Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Secondary Market Monetization and Willingness to Share Personal Data

Joy Wu

To cite this article:
Joy Wu (2025) Secondary Market Monetization and Willingness to Share Personal Data. Management Science 71(10):8471-8490. https://doi.org/10.1287/mnsc.2022.03423

https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 71, No. 10, October 2025, pp. 8471–8490
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Secondary Market Monetization and Willingness to Share Personal Data

Joy Wu[a]

[a]Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada
Contact: joy.wu@sauder.ubc.ca, https://orcid.org/0000-0002-5903-1958 (JW)

Received: November 5, 2022
Revised: October 2, 2023
Accepted: December 24, 2023
Published Online in Articles in Advance: January 31, 2025

https://doi.org/10.1287/mnsc.2022.03423

Abstract. People are often unaware that their personal data can serve as valuable inputs for economic activities in secondary data markets. However, whether secondary monetization of personal data determines privacy preferences remains unclear. I examine whether privacy decisions are motivated by the data recipient's ability to benefit from trading individuals' data with a third party. A large online laboratory experiment involving personally identifiable psychometric data is implemented with real data-sharing consequences and monetary benefits. I find that individuals decrease their willingness to share data—both in terms of their likelihood of participating in the data market and the prices demanded for such participation—when the recipient's ability to monetize the data through secondary trade is salient. Strategic responses to updated beliefs about the recipient's gain from the trade are ruled out via the chosen price elicitation. I find that increased data exposure (to more recipients) does not explain the significant revealed disutility from secondary monetization. These findings are also robust to controlling for the risk exposure differences between data recipients and third parties.

History: Accepted by Anindya Ghose, information systems.

Funding: This project was funded in part by the Institute for the Social Sciences, Cornell University.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2022.03423.

Keywords: privacy • personal data • data markets • economics experiment

1. Introduction

Recent events in the digital economy raise questions about whether individuals' data privacy choices are influenced by activities in secondary data markets. Cambridge Analytica collected Facebook users' psychometric data through a personality survey, and these data were used to create profiles and analytical services sold to U.S. presidential campaigns (Graham-Harrison and Cadwalladr 2018). Clearview AI harvests user-published images to fuel a facial recognition database (Hill 2020), which is sold as software to various government agencies (USAspending 2023). Personal data can also be utilized in secondary markets without allowing direct access, such as those harvested by data brokers to create data services sold to marketers. These examples motivate discourse on whether individuals demand privacy regulations over the use of their data in commercial activities.

The secondary monetization of personal data is a ubiquitous operational feature of data markets. After individuals make an initial decision to share data with a recipient, their data can be traded and monetized by the recipient with third parties. Once personal data enter secondary markets, the average consumer cannot control the economic activities that involve these data. Because personal data are linked to a person's identity and contain information about that person, individuals may have preferences over how others utilize data about them. For instance, users may feel continued ownership of the profiles that they share with a social media platform, so they demand some of the profits that accrue from the secondary use of such data. In another example, people might believe that it is unethical for anyone other than themselves to commercialize their facial images, and people are less willing to share such data if these ideals are violated.

To capture whether secondary markets negatively affect individuals, it is important to capture the "consumption value" of privacy that is enjoyed by individuals when they do *not* disclose their data.[1] If

the individual suffers disutility from a recipient's ability to monetize personal data in secondary markets, then the value of the privacy that the individual enjoys from nondisclosure increases. The aim of this study is to empirically isolate and examine this disutility, which is revealed through individuals' willingness to share data.

Using an economic experiment, I test whether increasing the salience of a data recipient's ability to monetize data in a secondary trade leads individuals to value their privacy more by being less likely to enter the data market or demanding greater benefits in exchange for data sharing. This study captures individuals' relative privacy preferences across various data-sharing conditions that contain differential information signals about secondary monetization. To reveal these preferences, subjects first generate personally identifiable psychometric data[2] and then face a privacy trade-off in choosing whether to release their data to a recipient in return for monetary payment. The consequences of data sharing and secondary monetization are real. The data recipients are anonymous and randomly selected from the study to earn profits from receiving the personal data of others. Thus, this study captures the valuations of the privacy that individuals expect to experience by not disclosing their personal data.

