For example, $\delta_{k0}$ and $\delta_{k1}$ may reflect not only consumers' higher-order belief, but also risk preferences that are common across types.

### 5.3 Estimation

I estimate the model under a Bayesian framework. A Bayesian model allows me to flexibly account for the heterogeneity and place theory-informed bounds on compensation sensitivity and belief parameters. I place the horseshoe prior for heterogeneity parameters (Carvalho et al. 2009), and a flat prior for the rest. Horseshoe is a form of continuous shrinkage prior; it accommodates the large number of parameters in the heterogeneity functions and avoids model over-fitting. Compared with other shrinkage priors such as Bayesian Lasso, Horseshoe yields estimates that are the closest to results from the Bayes Optimal Classifier. Intercepts of the heterogeneity functions are left unregularized to obtain unbiased estimates for the function mean. Due to regularization, the estimated heterogeneity will be smaller than the heterogeneity displayed in raw data. This is a necessary trade-off to avoid model overfitting.

I place non-negativity constraints on the sensitivity to compensation $\beta$, and bound constraints on $\delta$ such that they do not exceed the actual distribution support of consumer types. No sign constraints are placed on $c_k(X)$, thus allowing for the possibility that consumers have a "warm glow" in sharing insensitive data for improving research.

## 6 Estimation Results

Table 4 compares estimation results from models with different heterogeneity specifications. To compare model performance, I calculate the expected log predictive density (elpd) using the

Watanabe-Akaike information criterion (WAIC) approximation; a higher number indicates a better out-of-sample fit. Preference estimates are very different between the model without heterogeneity (Model 1) and the models that allow for heterogeneity in intrinsic preferences (Models 2 to 4). The latter exhibit better fits, as is demonstrated by higher elpd values. On the other hand, allowing for heterogeneity in belief or sensitivity to income does not improve out-of-sample fit: Estimation results are similar across Models 2, 3, and 4, and the elpd of Model 2 is the highest. Model 2 constitutes the basis for the main analysis.

Table 4: Intrinsic and Instrumental Preference for Privacy: Estimation Results Comparison

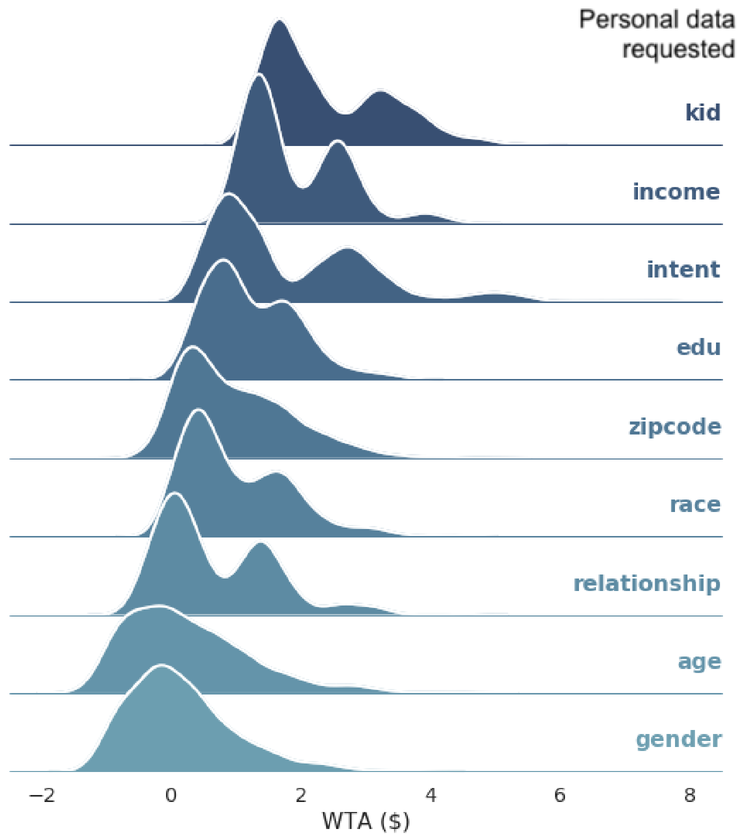
| | Model | 1. No Heterogeneity | | 2. Heterogeneous $c$ | | 3. Heterogeneous $c$ & $\delta$ | | 4. Heterogeneous $c$ & $\beta$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | mean | 95% CI | mean | 95% CI | mean | 95% CI | mean | 95% CI |
| intrinsic | $c_{income}$ | 0.57 | [0.43, 0.70] | 0.91 | [0.59, 1.32] | 0.93 | [0.58, 1.51] | 0.93 | [0.60, 1.39] |
| | $c_{intent}$ | 0.55 | [0.41, 0.70] | 0.83 | [0.42, 1.32] | 0.84 | [0.38, 1.38] | 0.87 | [0.41, 1.44] |
| | $c_{gender}$ | 0.02 | [-0.12, 0.15] | 0.19 | [-0.16, 0.66] | 0.24 | [-0.16, 0.95] | 0.20 | [-0.20, 0.75] |
| | $c_{age}$ | 0.06 | [-0.09, 0.20] | 0.26 | [-0.09, 0.73] | 0.29 | [-0.16, 0.91] | 0.28 | [-0.09, 0.82] |
| | $c_{education}$ | 0.37 | [0.23, 0.51] | 0.62 | [0.33, 1.05] | 0.65 | [0.29, 1.29] | 0.65 | [0.29, 1.24] |
| | $c_{relationship}$ | 0.20 | [0.06, 0.33] | 0.50 | [0.12, 1.01] | 0.55 | [0.11, 1.23] | 0.50 | [0.16, 1.04] |
| | $c_{kid}$ | 0.74 | [0.61 , 0.88] | 1.11 | [0.79, 1.46] | 1.09 | [0.71, 1.51] | 1.10 | [0.75, 1.55] |
| | $c_{zip}$ | 0.29 | [0.16, 0.43] | 0.56 | [0.23, 1.07] | 0.60 | [0.18, 1.22] | 0.61 | [0.19, 1.13] |
| | $c_{race}$ | 0.29 | [0.16, 0.42] | 0.60 | [0.29, 1.10] | 0.65 | [0.26, 1.26] | 0.65 | [0.30, 1.33] |
| instrumental | $w_{income}$ | 2.00 | [0.15, 3.87] | 2.12 | [0.11, 3.99] | 2.02 | [0.14, 3.92] | 1.90 | [0.04, 3.88] |
| | $w_{intent}$ | 2.63 | [1.07, 3.88] | 1.94 | [0.38, 3.76] | 1.97 | [0.29, 3.77] | 1.90 | [0.35, 3.70] |
| | $\widetilde{\delta}_{income,0}$ | 0.05 | [-0.19, 0.29] | 0.05 | [-0.19, 0.28] | 0.05 | [-0.19, 0.28] | 0.05 | [-0.19, 0.29] |
| | $\widetilde{\delta}_{income,1}$ | 0.05 | [-0.19, 0.29] | 0.04 | [-0.19, 0.28] | 0.05 | [-0.19, 0.29] | 0.04 | [-0.19, 0.28] |
| | $\widetilde{\delta}_{intent,0}$ | 0.08 | [-0.35, 0.39] | 0.06 | [-0.35, 0.38] | 0.07 | [-0.36, 0.38] | 0.06 | [-0.34, 0.39] |
| | $\widetilde{\delta}_{intent,1}$ | -0.05 | [-0.36, 0.31] | -0.05 | [-0.36, 0.32] | -0.05 | [-0.37, 0.31] | -0.04 | [-0.34, 0.32] |
| sensitivity to compensation | $\beta$ | 0.13 | [0.07, 0.21] | 0.15 | [0.07, 0.24] | 0.15 | [0.06, 0.24] | 0.15 | [0.07, 0.25] |
| log posterior | | -8015 | [-8022,-8010] | -7476 | [-7540,-7407] | -7433 | [-7501, -7352] | -7525 | [-7588, 7450] |
| elpd$_{WAIC}$ | | 6384 | [6358 , 6410] | 6460 | [6431 , 6489] | 6365 | [6337, 6394] | 6455 | [6427, 6484] |

*Note:* Variables are normalized using the Gelman method before estimation. Wherever heterogeneity is allowed, the table displays estimates on the intercept term only. The same seed is used for estimating different models. The model directly estimates $\widetilde{\delta_{ik}} \equiv \beta \cdot w_k \cdot \delta_{ik}$ instead of $\delta_{ik}$ for numerical stability. The distribution of $\delta_{ik}$ is later backed out from posterior draws.

## 6.1 Intrinsic Preferences

Figure 3 shows the predicted WTA heterogeneity distribution associated with intrinsic preferences (calculated as $\frac{c_k(X)}{\beta}$), separately for each personal variable requested. Table 5 summarizes the statistics corresponding to each distribution, and Table D.1 shows credible intervals associated with these estimates. Consumers' WTA are highly heterogeneous. The mean intrinsic preferences for sharing different personal variables range from \$0.14 for gender to \$2.37 for information about their children (in the follow-up survey question, many participants describe the request for information about their children as "irrelevant" and "improper"). In comparison, the 97.5% quantiles are more than twice as large as the mean valuations. The upper-tail values are worth attention, since these values are the prices that firms need to surpass to guarantee a representative dataset. For example, a data collector needs to pay $3.82 per customer for 97.5% of them to share their income data, and $5.08 per customer to get 97.5% of purchase-intent data. As a robustness check, Appendix E includes WTA estimates from Model 4, which allows for heterogeneity in both intrinsic utility and sensitivity to income. The WTA distribution is quantitatively and qualitatively similar to the main result produced by Model 2.

Figure 3: Posterior Predicted Density of WTA in Intrinsic Preference




I find that consumers with high intrinsic preferences overall tend to be male, non-white, low in digital product purchase intents, and less wealthy. In particular, Appendix F shows the bimodal pattern is mainly driven by the heterogeneity among different racial groups. Younger consumers are less willing to disclose their education history, but otherwise have similar levels of intrinsic preferences compared with their more senior counterparts.

Are these privacy-preferences high or low? One way to gauge the magnitude of intrinsic preferences is by calculating the WTA for a profile, which is essentially a bundle of different data. For example, if cookies used to identify online users are associated with different demographic variables examined above, the WTA for sharing the whole demographic profile will have a mean of $10.34 and a 97.5% quantile of $29.72. For more sensitive data such as browsing and location

Table 5: Posterior Predicted Distribution of WTA in Intrinsic Preference

| | mean | median | 2.5% | 97.5% |
|---|---|---|---|---|
| kid | 2.367 | 2.069 | 1.220 | 4.311 |
| income | 1.870 | 1.546 | 0.944 | 3.823 |
| intent | 1.825 | 1.352 | 0.398 | 5.078 |
| education | 1.228 | 1.051 | 0.228 | 2.845 |
| zipcode | 0.985 | 0.800 | -0.157 | 2.916 |
| race | 0.980 | 0.737 | -0.066 | 2.945 |
| relationship | 0.687 | 0.390 | -0.448 | 2.894 |
| age | 0.260 | 0.084 | -1.064 | 2.718 |
| gender | 0.142 | 0.006 | -1.043 | 2.187 |

*Note:* A wider spread of distribution indicates higher preference heterogeneity.

histories, the WTAs are possibly higher. Another way of sensing the magnitude of privacy preferences is by comparing them with the firm's willingness to pay for these data. This comparison is further discussed in Section 8.1, where I calculate the firm's valuation of personal data under different data acquisition strategies.

### 6.2 Beliefs that Generate Instrumental Preferences

Consumer belief on the type-dependent payoffs represents the degree of their attentiveness and sophistication. Taken as utility primitive, the first-order belief scales the magnitude of instrumental preference relative to the actual payoffs for a given information environment. In the experiment, a consumer whose type is one tier above can increase his probability of winning by 2 percentage points if he discloses his type to the firm. Mapped to the model, it means consumers' first-order beliefs are accurate if $w$ equals 2. Column 2 of Table 4 shows that consumers' beliefs about $w_{income}$ and $w_{intent}$ are correct on average. This result implies the magnitude of instrumental preferences will match the actual payoff even when the latter becomes endogenous.

Consumers' higher-order beliefs about the payoff of withholding data are much noisier, reflected by wide credible intervals for $\delta$. This pattern makes sense, given that consumers need to conjecture the firm's and other consumers' reasonings as well as other consumers' privacy preferences when forming this belief.

Overall, the belief estimates represent the level of consumer sophistication in making privacy choices *when fully informed*, as is required by GDPR and other similar regulations. My estimates suggest that with a transparent information environment, consumers are able to engage in strategic reasoning when making data sharing decisions, and their beliefs are accurate to the first order. In other policy regimes where firms are allowed to obfuscate information about how data will be used and accessed, consumers' beliefs may be further away from actual practices.

### 6.3 Dual Privacy Preferences and Selection in Shared Data

After measuring intrinsic and instrumental preferences respectively, we want to know how their relative magnitudes shift the selection pattern in shared consumer data, and how the selection pattern in turn affects the firm's view on consumers. To this end, I take the intrinsic preference and instrumental belief parameters as fixed while varying the magnitude of actual type-dependent payoff, then simulate consumers' data sharing choices and the firm's view on consumer "type" distribution. The latter depends on what the firm assumes about non-sharing consumers, which is reflected as the firm's imputation strategy when processing shared data. I examine two different imputation methods. The first method imputes the missing data using the median of observed data, consistent with a view that consumers care about privacy only intrinsically, and that people who share and who withhold their data have similar characteristics. The second method imputes missing data using the minimum of observed data, consistent with the view that privacy concerns are purely instrumental. This exercise is performed with the example of income sharing.

Figure 4 compares the distributions of *full* and *shared data* across a range of instrumental incentives and under different imputation methods. The mean instrumental incentive among consumers ranges from 0 to 2 dollars; the latter is chosen to match the mean intrinsic preference for income. As instrumental incentive increases, the composition of consumers sharing their data tilts increasingly towards high-income cohorts, indicated by the expansion of red and the shrinkage of blue regions from left to right in both Panel (a) and (b). However, the firm always ends up overestimating the proportion of low-income consumers, often in (a) and excessively so in (b). This bias is caused by the fact that high-income consumers have higher intrinsic preferences for privacy and are more willing to refrain from sharing. This selection bias is not fully offset by instrumental preference even when the mean instrumental and intrinsic preferences match. In Panel (a), the imputation value is median income among shared data, which is still lower than the population-level average. Meanwhile in Panel (b), the incorrect view that "low types are more willing to hide" exacerbates the bias in the firm's view about consumers.

In sum, taking a monolithic view about consumers' privacy preferences can result in misleading inferences about consumers and managerial decisions. Instead, firms need to either learn about the joint distribution of privacy preferences of their consumers, preferably via experimentation, or adopt data collection and analysis strategies that are agnostic about the joint preference distribution. The next two sections discuss these strategies more extensively.

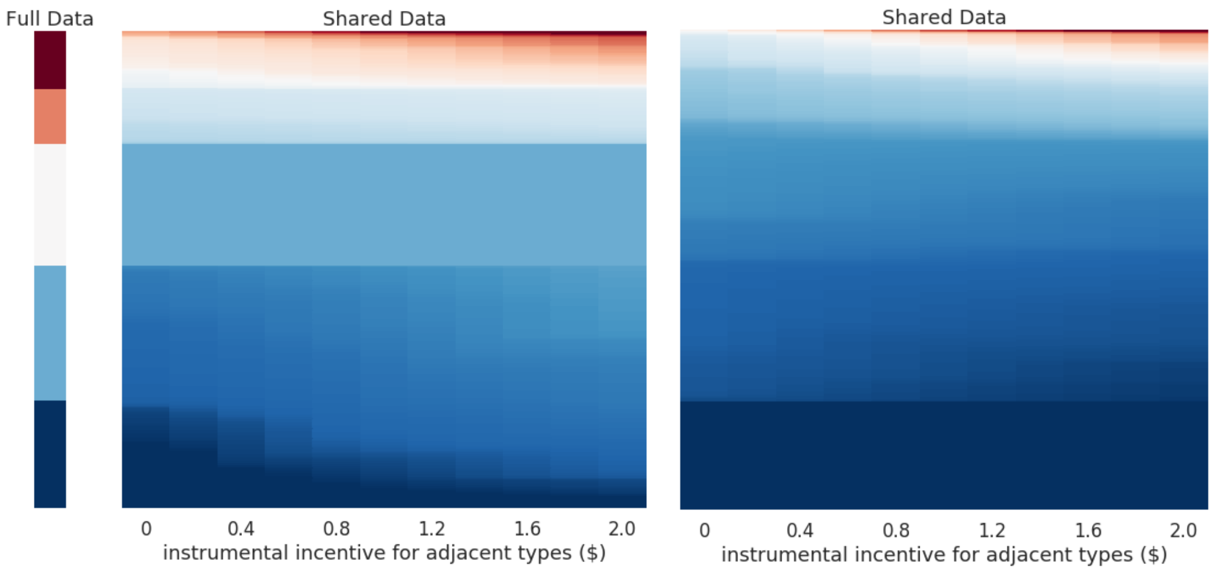
## 7 Replicating the Experiment in the Field

It is worthwhile to repeat the privacy preference measurement in the contexts of interest. Measuring the joint distribution of intrinsic and instrumental preferences will allow firms and researchers to understand consumers' data sharing decisions, even when their data utilization strategy (and

Figure 4: Full vs. Shared Data across Ranges of Instrumental Incentive

(a) Impute Missing Data with Median

(b) Impute Missing Data with Minimum




thus the instrumental preference) becomes endogenous. Below, I describe how firms and researchers can replicate my experiment to measure consumers' privacy preferences in the field.

To measure the selection in shared data, having a "ground truth" dataset is necessary. In my experiment, this is obtained by having a first stage. When such design is infeasible, the ground truth data can be obtained by having a treatment group in which consumers receive enough compensation so that everyone chooses to share. Alternatively, distribution-level statistics about the relevant type (price sensitivity, risk type, etc.) may be available from a market research company or a government agency (e.g., the census bureau).

Next is including treatments to induce exogenous variations of instrumental preferences. For example, if a firm intends to use consumer data for designing customized coupons, it can vary the depth of coupon across treatment arms, and inform consumers about the change. One challenge brought by the field setting is that instrumental preference is hard to remove completely. This is because consumers' beliefs about the consequences of revealing their personal information typically anchor on the firm's routine practices of using the data. Suppose a supermarket chain asks its customers for data and promises not to use these data for business purposes. Without additional legal guarantees, such a promise will not have commitment power: Users may still expect the supermarket owner to use these data to customize coupons and promotions.

Fortunately, by leveraging an additional assumption on consumer belief, we can still separate the two preference components, as long as variation in actual instrumental payoffs exists. In particular, assume the degree to which consumers internalize the instrumental payoff is stable. Consumers' privacy choices across treatments allow us to back out changes in instrumental preferences and compare them with changes in actual instrumental payoffs. By comparing the two, we can estimate how consumers internalize the actual instrumental consequences when forming privacy preferences. With the additional assumption, we can then calculate privacy preferences and data sharing choices in a hypothetical scenario where the instrumental preference is zero, thus backing out the intrinsic preference among consumers.

# 8 Counterfactuals

In this section, I examine data collection and analysis strategies when the firm does not know the distribution of consumers' privacy preferences. I specifically focus on two strategies that aim to correct for selection bias caused by heterogeneous privacy preferences:

1. **Sampling consumers during data collection.** By randomly sampling a subset of consumers to request data from and giving each of them higher compensation, the firm trades off data volume for representativeness. My goal is to see when this strategy can be effective in improving the efficiency of data collection.