The results provide consistent evidence that a data recipient's ability to profit from secondary market activities decreases individuals' willingness to share personal data. An aversion to secondary monetization is present in both extensive and intensive margins, including a decreased likelihood of participating in the data market (by not sharing data at all) and among those willing to share, a higher reservation price for data sharing. Concerns about data being made available to more parties are examined and found not to explain the disutility from secondary monetization. Moreover, any perceived differential risks in data access by recipients versus third parties also do not explain the reluctance to share data when secondary monetization activities are salient.

The idea that secondary market activities may incur privacy costs is not new. Varian (1996) explains that extrinsic nuisance costs to the individual can occur as a result of secondary transactions because there is a clear externality that arises from the misalignment of individual and third-party interests. On the other hand, the extant empirical research on privacy focuses on preferences regarding *access* to personal data (John et al. 2010, Tsai et al. 2011, Acquisti et al. 2013, Athey et al. 2017, Adjerid et al. 2019), which includes the risk of data access by third parties rather than only by second parties (Buckman et al. 2019). By predominantly focusing on access concerns, existing research fails to fully capture the externalities of secondary market activities and their impact on people's willingness to share and sell personal data.

The results of this study empirically challenge those privacy models that equate privacy to the value of secrecy and protection from unwanted observation. The claim that the distaste for surveillance is the only determinant of privacy preferences is theoretically contested. Nissenbaum (2009, p. 2) stipulates that what people care about most is not the restriction of information flows but rather, is "ensuring [information] flows appropriately." In this study, I present a framework for separating privacy preferences for data being experienced as a *product* (e.g., for others to "see" or "know") versus being utilized as an *asset* (e.g., for others to use or trade as a resource in other activities). This contributes to empirical works that support more expansive privacy models. For instance, Brandimarte et al. (2012), in their exploration of preferences for control, frame their study around the notion that "access" and "use" of personal information are two different dimensions of privacy considerations, reflecting whether recipients have access to the data and if so, what they can do with such information. Other works identify various underlying components of privacy preferences, such as the intrinsic versus instrumental factors in Lin (2022) or the larger combination of nonnormative factors in Acquisti et al. (2015).

In this experiment, the fact that people make imperfectly informed choices is both assumed and confirmed, thus contributing to the literature regarding the information disadvantage of individuals when valuing personal data privacy. Consistent with existing research, my results imply that individuals are vulnerable to suboptimal data-sharing choices in online settings. Individual valuations of data privacy are highly uncertain and unstable (Acquisti et al. 2016, Tomaino et al. 2023). Collis et al. (2021) find that real-world information about the market value of their data motivates people to revise the valuation of their data, with differential consequences for more or less economically vulnerable populations.

This study contributes to improving the methodology used in experiments that elicit data-sharing choices. I accommodate the pricing-out behavior common in real-world privacy trade-offs, where benefits are usually small and rely on all-or-nothing decisions. The experiment implements a repeated-measures design that balances carryover effects, reveals the relative privacy preferences that account for the idiosyncratic nature of personalized data, and lessens contextual challenges when forming valuations.

In this study, a privacy cost due to a data recipient's secondary market monetization capability is empirically isolated and disentangled from disutility arising from data access risks. This work extends empirical privacy research beyond secrecy concerns and assesses whether people find it appropriate for a recipient to benefit from utilizing individuals' personal data in

secondary markets. I contribute to the notion that preferences over sharing data exhibiting features of an input good (i.e., an asset) can differ from those over sharing data as a finished good (i.e., a product). The research design of this study guides future works that elicit privacy preferences by using prices as the instrument. Because people often lack the information needed to account for their preferences related to the operational features of data markets, the importance of secondary market activities can be easily overlooked. Overall, I find that the secondary monetization ability of data recipients is a critical determinant of individuals' valuations of their privacy.

2. Background

2.1. Inalienability, Nonrivaly, and Input Goods

Personal data are characterized by their *inalienability* (Koutroumpis et al. 2020): information that permanently refers to a specific individual. Individuals generate and contribute to digitizing information related to their attitudes, attributes, tastes, and behaviors. These contributions can be passive, such as the exhaust data collected as a by-product of other activities (e.g., web browsing). Active data generation activities include users creating content on information platforms (e.g., likes) or responding to online surveys (e.g., personality tests). All of these data are typically considered personal data as they are connected to or identifiable[3] to the originator of the data.[4] Because of inalienability, individuals may always perceive or feel psychological ownership of personal data (Spiekermann et al. 2013), even after they trade away personal data in return for goods and services.

Another aspect of data is their *nonrivalrous* nature, which means that utilizing some data does not prevent another entity from doing the same and thus, enables secondary markets for personal data with increasing returns (Jones and Tonetti 2020).[5] This feature prevents individuals from maintaining control over their data in secondary markets, and their agency over their personal data often begins and ends with generating and releasing that data to a recipient. In practice, users can face a series of cascaded choices beginning with an "upstream" choice to join a platform—or the preselection of privacy settings—before conducting "downstream" data generation and disclosure (Adjerid et al. 2019). However, individuals are usually excluded from economic participation in any further downstream transfers of their data (i.e., they cannot influence the manner in which their data flow in these markets). Because of the nonrivalry of data that enables, theoretically, infinite transfers, individual preferences for whether and how their data are accessed and utilized in secondary markets should be examined.

Finally, data are not often finished goods (i.e., products consumed and experienced by a recipient) but rather, *input goods* (e.g., assets utilized in a commercial activity). Once released to a single recipient, data can be repeatedly traded. However, the originator of the data no longer receives the surplus accrued from their data in secondary uses. This situation leads to a valid debate over how and whether consumers should receive data dividends (Arrieta-Ibarra et al. 2018) as the original suppliers of personal data that are utilized in secondary markets.

The inalienability, nonrivalry, and input good nature of personal data underscore the importance of broadening empirical privacy research to understand the role of individuals as the suppliers of data inputs for the digital economy. This study contributes to the economics of data markets by examining individuals' privacy preferences regarding their inalienable data as an *asset* utilized for profit in secondary trades rather than a *product* to be experienced or consumed.

2.2. Unique Privacy Concerns for Secondary Data Markets

Seminal privacy research in information systems documents an association between individuals' privacy attitudes and the secondary use of their personal information. According to Culnan (1993), those less concerned about secondary use are also less concerned about other privacy features, including control over access to personal information and the nuisance of privacy invasions. Sutanto et al. (2013) finds that an information technology solution that prevents third-party data sharing reduces perceived privacy intrusions. Recent experiments also feature data with either finished or input good characteristics. Athey et al. (2017) measure people's willingness to prevent surveillance (i.e., framing personal data as a finished good) and find weak revealed preferences, despite strong stated concerns. Buckman et al. (2019) frames personal data as an input good but finds no evidence of changes in behavior when participants are informed about the risk of data being distributed to third parties. These studies offer some clues about privacy preferences related to the nature of personal data markets.

Studies focusing on exposure concerns—or borrowing terminology from Nissenbaum (2009, p. 2), to how many parties the data are "flowing"—find weak revealed preferences, such as the value of preventing surveillance or the tolerance for data being accessed by a third party. In addition, studies often include explicit or implied data usage risks that can influence the instrumental value of privacy as opposed to the intrinsic value of privacy (Lin 2022). This makes it difficult to isolate whether the operational features of the secondary data market are undesirable in themselves rather than the consequences of such features. Therefore, little is known about whether individuals might find certain forms of data transfer, such as

secondary monetization, more appropriate than other forms.

Data markets are an important context for studying more expansive privacy models by considering how features of data markets can influence privacy choice making. External data sharing is now a common and profitable digital business strategy among firms. Data-based consumer analytics have evolved beyond the nuisance costs of unwanted solicitation theorized in Varian (1996) and into issues of targeted advertising and even digital mass persuasion (Matz et al. 2017). Today, some of the largest and most profitable digital companies are built on personal data. As consumers become more aware of how the data economy operates, their privacy concerns may also focus beyond the issue of unwanted surveillance.