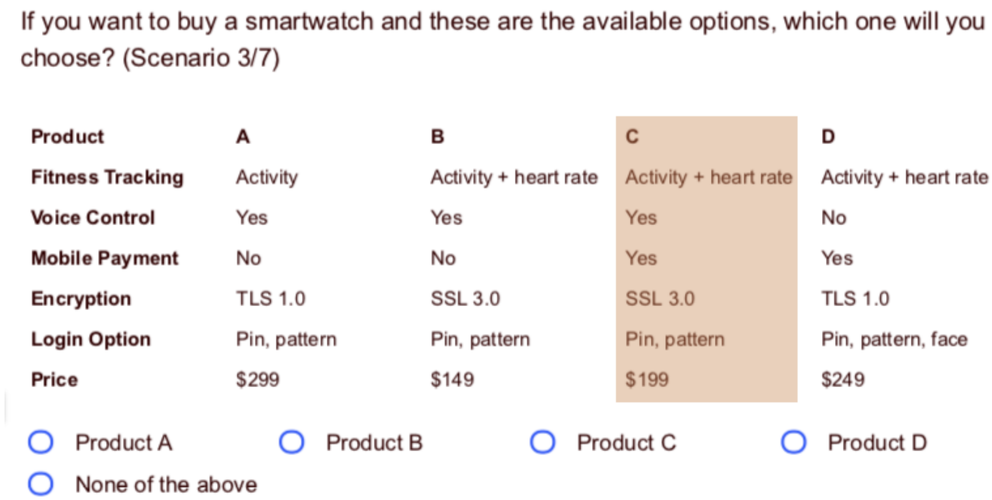
2. **Tagging non-sharing consumers when building models based on shared data.** The firm can add consumers' data-sharing decision as an additional indicator variable when building models about consumer behavior. My goal is to see whether this strategy can mitigate the selection bias and improve prediction accuracy for targeting purposes.

I evaluate these strategies in the context of price targeting. Pricing is one of the areas where firms actively use data to deliver customized offers to consumers (e.g., Dubé & Misra 2019). For the counterfactual simulation, I specify the focal firm as the third-party company featured in the experiment's second stage. I take a choice scenario featured in the first-stage conjoint survey to serve as the market environment (Task 3) and the product that the firm sells (Option C); they are displayed in Figure 5. Consumers' valuation of product features and price sensitivity are calculated from their conjoint responses. I assume the marginal cost of a smartwatch is $50, the average of two popular products on the market.[14] Data sharing choices and their impact on firms are evaluated in a GDPR-like policy regime, where firms need to seek opt-in consent before collecting data. The value of a data utilization strategy for the firm is the difference in profits with or without adopting this strategy. To calculate profits, I estimate consumer demand based on the *full data* from Stage one and view this demand as the ground truth.

To construct *firm data* in the data collection counterfactual, I first simulate 300 privacy choice draws for each compensation level, then construct a shared dataset separately for each draw: If a consumer decides not to share variable $k$, the value of $k$ is left empty. Firm data also contain a

[14]This amount is the average of the estimated production cost for Apple Watch ($83.70) and the cost of Fitbit Flex ($17.36). See https://www.forbes.com/sites/aarontilley/2015/04/30/the-apple-watch-only-costs-83-70-to-make/#6e981e8d2f08, and https://electronics360.globalspec.com/article/3128/teardown-fitbit-flex.

Figure 5: Screenshot of the Conjoint Task and Focal Product Used for Price Optimization



If you want to buy a smartwatch and these are the available options, which one will you choose? (Scenario 3/7)

| Product | A | B | C | D |
|---|---|---|---|---|
| Fitness Tracking | Activity | Activity + heart rate | Activity + heart rate | Activity + heart rate |
| Voice Control | Yes | Yes | Yes | No |
| Mobile Payment | No | No | Yes | Yes |
| Encryption | TLS 1.0 | SSL 3.0 | SSL 3.0 | TLS 1.0 |
| Login Option | Pin, pattern | Pin, pattern | Pin, pattern | Pin, pattern, face |
| Price | $299 | $149 | $199 | $249 |

○ Product A ○ Product B ○ Product C ○ Product D
○ None of the above

*Note:* Highlights are added to illustrate the focal product used for the counterfactual. They were not present in the actual experiment.

privacy choice indicator for each data-sharing decision, which equals 1 when the consumer chooses not to share $k$, and 0 otherwise. I assume the firm imputes missing variables using mean values among the shared data, and take competitors' prices as given when doing price optimization.[15]

## 8.1 When and How to Buy Data from Consumers

To evaluate data buying strategies, I start by calculating the value of consumer data shared with the firm under different levels of compensation. The firm's value of consumer data is the basis for assessing the value of data buying plans. It is also a legal prerequisite for buying consumer data in recent privacy regulations. For example, the final text of the CCPA Regulation states,[16]

> *If a business is unable to calculate a good-faith estimate of the value of the consumer's data or cannot show that the financial incentive or price or service difference is reasonably related to the value of the consumer's data, that business shall not offer the financial incentive or price or service difference.*

Afterwards, I decompose the value of consumer data based on the role they play for learning about consumers: *model estimation* and *profiling*. The goal of this decomposition is to study the role of information externality in different stages of data-driven decisions, and how it can help the firm improve the data buying plan.

Nevertheless, estimating the value of consumer data is challenging. Conceptually, it depends on what other data are already available, and is model and application specific. Computationally,

[15]If a consumer chooses not to share the choice task responses, the outcome variable of the pricing model is missing. In this case, I assume the firm imputes the missing outcome using observed conjoint choices from consumers who are demographically similar in the shared data. A real-world analog is the look-alike model commonly adopted in the industry.

[16]https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf

estimating the value of data requires simulating many different datasets and estimating a separate model for each data, in order to smooth out the idiosyncratic noises in data sharing decisions. With one single dataset, the noises inherent in the discrete privacy choices across many consumers would have rendered the value estimate imprecise and unusable. However, this requirement brings a heavy computational burden when the researcher wants to search over different compensation levels for a desirable data buying strategy. The computational problem is exacerbated by the need to incorporate the endogeneity of instrumental privacy preference in the data sharing simulation.

Below, I provide one way to calculate the value of additional consumer data for a given model, application domain, and dataset already available to the firm. Each data buying strategy leads to a different dataset shared to the firm, denoted as $d$; the full dataset is indicated as $d_0$. For a given dataset $d$, the profit loss from not obtaining full data is

$$\Delta\pi_{total} = \pi\left(P_{d_0}(d_0)\right) - \pi\left(P_d(d)\right).$$

I choose the profit level with full data as the comparison benchmark, since this is what the firm would have obtained in the old policy regime. Note that in a new privacy regulatory regime, a firm normally only observes the shared data $d$ but not the full data $d_0$. However, the firm may still learn about the value of full data via a market research company that has access to full data, or by running an experiment that collects full data.

This value is then decomposed into two parts. The first part indicates the value of data in *estimating the model*. The second part indicates its value in *profiling consumers*: that is, gathering individual consumer profiles to deploy the targeting model, taking model parameters as fixed:

$$\Delta\pi_{total} = \Delta\pi_{model} + \Delta\pi_{profile};$$

$$\Delta\pi_{model} = \pi\left(P_{d_0}(d)\right) - \pi\left(P_d(d)\right); \ \Delta\pi_{profile} = \pi\left(P_{d_0}(d_0)\right) - \pi\left(P_{d_0}(d)\right). \tag{4}$$

Here, $P_{d_0}(d)$ is the firm's pricing model trained using $d_0$ and taking $d$ as input. $\pi$ is the true profit, which is a function of the pricing strategy. At the model estimation stage, consumers impose information externality to each other via their data sharing decisions: data used for building the model will affect every consumer for whom the model is used to generate targeting decisions. On the other hand, externality is absent in the profiling stage, because one consumer's profile does not inform the firm about the profile of another consumer.

**The value of personal data to firms.** As a starting point, I calculate the value of data when consumers only have intrinsic preferences. This situation may occur when consumers who receive requests for data provision do not experience the direct economic impact from the firm's data analysis. For example, Nielson and ComScore maintain a panel of consumers and provide the data to other firms for analysis, but these firms' focal customers may not overlap with the panel. As another example, a wedding planner has one-off transactions with most of its customers, and those who already use its service will not expect direct economic consequences from sharing their data. The first two rows of Table 6 show the posterior mean and credible intervals of the profit losses at different levels of compensation. Having to seek consent results in a profit loss of \$1,440 per thousand customers when no compensation is given, which is 3% of the total profits that could have been obtained using full data. Having incomplete data in the model building stage contributes to 61.4% of the total profit loss, whereas having incomplete data for profiling results in the other one third.

Table 6: Profit Loss (\$/1,000 customers) with Different Prices for Data

| Privacy Preference | Role of Data | Price per Variable (\$) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | | 1 | | 2 | |
| Intrinsic | Model+Profile | 1,440 | [657, 3,040] | 1,126 | [536, 2,400] | 883 | [382, 2,193] |
| | Profile | 556 | [379, 716] | 465 | [317, 636] | 380 | [218, 592] |
| Intrinsic+Instrumental | Profile | 862 | [799, 892] | 857 | [779, 892] | 852 | [748, 890] |

*Note:* This table reports posterior mean estimates, with 95% credible intervals in brackets. Total profit loss is calculated as $\Delta\pi_{total} = \pi(P_{d_0}(d_0)) - \pi_d(d)$; profit loss associated with profiling is $\Delta\pi_{profile} = \pi(P_{d_0}(d_0)) - \pi(P_{d_0}(d))$.

Instrumental preferences are otherwise present, especially when a firm solicits data from its own consumers and applies its model to them. In the pricing context, the instrumental incentives are the payoff differences that consumers *expect to* receive when sharing versus withholding their data. Below, I calculate the instrumental preferences by combining the endogenous instrumental incentive and the consumer belief estimation results.

I assume consumers have "approximately rational" beliefs. Previous estimation results show that consumers are first-order rational when making data sharing decisions. However, there is no sufficient evidence that they also conduct higher-level reasoning. The requirement to sustain higher-order rationality is unlikely to be satisfied: Consumers need to believe everyone else in the market is rational and to know the distribution of other consumers' privacy preferences. To further simplify the analysis, I focus on the case in which the firm has previously trained its pricing model using a set of representative data from other customers. In other words, I only calculate the value of shared data for profiling when consumers have both intrinsic and instrumental preferences. Taking the pricing model as given, consumer $i$ expects to receive different prices when he withholds or shares data $k$:

$$E[P_i|s_{ik} = 0] = \overline{P}_{i'}; \text{ and } E[P_i|s_{ik} = 1, d_{ik}] = \overline{P}_{i', \forall d_{i'k} = d_{ik}}.$$

Here, $i'$ denotes all other consumers in the market; $\overline{P}_{i'}$ is the mean price for all other consumers, and $\overline{P}_{i', \forall d_{i'k} = d_{ik}}$ is the average price for all other consumers with the same attribute $d_{ik}$. Given that consumer $i$ can always choose the outside option when the price is too high, his instrumental preference is the difference in log sums:

$$E[\Delta U] = \frac{1}{\beta_i}\left[\log(1+\exp(v_i - \beta_i \overline{P}_{i'})) - \log(1+\exp(v_i - \beta_i \overline{P}_{i',\forall d_{i'k}=d_{ik}}))\right], \tag{5}$$

where $\beta_i$ is $i$'s price sensitivity and $v_i$ is his valuation for the product.

The last row of Table 6 shows that when consumers have instrumental preference, the loss from not obtaining the full data is larger—in this case, around twice as large as when they only have intrinsic privacy concerns. The larger magnitude is driven by a more severe sample bias in the shared data. It also shows that compensation for data sharing is less effective in overcoming instrumental incentives. The reason is that the expected difference in log sums due to revealing private information ranges from \$20 to \$50 for each data-sharing decision.

To analyze the scenario in which consumers have instrumental preferences when sharing data for both modeling and profiling, one needs to solve the full equilibrium, because the pricing model and the data shared now depend on each other. This task is computationally daunting, as each iteration involves simulating many different dataset draws and computing the firm's pricing model for each of these datasets. This analysis is part of my future work. Based on the results from the intrinsic-only case in Table 6, I conjecture that when consumers have both intrinsic and instrumental preferences, the economic loss of having incomplete data for modeling is much larger than its impact solely on profiling. In this case, the firms may achieve efficiency gain by using a separate consumer panel for modeling. This way, the firm can "insulate" the modeling sample from instrumental concerns.

**Information externality and firm's valuation of consumer data.** One way to see if a data buying strategy is worth adopting is to compare the firm's WTP for buying data and consumers' WTA for sharing data. For example, if the firm's WTP is lower than most consumers' WTA, it implies that matching the price for data to consumers' WTA will lead to a loss in profits, thus no additional data buying should take place.

The firm's WTP for obtaining $d_0$ given existing data $d$ is calculated as the profit difference divided by the unit difference between the two datasets:

$$WTP_{firm} = \frac{\Delta\pi}{N \cdot \overline{K}.}$$

Here, $\Delta\pi$ can be either $\Delta\pi_{total}$ or one of its subparts from (4), depending on the data buying strategy being considered. $N$ is the number of consumers from whom the firm wants to collect data; with a mass-collection strategy, $N$ equals the market size. $\overline{K}$ is the average number of variables withheld per consumer in dataset $d$. The resulting WTP is the break-even price that the firm is willing to offer per consumer and personal variable.

The metric above has a caveat that it represents the average value rather than the marginal value of data. However, tracing out the marginal value along different compensation levels is computationally demanding, for the reasons described at the beginning of Section 8.1. Assuming the marginal value of data decreases with the volume of data already available, the average value calculated above will serve as a lower bound for the marginal value evaluated at $d$.

Consider two different data buying strategies:

(a) *The firm buys data from all consumers and uses the data for both modeling and profiling.*

(b) *The firm allocates resources to collect data for modeling. In doing so, it randomly samples 1% of consumers, and only compensates them for sharing data.*

With strategy (a), the firm has to request data from every consumer; but with strategy (b), it can randomly sample a subset of total customers and offer to buy data only from this sample. This is because information externality exists in the model building but not at the profiling stage. When data are used to learn a systematic relationship between optimal prices and personal characteristics, data coming from one consumer also improves the inferred optimal prices for other consumers. On the other hand, knowing the characteristics of consumer A does not tell the firm about the characteristics of other consumers.

To evaluate these two strategies, I calculate the firm's WTP for additional data when it already obtains some data that consumers are willing to share when no compensation is given. The results indicate substantial improvements from leveraging information externality at the modeling stage. On average, a consumer withholds 5.31 variables without compensation. With strategy (a), $WTP_{firm} = \$1.440/5.31 = \$0.27$. In comparison, consumers' mean WTA ranges from \$0.14 to \$2.37. This result indicates that collecting data from all consumers for both modeling and profiling is not viable. With strategy (b), however, $WTP_{firm}$ = (\$1.440 − \$0.556) $\times 100/5.31$ = \$16.65, four times as much as the 97.5% quantile of consumer WTA even for the most precious data. Although this calculation may be simplistic, the qualitative pattern is general. Recent work highlights the presence of information externality in privacy choices (Acemoglu et al. 2019, Bergemann et al. 2020, Choi, Jeon & Kim 2019). My paper shows that to leverage information externality when collecting consumer data, firms need to know the stage at which it is present.

In reality, the performance of a model will increase with the size of the estimation sample. For sampling to improve the efficiency of data collection, estimation data should exhibit decreasing returns. This condition is supported by recent empirical findings, such as Bajari et al. (2019) and Claussen et al. (2019). The marginal return to data will diminish more slowly with more complex models and greater heterogeneity that the model intends to capture; in these cases, the sampling percentage should increase accordingly.

## 8.2 Learning about Selection: What Do Privacy Choices Reveal?

At the stage when available consumer data is a given, what can firms do to improve the quality of their inference? Incorporating consumers' privacy choices into the model may help. *Privacy choice* means the data sharing decision per se, apart from values of the data. An example of the privacy choice variable is the do-not-track header: It is generated when a user declines to be tracked by third parties, and remains visible to websites.[17] In a world where privacy preferences were purely instrumental, privacy choices would completely reveal a consumer's hidden type. If this were true, firms would be able to use privacy choices as additional targeting variables and substantially improve the targeting performance, even without observing the actual contents of personal data for these consumers. However, the information value of privacy choices changes substantially when consumers have both types of privacy preferences.

In what follows, I examine what firms can and cannot learn by adding privacy choices to their model. Fixing the dataset shared by consumers, I compare the performances of two pricing models. In the *without-indicator* model, the firm sets prices based on the content of data provided by consumers, but not their privacy choices. In the *with-indicator* model, the firm sets prices based on both the content of available data and consumers' privacy choices. I take $d$ as actual data shared from the experiment's second stage, instead of simulating counterfactual datasets, which is only needed when evaluating data buying strategies and would add to the sampling error. The metrics for evaluating pricing performances are

$$\Delta\pi_{without-indicator} = \pi\left(P_{d_0}(d_0)\right) - \pi\left(P_d(d)\right),\ \Delta\pi_{with-indicator} = \pi\left(P_{d_0}(d_0)\right) - \pi\left(P_{d+c}(d+h)\right),$$

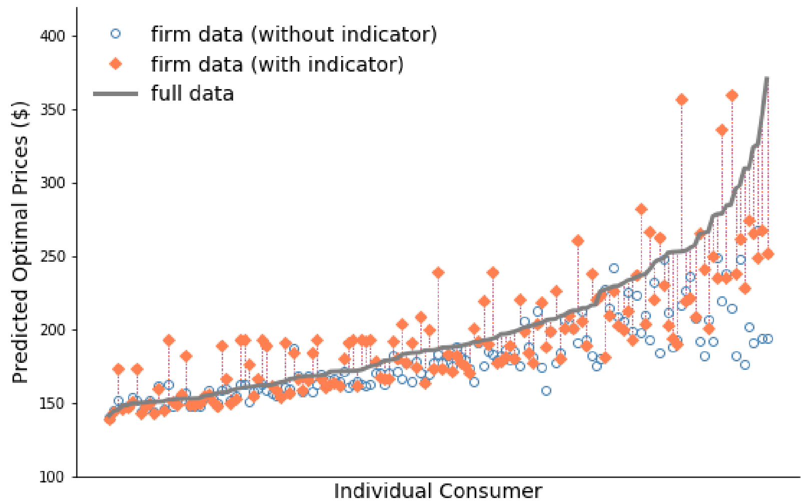
where $h$ refers to the privacy-choice indicator.

Figure 6 compares the individual-level optimal prices predicted by these two models when applied to firm data, and the benchmark prices calculated based on full data. Prices predicted by the *with-indicator* model are less biased. The mean price that consumers receive under this model is \$194.26, close to the mean price \$199.05 when the firm has all data; in comparison, the mean price under the *without-indicator* model is \$179.22. In other words, a model with privacy choice indicators can serve as a bias correction tool: By comparing average prices with and without the privacy choice indicator, the firm can learn about the direction and magnitude of selection in the shared personal data. This information is useful not only for interpreting insights obtained from the data, but also for designing and evaluating data acquisition schemes.

On the other hand, Figure 6 also shows that predictions generated by the *with-indicator* model are not necessarily more accurate. The *with-indicator* model surpasses the alternative model when predicting prices for consumers who have high valuations for the product, but performs worse at the opposite end of the spectrum. To quantify the impact of adding privacy choices on the

[17] https://www.eff.org/issues/do-not-track.

Figure 6: Inferred Optimal Prices with or without Privacy Choice Indicator




performance of price targeting, Table 7 compares the profit losses from using each pricing model. Adding privacy choices improves overall targeting performance, but only to a small extent (see column 1). The performance gain mainly comes from a better calibration of prices offered to privacy-sensitive consumers (column 3). The prediction accuracy for consumers who already share lots of data can suffer (column 2), because privacy choices add little additional explanatory power for the price sensitivity of these consumers.