The user-generated data appended to personal identifiers can be highly valuable in commercial data markets. From the perspective of parties interested in utilizing data, the value of consumer data is more than simply a function of the ability to identify individuals and is not necessarily correlated with the degree of secrecy or sensitivity of the content.[6] For example, psychometric data based on the five-factor model have the ability to understand, predict, and discriminate the attitudes and behaviors of individuals (Goldberg 1992, McCrae and John 1992, Junglas et al. 2008, Matz et al. 2017, Li et al. 2019).[7] Miller and Tucker (2018) make a similar characterization about the predictive power of a person's genetic data on future health risks. These data (unlike phone numbers or email addresses) contain much information about other behaviors and traits that have consequences for long-term welfare.

There is growing theoretical and empirical support for the idea that privacy preferences are combinations of more primitive ideals. Individuals have idiosyncratic intrinsic tastes for privacy separate from the instrumental value of privacy (Lin 2022). This can be conceptualized by viewing privacy as a commodity enjoyed for its own sake and has various attributes valued more or less by each individual. As described by Farrell (2012), some of these attributes make privacy seem like a final good (i.e., caring about privacy for its own sake in an intrinsic way), and others make it seem like an intermediate good (i.e., caring about privacy in a more instrumental way).

A person's aversion to a data recipient's ability to conduct secondary data monetization can be consistent with privacy as a final and intermediate good.[8] For instance, individuals may find it unfair that the trading of their privacy (as a final good) is profitable to others rather than only to themselves. On the other hand, a desire for control and self-determination enabled by privacy (as an intermediate good) can explain an aversion to secondary transactions of their data. Some people may find secondary monetization to be unethical or untrustworthy (Culnan and Armstrong 1999). Other people may feel that they have ownership rights over the data that others have about them (Spiekermann et al. 2013) and desire a share of the profits gained from the use of their personal data. Overall, any combination of these ideals and beliefs may explain individuals' revealed preferences when faced with salient secondary monetization capabilities of data recipients, which can be separated from purely objective consequences of sharing personal information (e.g., price discrimination or identity theft).

Overall, more expansive privacy models motivate an empirical examination of privacy preferences related to secondary markets, where data are usually bought and sold as inputs or assets rather than finished goods or products. Although this study measures a privacy trade-off (i.e., willingness to share data), these choices result from unobserved preferences over the attributes of a privacy commodity. By examining the relationship between data sharing and secondary monetization, this study explores theories of privacy preferences that extend beyond secrecy concerns and into preferences related to operational features of secondary data markets, including but not limited to notions of fairness ideals, self-determination, and perceived ownership.

2.3. Unstable and Uninformed Revealed Privacy Preferences

Following the "privacy calculus" framework of Laufer and Wolfe (1977), the privacy literature provides a wide range of costs and benefits that can enter into each disclosure decision (Culnan and Armstrong 1999, Dinev and Hart 2006). Within the realm of privacy economics, the disclosure decision is generalized as the individual's trade-off in utility over wealth and privacy: $u(w, p)$ (see Acquisti et al. 2013, Buckman et al. 2019). This model describes a person with p^+ amounts of privacy considering entering a state with $p^- < p^+$ amounts of privacy. A perfectly optimized decision would demand price r such that $u(w, p^+) = u(w + r, p^-)$. The challenge with this rational privacy choice is that individuals are burdened with estimating r. In reality, an individual's estimate, $\hat{r}$, is likely biased and incorrect because of various inattention, environmental, or nonnormative factors. In this study, therefore, it is never assumed that individuals have perfectly informed choices. Instead, I assume that individuals are motivated to improve their estimate of the benefits they demand, $\hat{r}$, once they become more informed.