Table 7: Profit Loss When Using Firm Data ($/1000 Consumers)

| | Consumer Subset | | | | | |
|---|---|---|---|---|---|---|
| Model | All consumers | | Share all data | | Share no data | |
| without-indicator | 2,441 | [917, 5,113] | 2,348 | [926, 5,721] | 2,492 | [1,229, 4,186] |
| with-indicator | 2,384 | [957, 5,229] | 2,405 | [877 , 5,193] | 2,419 | [1,138, 3,854] |

*Note:* This table reports posterior mean estimates, with 95% credible intervals in brackets. Profit loss is calculated as $\Delta\pi = \pi\left(P_{d_0}(d_0)\right) - \pi\left(P_d(d)\right)$; a lower number indicates a better performance.

Taken together, these results paint a nuanced picture of the information value of privacy choices. Incorporating consumers' privacy choices into a firm's decision model can reveal the direction and degree of sample selection, provided that the untruncated distribution of outcome (e.g., individual-level sales across all customers) is observed. On the other hand, the information value is limited when the goal is to improve individual-level pricing. Intuitively, privacy choices capture systematic differences in price sensitivity *between* consumers who share and who withhold their data, but will not reflect the heterogeneity in price sensitivity *within* the withholding consumers. The former reflects the impact of selection, while the latter is more useful for targeting. With greater heterogeneity in intrinsic preferences, consumers who decline sharing data are more likely to exhibit heterogeneous price sensitivity, and privacy choices become less useful in improving targeting as a result.

### 8.3 Summary

The counterfactual studies show that firms can improve their inference on consumers and targeting decisions by sampling consumers at the data collection stage, and by tagging non-sharing consumers at the data analysis stage. Both strategies improve inference and targeting performance by correcting for sample selection bias, without invoking assumption on consumers' privacy preference distribution. Although the quantitative results depend on the application context and categories of personal data under consideration, the qualitative findings are general. They can be applied not only when consumer data are used for targeting, but also in other managerial decisions and research settings where analyzing self-selected consumer data is necessary.

## 9 Conclusion

Privacy choices are motivated by both intrinsic preference—a taste for privacy, and instrumental preference—the expected change in payoffs from disclosing one's private information relevant to the specific market environment. While the intrinsic preference is a utility primitive, the instrumental preference is endogenous to how the firm uses consumer data to generate targeting outcomes. Separating these two preference components can help us understand how consumers self-select into sharing data, and how this selection pattern reacts to changes in the firm's data utilization strategies. Ultimately, understanding the selection in voluntarily shared data is crucial for obtaining valid insights when collecting and analyzing consumer data.

By separating intrinsic and instrumental motives using experimental variation, I establish the following findings. Consumers' WTA corresponding to intrinsic preferences are highly heterogeneous and skewed to the right: The mean valuation for sharing a demographic profile is \$10, while the 97.5% quantile is \$30. When information on data usage is delivered in a transparent manner, consumer beliefs about the instrumental consequences are first-order correct. The direction and magnitude of selection in shared data are jointly determined by the heterogeneity and correlation of the two preference components. Firms and researchers can adopt several strategies to account for the impact of privacy-induced selection when making inferences and decisions. They can run an experiment to measure the joint preference distribution among consumers, and use this information to understand the types of consumers who withhold data. Alternatively, they can adopt strategies that are agnostic about the preference distribution. Ex ante, they can allocate resources to buying a more representative dataset rather than simply increasing its volume, whenever information externality exists. Ex post, incorporating privacy choices into models can effectively debias the inference and prediction results.

Privacy preferences are known to be context-specific. My model captures the context dependence that comes from changes in perceived instrumental consequences across scenarios. In addition, privacy choices can also be influenced by various psychological shifters. I show how the experiment can be replicated in the field to address the second type of context dependence. Different versions of the experiment can be used for unpacking how consumers' intrinsic preferences respond to psychological shifters, and for examining consumers' beliefs about the instrumental consequences when firms obfuscate their data usage. The model can be extended to discuss cases in which consumers can manipulate the contents of shared information at a cost, and when sharing data improves the quality of products offered by the firm.

Although this paper does not directly discuss welfare, measuring intrinsic and instrumental preferences separately is useful for welfare calculations. First, separating these two components can help us understand the extent to which privacy preferences change endogenously with firms' strategy to use consumer data. In addition, the relative magnitudes of these two preference components have distinct welfare implications. The intrinsic preference implies a pure loss of consumer welfare caused by data collection; the instrumental preference implies welfare transfer between consumers and firms, as well as among consumers.

Future analysis will enrich the model and further explore the implications of the dual-preference framework. One direction is to investigate the substitution and complementarity among privacy choices. Another direction is to develop better models to extract information from consumers' data-sharing decisions. A third extension is to explore optimal data acquisition strategy.

# References


Acemoglu, D., Makhdoumi, A., Malekian, A. & Ozdaglar, A. (2019), 'Too much data: Prices and inefficiencies in data markets', *NBER Working Paper No. 26296* .

Acquisti, A., John, L. K. & Loewenstein, G. (2012), 'The impact of relative standards on the propensity to disclose', *Journal of Marketing Research* **49**(2), 160–174.

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

Adjerid, I., Acquisti, A. & Loewenstein, G. (2019), 'Choice architecture, framing, and cascaded privacy choices', *Management Science* **65**(5), 2267–2290.

Adjerid, I., Acquisti, A., Telang, R., Padman, R. & Adler-Milstein, J. (2015), 'The impact of privacy regulation and technology incentives: The case of health information exchanges', *Management Science* **62**(4), 1042–1063.

Al-Shahi, R., Vousden, C. & Warlow, C. (2005), 'Bias from requiring explicit consent from all participants in observational research: prospective, population based study', *The BMJ* **331**(7522), 942.

Aridor, G., Che, Y.-K., Nelson, W. & Salz, T. (2020), 'The economic consequences of data privacy regulation: Empirical evidence from gdpr', *Available at SSRN 3522845* .

Athey, S., Catalini, C. & Tucker, C. (2017), 'The digital privacy paradox: Small money, small costs, small talk', *NBER Working Paper No. 23488* .

Bajari, P., Chernozhukov, V., Hortaçsu, A. & Suzuki, J. (2019), The impact of big data on firm performance: An empirical investigation, *in* 'AEA Papers and Proceedings', Vol. 109, pp. 33–37.

Batikas, M., Bechtold, S., Kretschmer, T. & Peukert, C. (2020), 'European privacy law and global markets for data', *Available at SSRN 3560392* .

Becker, G. S. (1980), 'Privacy and malfeasance: A comment', *Journal of Legal Studies* **9**(4), 823–826.

Ben-Shahar, O. & Chilton, A. (2016), 'Simplification of privacy disclosures: An experimental test', *The Journal of Legal Studies* **45**(S2), S41–S67.

Benndorf, V. & Normann, H.-T. (2018), 'The willingness to sell personal data', *The Scandinavian Journal of Economics* **120**(4), 1260–1278.

Bergemann, D., Bonatti, A. & Gan, T. (2020), 'The economics of social data', *Cowles Foundation Working Paper* .

Bernheim, B. D., Fradkin, A. & Popov, I. (2015), 'The welfare economics of default options in 401 (k) plans', *American Economic Review* **105**(9), 2798–2837.

Bernheim, B. D. & Rangel, A. (2010), Behavioural public economics, *in* 'Behavioural and Experimental Economics', Springer, pp. 51–58.

Boas, T. C., Christenson, D. P. & Glick, D. M. (2018), 'Recruiting large online samples in the united states and india: Facebook, mechanical turk, and qualtrics', *Political Science Research and Methods* pp. 1–19.

Burtch, G., Ghose, A. & Wattal, S. (2015), 'The hidden cost of accommodating crowdfunder privacy preferences: A randomized field experiment', *Management Science* **61**(5), 949–962.

Carvalho, C. M., Polson, N. G. & Scott, J. G. (2009), Handling sparsity via the horseshoe, *in* 'Artificial Intelligence and Statistics', pp. 73–80.

Choi, J. P., Jeon, D.-S. & Kim, B.-C. (2019), 'Privacy and personal data collection with information externalities', *Journal of Public Economics* **173**, 113–124.

Choi, W. J., Jerath, K. & Sarvary, M. (2019), 'Customer purchase journey, privacy choices, and advertising strategies', *Available at SSRN 3501236* .

Claussen, J., Peukert, C. & Sen, A. (2019), 'The editor vs. the algorithm: Economic returns to data and externalities in online news', *Available at SSRN 3479854* .

Cooper, J. C. (2017), 'Separation anxiety', *Va. JL & Tech.* **21**, 1.

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N. & Chaintreau, A. (2020), 'Biased programmers? or biased data? a field experiment in operationalizing ai ethics', *Available at SSRN 3615404* .

DellaVigna, S. (2009), 'Psychology and economics: Evidence from the field', *Journal of Economic Literature* **47**(2), 315–72.

Dubé, J.-P. & Misra, S. (2019), 'Personalized pricing and customer welfare', *Available at SSRN 2992257* .

Egelman, S., Tsai, J., Cranor, L. F. & Acquisti, A. (2009), Timing is everything? the effects of timing and placement of online privacy indicators, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, pp. 319–328.

Farrell, J. (2012), 'Can privacy be just another good', *Journal on Telecommunications and High Technology Law* **10**, 251.

Goldberg, S., Johnson, G. & Shriver, S. (2019), 'Regulating privacy online: An early look at europe's gdpr', *Available at SSRN 3421731* .

Goldfarb, A. & Tucker, C. (2011), 'Online display advertising: Targeting and obtrusiveness', *Marketing Science* **30**(3), 389–404.

Goldfarb, A. & Tucker, C. (2012*a*), 'Privacy and innovation', *Innovation policy and the economy* **12**(1), 65–90.

Goldfarb, A. & Tucker, C. (2012*b*), 'Shifts in privacy concerns', *American Economic Review* **102**(3), 349–53.

Goldin, J. & Reck, D. (2018), Optimal defaults with normative ambiguity, *in* 'AEA Papers and Proceedings', Vol. 108, pp. 98–102.

Goswami, I. & Urminsky, O. (2016), 'When should the ask be a nudge? the effect of default amounts on charitable donations', *Journal of Marketing Research* **53**(5), 829–846.