Individuals' lack of awareness regarding data-sharing consequences is a challenging issue to address. Consequences may result from failed-to-imagine scenarios and incorrect presumptions concerning data control rights. Moreover, the complexity of data markets and confusion regarding the terms and conditions of releasing personal data exacerbate the salience challenges

to individuals' privacy decisions. Behavioral research already documents the instability of people's valuations of their privacy (Adjerid et al. 2013, Acquisti et al. 2015). Most of this instability is attributed to a lack of awareness and incomplete information concerning disclosure outcomes (Acquisti and Grossklags 2005).

Many other behavioral affects and heuristics regarding privacy choices can come into play (Acquisti and Grossklags 2008). For instance, individuals often obtain tangible, immediate benefits in return for generating data, and individuals may have tendencies to overweigh immediate payoffs (O'Donoghue and Rabin 1999). Moreover, suppose that individuals myopically focus on the immediate rewards of information sharing while ignoring the less vivid downstream costs of data trading. In this case, they do not accurately reveal their willingness to accept all—and especially the more opaque—of the privacy consequences.

Uninformed privacy choices are consequential for consumer welfare, motivating research on information interventions to help people make better (i.e., utility-enhancing) privacy choices. In most data markets, individuals cannot reappropriate the information that they disclose.[9] Often, their only available privacy choice rests in their initial decision to disclose raw information that can be digitized. To correct consumer inattention toward data-sharing consequences, policy-oriented research examines the effects of notice-and-consent policies (Tsai et al. 2011, Athey et al. 2017). However, the challenge here is identifying which opaque features of data markets can affect privacy preferences. Therefore, this study contributes to the work on notice-and-consent policies by providing novel evidence of a feature that consumers both care about and are inattentive toward—namely, the secondary monetization capability of the data recipient.

2.4. Interpreting the Monetary Values of Personal Data

As used in economic experiments, monetary rewards are an appropriate medium for measuring the value of privacy, even when real data markets resemble a bartering economy (i.e., data in return for goods and services). Empirical privacy works show that monetary rewards are effective at obtaining private information (Hui et al. 2007, Xu et al. 2010). In contrast, the use of other goods in exchange for data has severe interpretability limits and biased estimates of the value of privacy (Tomaino et al. 2023). Prices can be modeled into all forms of data sharing as an economic trade-off, where—even when there are no explicit monetary amounts involved—there is always an implicit price for a person's loss of privacy from disclosing personal data. Money is comprehensively used as the numeraire in a vast majority of real-world transactions; therefore, it is relatively easy for individuals to estimate their preferences regarding money. Money is easily divisible, and individuals nearly always prefer more than less.

Unlike the ease of understanding how individuals value money, the interpretation of privacy valuations can be challenging. Preferences for sharing personal data are highly idiosyncratic across individuals, which can lead to uninformative average valuations. Idiosyncrasy is not challenging to interpret when there are differences only in preferences concerning the same commodity; however, complexity arises when the commodity differs from one person to the next. Individual data are, after all, personalized. For instance, people might perceive that the disclosure of certain personality types is neither intrinsically invasive nor instrumentally harmful, whereas the release of other personality types is highly embarrassing or risky. Depending on their personality score, individuals may be more or less reluctant to share their psychometric data.

Another complication is that individuals might have relative rather than absolute valuations of the state of privacy that they enjoy from not participating in data markets. For instance, they may prefer to disclose data under some conditions more than others; however, they may make very imprecise estimations of the benefits (e.g., prices) that they are willing to accept under any condition in isolation. Experimental evidence from Adjerid et al. (2018) already shows that reference dependence is important and present in privacy decision making, particularly in actual choice contexts.

To circumvent these challenges in empirical privacy research, allowing individuals to make relative privacy choices can be helpful. In essence, studies can observe individuals' changes in willingness to share data in response to a common environmental factor. To understand how external factors influence personal data sharing, observing repeated measures of individuals' privacy behavior can capture more informative privacy responses, which controls for idiosyncratic data and supports reference-dependent decision making. A discussion of the methods for implementing the within-subject design used in this study is presented in Section 4.5.