Grossman, S. J. & Hart, O. D. (1980), 'Disclosure laws and takeover bids', *The Journal of Finance* **35**(2), 323–334.

Harper, J. & Singleton, S. (2001), 'With a grain of salt: What consumer privacy surveys don't tell us', *Available at SSRN 299930* .

Heen, M., Lieberman, J. D. & Miethe, T. D. (2014), 'A comparison of different online sampling approaches for generating national samples', *Center for Crime and Justice Policy* **1**, 1–8.

Hsee, C. K. (1996), 'The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives', *Organizational Behavior and Human Decision Processes* **67**(3), 247–257.

Hsee, C. K. & Zhang, J. (2010), 'General evaluability theory', *Perspectives on Psychological Science* **5**(4), 343–355.

Hui, K.-L. (2007), 'Consumer disclosure: The effects of company information presentation and question sequence', *Working Paper* .

Jin, Y. & Vasserman, S. (2018), 'Buying data from consumers: The impact of monitoring programs in u.s. auto insurance', *Working Paper* .

John, L. K., Acquisti, A. & Loewenstein, G. (2010), 'Strangers on a plane: Context-dependent willingness to divulge sensitive information', *Journal of Consumer Research* **37**(5), 858–873.

Johnson, E. J., Bellman, S. & Lohse, G. L. (2002), 'Defaults, framing and privacy: Why opting in-opting out', *Marketing Letters* **13**(1), 5–15.

Johnson, G. (2013), 'The impact of privacy policy on the auction market for online display advertising', *Simon School Working Paper* (FR 13-26).

Johnson, G. A., Shriver, S. & Du, S. (2020), 'Consumer privacy choice in online advertising: Who opts out and at what cost to industry?', *Marketing Science* **39**(1), 33–51.

Kahneman, D. (1979), 'Prospect theory: An analysis of decisions under risk', *Econometrica* **47**, 278.

Karlan, D., McConnell, M., Mullainathan, S. & Zinman, J. (2016), 'Getting to the top of mind: How reminders increase saving', *Management Science* **62**(12), 3393–3411.

Kummer, M. & Schulte, P. (2019), 'When private information settles the bill: Money and privacy in google's market for smartphone applications', *Management Science* .

Lee, Y. (2019), 'Revealed privacy preferences: Are privacy choices rational?', *Working Paper* .

Madrian, B. C. & Shea, D. F. (2001), 'The power of suggestion: Inertia in 401 (k) participation and savings behavior', *The Quarterly Journal of Economics* **116**(4), 1149–1187.

Martin, K. & Nissenbaum, H. (2016), 'Measuring privacy: An empirical test using context to expose confounding variables', *Columbia Science and Technology Law Review* **18**, 176.

Milgrom, P. R. (1981), 'Good news and bad news: Representation theorems and applications', *The Bell Journal of Economics* pp. 380–391.

Miller, A. R. & Tucker, C. (2017), 'Privacy protection, personalized medicine, and genetic testing', *Management Science* pp. 1–21.

Palmeira, M. M. & Srivastava, J. (2013), 'Free offer≠ cheap product: A selective accessibility account on the valuation of free offers', *Journal of Consumer Research* **40**(4), 644–656.

Posner, R. A. (1981), 'The economics of privacy', *American Economic Review* **71**(2), 405–409.

Rainie, L. & Duggan, M. (2015), 'Privacy and information sharing', *Available at: http://www.pewinternet.org/2016/01/14/2016/Privacy-and-Information-Sharing* .

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

Soleymanian, M., Weinberg, C. & Zhu, T. (2019), 'Sensor data, privacy, and behavioral tracking: Does usage-based auto insurance benefit drivers?', *Marketing Science* **38**(1).

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', ACM, pp. 38–47.

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

Stutzman, F., Gross, R. & Acquisti, A. (2013), 'Silent listeners: The evolution of privacy and disclosure on facebook', *Journal of Privacy and Confidentiality* **4**(2).

Tang, H. (2019), The value of privacy: Evidence from online borrowers, Technical report, HEC Paris Working Paper.

Tannenbaum, D. & Ditto, P. H. (2012), 'Information asymmetries in default options', *Working Paper* .

Thaler, R. (1980), 'Toward a positive theory of consumer choice', *Journal of Economic Behavior & Organization* **1**(1), 39–60.

Tirole, J. (2020), 'Digital dystopia', *13th Digital Economics Conference, Toulouse* .

Tucker, C. E. (2014), 'Social networks, personalized advertising, and privacy controls', *Journal of marketing research* **51**(5), 546–562.

Urminsky, O. & Kivetz, R. (2011), 'Scope insensitivity and the 'mere token' effect', *Journal of Marketing Research* **48**(2), 282–295.

Varian, H., Wallenberg, F. & Woroch, G. (2005), 'The demographics of the do-not-call list', *IEEE Security and Privacy* **3**(1), 34–39.

Warren, S. D. & Brandeis, L. D. (1890), 'Right to privacy', *Harvard Law Review* **4**, 193.

# A Proof for Proposition 1

First, define the notation for the means and covariances of preference components: $E[c_i] = \mu_c$, $Var[c_i] = \sigma_c^2$; $E[-T(d_i)] = \mu_t$, $Var[-T(d_i)] = \sigma_t^2$; $Corr(c_i, -T(d_i)) = \rho$. Note that $\Delta T(d_i) = T(F_d(d|s=0)) - T(d_i)$, where $T(F_d(d|s=0))$ does not vary across consumers. Therefore, $Var[\Delta T(d_i)] = \sigma_t^2$ and $Corr(c_i, \Delta T(d_i) = \rho$. $\sigma_c^2$ and $\sigma_t^2$ respectively represent the heterogeneity of the intrinsic and instrumental preference components.

Denote the total preference for privacy as $g_i$. Then,

$$Corr(g_i, \Delta T(d_i)) = Corr(c_i + \Delta T(d_i), \Delta T(d_i)) = \frac{Cov(c_i + \Delta T(d_i), \Delta T(d_i))}{\sqrt{Var[c_i + \Delta T(d_i)] \cdot Var[\Delta T(d_i)]}} = \frac{\rho\sigma_c + \sigma_t}{\sqrt{\sigma_c^2 + \sigma_t^2 + 2\rho\sigma_c\sigma_t}}. \quad \text{(A.1)}$$

$Corr(g_i, \Delta T(d_i))$ captures the degree to which privacy decisions can be explained by the instrumental preference $\Delta T(d_i)$. Because a one-to-one mapping exists between instrumental preference and a consumer's type (conditional on a fixed offer to non-disclosing consumers $T(F_d(d|s=0))$), $Corr(g_i, \Delta T(d_i))$ is a direct assessment of the information value of non-sharing decisions for inferring consumer types. The following observations hold:

1. $Corr(g_i, \Delta T(d_i)) > 0$ if and only if $\rho + \frac{\sigma_t}{\sigma_c} > 0$.
2. $Corr(g_i, \Delta T(d_i))$ increases with $\frac{\sigma_t}{\sigma_c}$, and strictly increases with $\frac{\sigma_t}{\sigma_c}$ if $|\rho| < 1$.
3. $Corr(g_i, \Delta T(d_i))$ increases with $\rho$ iff $\sigma_c + \rho\sigma_t > 0$, and decreases with $\rho$ if $\sigma_c + \rho\sigma_t < 0$.

Observation 3 reveals a more nuanced relationship between the explainability of instrumental preference and the correlation between the two preference components. In particular, if $\sigma_t > \sigma_c$, a regime $\rho \in [-1, -\frac{\sigma_c}{\sigma_t}]$ exists where an increase in $\rho$ leads to a decrease in $Corr(g_i, \Delta T(d_i))$. The reason is that when $\rho$ is close to -1, the variation in instrumental preference dominates intrinsic preference ($\sigma_t > \sigma_c$), leading to a perfect correlation between total preference $g_i$ and instrumental preference $\Delta T(d_i)$). Once $\rho$ deviates away from $-1$, this relationship is loosened. Note that when $\sigma_t < \sigma_c$, $Corr(g_i, \Delta T(d_i))$ always increases with $\rho$.

The proof goes through regardless of the level of $T(F_d(d|s=0))$. In particular, consumers need not have rational expectations, such that their beliefs about $T(F_d(d|s=0))$ are consistent with the actual transfer that the firm gives to consumers who withhold their data. By the same token, firms need not have correct inference about consumers who choose not to share data. In other words, the conclusions above are robust to scenarios where firms actively experiment or where information is inadequate for consumers or firms to form rational beliefs. The proof also remains valid when compensation for data sharing is present.

# B Displayed Compensation Schedules across Treatments

Figure B.1: Displayed Compensation Schedule: Intrinsic Treatment

(a) Main Screen

You will have the chance to win a $20 gift card if you choose to share your responses with Odde, our corporate partner. Odde is a high-end smart device manufacturer; it hopes to use the survey data to inform product development. To encourage participants to share their feedback, **it decides to increase the probability of winning for participants who share more information** (see details).

You can choose what information to share with Odde by selecting or deselecting the boxes below. Note that only the questions that you previously chose to give an informative answer to (i.e. not stating "prefer not to say") can potentially be shared. *Any information that you choose not to share with Odde will not be obtained by the company.*

| | | |
|---|---|---|
| ☐ Choice task responses | ☐ Ethnicity | ☐ Age |
| ☐ Income | ☐ Marital status | ☐ Gender |
| ☐ Zipcode | ☐ Education | ☐ Kids at home |

(b) Details Screen

A participant's winning probability is calculated by the following formula:

Probability of winning = number of boxes checked × 1%

For example, if you decide to share your responses to 5 questions that you previously gave, your probability of winning will be 5%.

Figure B.2: Displayed Compensation Schedule: Instrumental Treatment

(a) Main Screen

You will get the chance to win another $50 reward if you choose to share your responses with Odde, our corporate partner. Odde is a high-end smart device manufacturer; it hopes to use the survey data to inform product development.

Your probability of getting the reward will increase with the amount of information that you share. Meanwhile, Odde is designing a new smartwatch geared towards tech-savvy, high-income consumers, and wants to get more feedback from this group of people. As a result, it chooses to assign higher winning probabilities to participants who fit into this profile. In particular, **if it infers you to be wealthy or likely to purchase a smartwatch in the near future, the probability of you winning the reward will increase substantially** (see details).

You can choose what information to share with Odde by selecting or deselecting the boxes below. Note that only the questions that you previously chose to answer (i.e. not stating "prefer not to say") can potentially be shared and therefore be displayed. *Any information that you choose not to share with Odde will not be obtained by the company, and therefore will not be used for determining the winning probability.*

- ☐ Choice task responses
- ☐ Education
- ☐ Ethnicity
- ☐ Marital status
- ☐ Gender
- ☐ Kids at home
- ☐ Income
- ☐ Age

(b) Details Screen

Your winning probability is determined both by the baseline probability and by the adjustment terms. The baseline winning probability is calculated as follows:

$$\text{Baseline probability of winning} = \text{Number of boxes checked} \times 1\%$$

This baseline probability is then adjusted to encourage response sharing from the customer group that Odde intends to serve, as shown in the following chart:

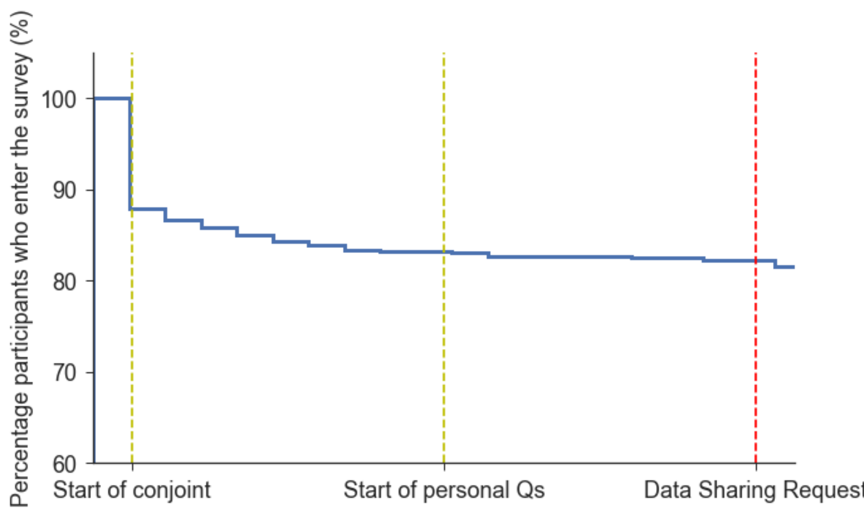
| Income | < $50,000 | $50,000 – $75,000 | > $75,000 |
|---|---|---|---|
| **Adjustment** | **-2%** | **Unchanged** | **+2%** |
| Plan to purchase a smartwatch in the next 3 months | Somewhat or extremely unlikely | Neither likely nor unlikely | Somewhat or extremely likely |
| **Adjustment** | **-2%** | **Unchanged** | **+2%** |

For example, if you have checked 5 boxes, then your baseline winning probability will be 5%. In addition, if the information you share indicates that your annual income is between $75,000 and $100,000, but you are unlikely to buy a smartwatch in the short run, then your final probability of winning will be 5 + 2 - 2% = 5%. The final winning probability never goes below zero.

Any information that you choose not to share with Odde will not be accessed by the company, and therefore will not be used to adjust your winning probability. Meanwhile, *Odde might still be able to use the information that you choose to share (e.g. zipcode, age, education) to infer your income level and your willingness to purchase*.

# C Attrition

Figure C.1: Percentage of Participants Remained Throughout the Survey




# D Credible Intervals for Intrinsic Preference Estimates (WTA)

Table D.1: Posterior Estimates of Mean and Standard Deviation of the Intrinsic WTA

(a) WTA Mean

| | mean | 95% CI |
|---|---|---|
| income | 1.870 | [1.012, 3.518] |
| intent | 1.825 | [0.981, 3.534] |
| gender | 0.142 | [-0.285, 0.709] |
| age | 0.260 | [-0.172, 0.805] |
| education | 1.228 | [0.619, 2.337] |
| relationship | 0.687 | [0.249, 1.454] |
| kid | 2.367 | [1.337, 4.523] |
| zipcode | 0.985 | [0.450, 1.992] |
| race | 0.980 | [0.437, 2.008] |

(b) WTA Standard Deviation

| | mean | 95% CI |
|---|---|---|
| income | 0.906 | [0.379, 1.840] |
| intent | 1.337 | [0.702, 2.615] |
| gender | 0.929 | [0.438, 1.965] |
| age | 1.078 | [0.536, 2.173] |
| education | 0.805 | [0.330, 1.602] |
| relationship | 0.998 | [0.477, 1.973] |
| kid | 1.001 | [0.465, 1.990] |
| zipcode | 0.982 | [0.455, 1.953] |
| race | 0.906 | [0.406, 1.801] |

# E Intrinsic WTA Estimates with Heterogeneous Sensitivity to Income

As a robustness check, I also calculate consumers' WTA distribution corresponding to Model 4, which allows consumers to have heterogeneous preferences in both the intrinsic value for privacy and monetary compensation. The estimated sensitivity to income is not very different among consumers. The median sensitivity is 0.15; for consumers at the bottom 2.5% quantile, $\beta = 0.13$, while for the top 2.5% quantile, $\beta = 0.18$. Table E.1 reports the posterior distribution of intrinsic WTA from Model 4. Compared to the main results in Table 5 and Figure 3, the estimated WTA distribution from Model 4 exhibits slightly larger heterogeneity among high-value variables and smaller heterogeneity among low-value ones. That being said, overall the two sets of estimates are similar both qualitatively and quantitatively.

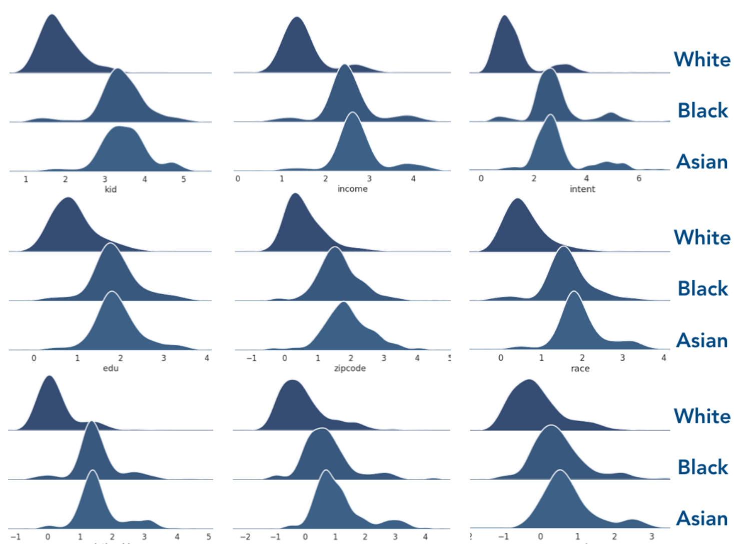
Table E.1: Posterior Distribution of WTA in Intrinsic Preference (with Heterogeneous Sensitivity to Income)

| | mean | median | 2.5% | 97.5% |
|---|---|---|---|---|
| kid | 2.253 | 2.007 | 1.051 | 4.453 |
| income | 1.784 | 1.533 | 0.794 | 3.882 |
| intent | 1.742 | 1.261 | 0.341 | 5.097 |
| education | 1.189 | 1.008 | 0.224 | 2.960 |
| zipcode | 0.959 | 0.740 | -0.114 | 2.971 |
| race | 0.951 | 0.734 | -0.059 | 2.919 |
| relationship | 0.691 | 0.404 | -0.359 | 2.870 |
| age | 0.271 | 0.081 | -0.927 | 2.647 |
| gender | 0.149 | -0.010 | -0.897 | 2.150 |

*Note:* Numbers in this table refer to statistics associated with the estimated WTA distribution among consumers; these are measures of preference heterogeneity.

# F Intrinsic Preference Distribution: Bimodal Pattern Decomposition

Figure F.1: WTA in Intrinsic Preference Distribution across Racial Groups

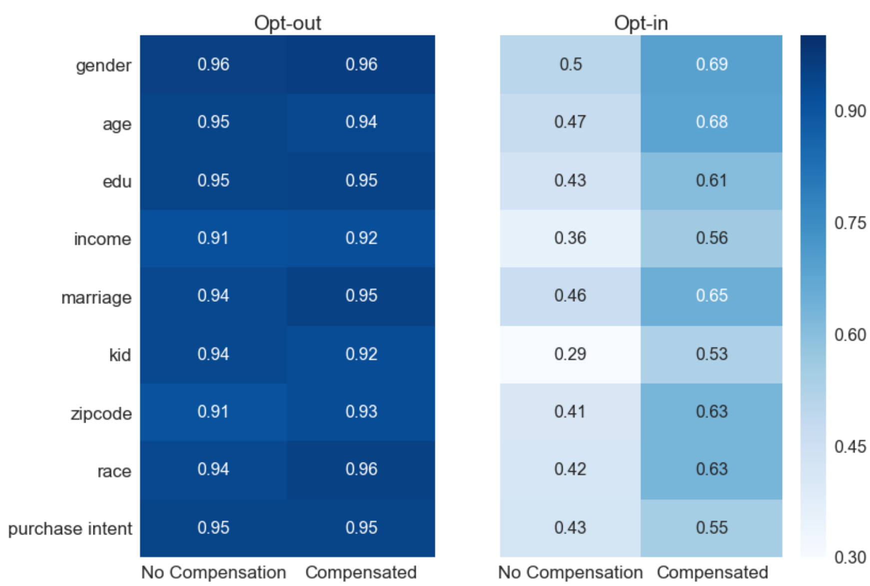


# G Psychological Factors

## G.1 The Default Frame

Figure G.1 visualizes the data-sharing frequency in different default regimes. Under the opt-out regime, almost everyone shares everything, regardless of the amount and format of compensation. The lack of choice variation in the opt-out regime does not per se imply a weaker preference for privacy or economic incentives; it simply means the impact of a "share-all" frame is strong enough to dominate other components in utility.