Depending on the choice architecture, reservation prices can manifest as choices to participate in a data market (i.e., the selection decision) and conditional on participating, the benefits demanded (i.e., the price decision). Assuming that a person has some degree of inattention toward consequences that influence the consumption value of not sharing data, $\hat{r}$ (i.e., the reservation price to compensate for privacy loss) can shift with an increasing awareness of any relevant consequences as a result of more salient environmental signals. Holding fixed the available market prices for data, I can capture relative preferences using the differential inattention that individuals have toward certain aspects of data markets (for further details, see Section 3.1).

3. Framework

This study is focused on measuring an individual's willingness to share her data or equivalently, her valuation of the privacy enjoyed from *not* sharing data. This is captured through observing how much compensation a person is willing to accept to trade away privacy. The consumption value of privacy, revealed by the person's willingness to share, represents the utility gained or lost when the individual discloses her data with a set of recipients.

Many different consequences of data sharing can influence how much an individual gains or loses in utility. Brandimarte et al. (2012, p. 341) describes that the action to share data with some set of recipients is a necessary precondition for the "access, use, and potential misuse" of data. Disutility from disclosure can emerge from the recipient accessing the data as well as sharing the data with others. In sharing the individual's data with others, the recipient may economically benefit from the secondary monetization of personal data.

I organize two possible components for the individual's value of privacy in preventing the recipient from accessing personal data. Suppose that some possible amount of disutility that the individual suffers when her personal data are released to recipients is $V = v(e) + o(e)$, where $e \geq 0$ is the number of data recipients.[10] The first component v is the disutility from sharing her data to be experienced as a *product* by data recipients. The second component o is the disutility from sharing her data to be utilized as an *asset* by recipients.[11] For example, v can reflect unwanted observations or judgments by others. As discussed in Section 2.2, the o component can include or result from a desire for control over personal data that is violated if a recipient participates in secondary data markets, a distaste for a recipient profiting from personal data about her, or a perception of ownership over the personal data released to a recipient.

3.1. Manipulating Salience to Identify Disutility from Secondary Monetization

It is challenging to observationally disentangle preferences over sharing data as a product versus an asset. The decision to release data includes the preferences both for tolerating data being made available to be experienced *and* for tolerating data being utilized by the recipient. The solution proposed by this study is to exploit the inattention that individuals might have toward the asset nature of data by manipulating its salience and exogenously influencing individuals' awareness about the secondary monetization abilities of their data recipient.

To organize and demonstrate how privacy behavior can change in response to information signals, I adopt the inattention framework from DellaVigna (2009, p. 349). Suppose that the individual is inattentive to features related to data as an asset, the o component. Then, she perceives that

$$\hat{V} = v(e) + [1 - \theta(s)]\, o(e),$$

where θ is the inattention parameter as a function of salience $s \in [0,1]$ of o. Assuming that $\theta'(s) < 0$, $\theta(1) = 0$ is full awareness with a fully salient signal, and $\theta(0) = 1$ is complete blindness with no salient signal (which follows psychological theory positing that information attention is nondecreasing in salient signals). Therefore, varying the salience of the information about the recipient's secondary monetization ability should reveal the inattentive individual's preferences related to her data being utilized as an asset.

The main prediction is that individuals decrease their willingness to share data when it is more salient that their recipient can monetize that data in a secondary market. However, there are consequences and attributes entangled with a recipient's secondary monetization ability that can potentially increase other privacy concerns related to data being experienced by others as a product (the v component). These factors include increased exposure and differential exposure risks. Additional considerations for the conditions under which individuals make data-sharing choices can be used to isolate the importance of secondary monetization in individuals' privacy choices.

3.2. Exposure Concerns Due to Secondary Monetization

Secondary market monetization necessarily makes data available to more recipients. Any revealed disutility from secondary market transactions can potentially be explained by an individual's dislike of additional parties experiencing her data as a product rather than by her dislike of her data being utilized as an asset. For instance, if secrecy is the main privacy concern of individuals, then this can explain the dislike of the secondary monetization activities of data recipients (as opposed to a dislike specific to the ability of their recipients to benefit from such secondary transactions). To control for this factor, secondary monetization signals can be explicit about the presence of a single third party with which the recipient can transact, thus limiting the exposure level to only two parties (the recipient and the third party). Then, this context can be compared with how individuals value privacy from many parties but in the absence of secondary monetization signals. If individuals are more likely to share data at higher exposure levels with *no* signals about secondary transaction ability, then concerns about secrecy are unlikely to explain a decreased willingness to share data when secondary monetization is salient.