Figure G.1: Frequency of Data Sharing under Different Policy Regimes




**Interaction between the default regime and privacy preferences.** The literature has widely acknowledged the fact *that* default frame influences choices (Kahneman 1979, Thaler 1980, Johnson et al. 2002). However, little consensus exists on *how* or *how much* default affects choices. To flexibly characterize how default influences privacy choices, I estimate separate models for each default frame. Table G.1 displays the estimated privacy preferences under opt-in and opt-out regimes. In the comparison below, I acknowledge the scaling differences across the models, and normalize parameters to the same (dollar) scale when needed. The scaling does not affect the sign of parameters, nor the sensitivity ranking across categories of data within the same model. The comparison of belief parameters $w$ and $\delta$ are not affected by the scaling either, since these parameters directly apply to the sensitivity to compensation parameter $\beta$.[18]

To compare intrinsic-preference parameters across models, Figure G.2 displays the willingness to pay (WTP) of intrinsic preferences, which are heavily influenced by the default frame. The

[18]To see this point, note that if the instrumental utility is $w \cdot \beta \cdot \Delta E[d]$ in the utility space, then its dollar value is simply $w \cdot \Delta E[d]$.

Table G.1: Privacy Preferences across Default Frames

| | Default Frame | Opt-In | | Opt-Out | |
|---|---|---|---|---|---|
| | | mean | 95% CI | mean | 95% CI |
| intrinsic | $c_{income}$ | 0.906 | [0.588, 1.323] | -1.903 | [-2.705, -1.134] |
| | $c_{intent}$ | 0.826 | [0.419, 1.322] | -2.704 | [-3.653, -2.127] |
| | $c_{gender}$ | 0.189 | [-0.162, 0.664] | -2.988 | [-3.956, -2.184] |
| | $c_{age}$ | 0.262 | [-0.088, 0.733] | -2.429 | [-3.127, -1.729] |
| | $c_{education}$ | 0.624 | [0.329, 1.051] | -2.739 | [-3.301, -2.161] |
| | $c_{relationship}$ | 0.497 | [0.124, 1.010] | -2.734 | [-3.331, -2.105] |
| | $c_{kid}$ | 1.109 | [0.790, 1.461] | -2.143 | [-2.692, -1.380] |
| | $c_{zip}$ | 0.560 | [0.227, 1.066] | -2.093 | [-3.448, -1.328] |
| | $c_{race}$ | 0.604 | [0.285, 1.104] | -2.660 | [-3.518, -1.805] |
| instrumental | $w_{income}$ | 2.118 | [0.108, 3.989] | 1.994 | [0.136, 3.893] |
| | $w_{intent}$ | 1.942 | [0.383, 3.762] | 1.995 | [0.109, 3.909] |
| | $\widetilde{\delta}_{income,0}$ | 0.047 | [-0.186, 0.282] | 0.054 | [-0.183, 0.280] |
| | $\widetilde{\delta}_{income,1}$ | 0.037 | [-0.192, 0.284] | 0.052 | [-0.185, 0.286] |
| | $\widetilde{\delta}_{intent,0}$ | 0.059 | [-0.352, 0.379] | -0.121 | [-0.391, 0.350] |
| | $\widetilde{\delta}_{intent,1}$ | -0.049 | [-0.362, 0.324] | -0.129 | [-0.384, 0.281] |
| sensitivity to compensation | $\beta$ | 0.146 | [0.070, 0.2359] | 0.046 | [0.001, 0.141] |
| log posterior | | -7476 | [-7540, -7407] | -2075 | [-2166, -1981] |

*Note:* The models are estimated separately for each default frame. Variables are normalized using the Gelman method before estimation. Both models allow for heterogeneity in intrinsic preferences.

negative WTPs imply that once data are obtained by companies, consumers will not take back their control over personal data, unless they are incentivized by the amount indicated by the WTP. In my data, the gap between median WTA and median WTP amounts to \$69.18 (income) to \$88.06 (gender). In comparison, previous literature estimates dollar values of default in 401(k) plan enrollment decisions that range from \$37–\$54 (Bernheim et al. 2015) to \$1,200 (DellaVigna 2009). However, the WTP estimates are very noisy, due to the fact that the estimated sensitivity to compensation in the opt-out regime is close to zero (see Table G.2 for credible interval estimates).

Interestingly, Table G.1 shows that the default frame does not heavily influence consumer beliefs about the instrumental payoff. The differential impacts of default suggest that while subjective evaluations are more susceptible to the influence of the default condition, objective evaluations—beliefs about the instrumental payoff—are less so. In view of this fact, distinguishing between the intrinsic and instrumental preferences also reveals how default (and potentially other psychological factors) influences different privacy motives differently.

The managerial implication is immediate. With a regulation that mandates opt-out consent, firms can still collect most customer data even if consumers are fully informed when making privacy choices. However, once the firm moves to an opt-in regime, it will incur substantial losses

Figure G.2: Posterior Predicted Density of WTP in Intrinsic Preference

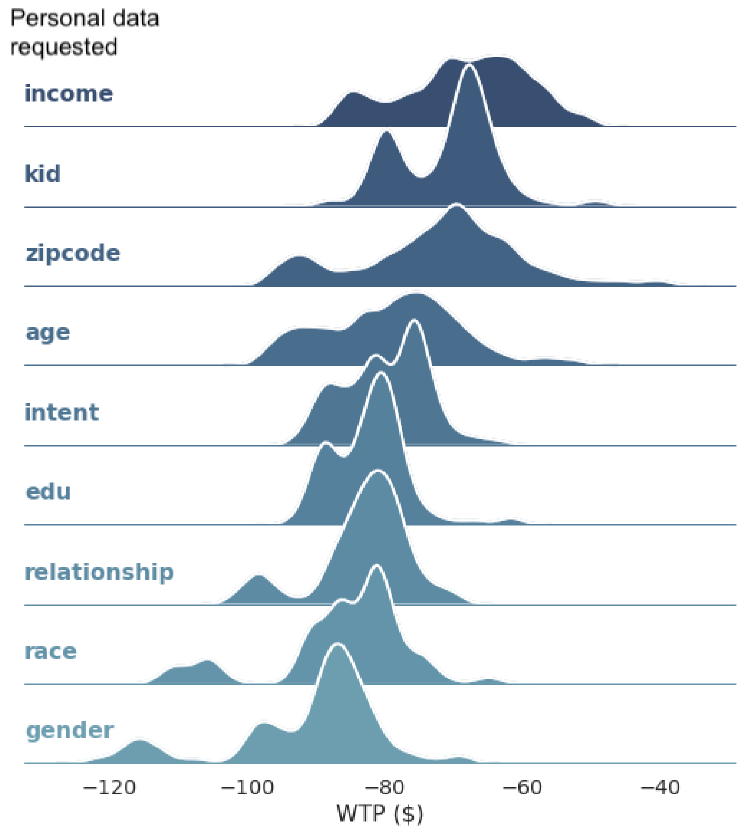



Table G.2: Posterior Estimates of Parameters Associated with Intrinsic WTP Distribution

(a) WTP Mean

| | mean | 95% CI |
|---|---|---|
| income | -66.59 | [-621.55, -6.92] |
| intent | -78.87 | [-733.52, -8.29] |
| gender | -89.84 | [-866.19, -9.28] |
| age | -76.57 | [-722.37, -8.03] |
| education | -81.10 | [-767.11, -8.41] |
| relationship | -82.10 | [-773.98, -8.63] |
| kid | -70.15 | [-634.81, -7.34] |
| zipcode | -70.52 | [-653.87, -7.46] |
| race | -86.69 | [-834.71, -8.97] |

(b) WTP Standard Deviation

| | mean | 95% CI |
|---|---|---|
| income | 70.23 | [2.41, 479.59] |
| intent | 35.52 | [0.96, 249.39] |
| gender | 57.36 | [1.92, 360.06] |
| age | 76.40 | [2.70, 522.52] |
| education | 34.59 | [1.22, 248.68] |
| relationship | 43.54 | [1.22, 286.02] |
| kid | 29.53 | [1.37, 185.01] |
| zipcode | 57.63 | [3.04, 405.01] |
| race | 28.37 | [1.30, 213.90] |

in the amount of data collected. The default paradigm is also useful for thinking about the real impact of data-portability rights.[19] Taking the incumbent as the default choice, consumers are less likely to opt out of incumbent tracking and transfer data to its competitors, unless the expected utility gain from switching is substantially large.

[19] GDPR Article 20 and CCPA Title 1.81.5, Section 1798.100 (d).

### G.2 Other Psychological Factors

The model includes a behavioral response term $m \cdot (p_i \geq 0) \cdot s_i$, to account for a combination of a mere-incentive effect and potential anchoring effects at the start of the survey. Behavioral response to the mere presence of incentives is well documented in the psychology literature (Shampanier et al. 2007, Urminsky & Kivetz 2011, Palmeira & Srivastava 2013), which can be explained by the theory that people are insensitive to scopes when evaluating options separately (Hsee 1996, Hsee & Zhang 2010). In treatment groups that distribute positive amounts of compensation, participants are told at the beginning that they can enter a gift-card lottery upon finishing the survey. This information may inadvertently create an additional anchoring effect, making all participants in these groups more inclined to share their data in order to get the anticipated gift-card rewards. The parameter $m$ captures the combination of these two forces. Under the second mechanism, the additional anchoring effect will be stronger for participants in the opt-in group (because an opt-out condition per se also has a substitutive anchoring effect); this possibility is accounted for by having separate $m$'s for different default conditions.

In the opt-in frame, the point estimate for $m$ is 0.76, with the 95% credible interval being $[0.65, 0.87]$. In the opt-out frame, the point estimate is 0.07, with the credible interval being $[-0.17, 0.30]$. The strong effect asymmetry and the fact that the effect is almost non-existent in the opt-out condition suggest anchoring is more likely to be the main driver of this effect.