On the other hand, individuals may not strictly perceive that their recipient's transaction with a single third party is equivalent to making data available to one additional recipient. For example, they may believe

that a third party is an entity of several individuals. Additional conditions can further disentangle or control these expectations about data availability. First, information about the monetization abilities of the recipient that does not contain information about the third party can be explored. If individuals decrease their willingness to share data under a recipient's monetization potential alone compared with sharing with many recipients that do not have monetization abilities, then increased exposure is unlikely to be the primary mechanism behind the increase in the value of privacy under secondary market monetization.

Moreover, choices can be examined under high exposure to many data recipients that can *each* engage in secondary monetization. Then, individuals' data-sharing choices can be compared with those in environments with a singular data recipient with the same secondary monetization ability. If exposure to additional parties is the primary driver of disutility from secondary monetization, then there should be a significant decrease in the willingness to share with many recipients under secondary monetization. If such a situation does not occur, then this challenges the importance of exposure concerns in their privacy choices.

Finally, data-sharing choices can be elicited in an environment where the third party is drawn from the same anonymous and indistinguishable pool of recipients. For instance, the choice to release data to many recipients can be compared with the choice to release data to one recipient that transacts that data with one other recipient. The differential exposure concerns can be effectively controlled in a market that disables unsanctioned data trades among recipients. If individuals are less willing to share data with 2 recipients that can benefit through trade with each other than they are to share data with 30 recipients, with no signals about their secondary transactions, then concerns related to differential exposure risks are unlikely to be the driving force behind any revealed aversion to secondary monetization.

3.3. Differential Risk Between Recipients and Third Parties

Secondary market activities can be perceived as riskier depending on the type of exposure. For instance, indirect access by a third party can incur greater perceived instrumental costs for the individual than direct access by a recipient. To explore this mechanism, data-sharing behavior under weaker information or more controlled exposure can be examined. First, the choices made under an understanding of only the recipient's monetization ability are informative. This approach allows signals about the asset goods nature of data to be vivid while keeping information about third-party access opaque. This more narrowly elicits the individual's consumption value of a privacy commodity that prevents the recipient from monetizing the individual's personal data.

In addition, the third party can be made indistinguishable from another data recipient. Following the same method of disentangling exposure concerns from secondary monetization, privacy concerns associated with third parties can be controlled by drawing the third party from the set of potential recipients. Thus, individuals choose how willing they are to share their data when monetization can occur in a secondary market among their data recipients. Perceived privacy risks unique to third parties can be ruled out if individuals reveal disutility from secondary monetization when the risk profile of the third party is identical to that of the recipient.

4. Experimental Design

The online experiment involves a personal data market where individuals generate their psychometric data and face real decisions regarding whether to share those data with others in return for benefits. The study is broadly categorized into three stages. First, in the data generation stage, subjects generate psychometric data that are personally identifiable. Then, in the data-sharing stage, individuals choose to release their data to anonymous recipients in the study in return for monetary benefits. In the final stage, real outcomes from data sharing are realized by both the subjects and the recipients, where the subjects receive earnings from any data sharing and the recipients earn profits from user data through secondary monetization.

4.1. Subjects

The study was conducted at Cornell University's Business Simulation Laboratory (hereafter referred to as the "laboratory"), which maintains an institutional review board-approved subject pool for online studies. One benefit of running a study using this form of research laboratory is the ability to tightly control the setting and implement real trades between subjects. There are credible consequences for subjects as this experiment is similar to other laboratory studies in which people are randomly and anonymously paired to conduct real transactions.[12]

Another reason that the laboratory is well positioned for this study is its high-quality maintenance of the subject pool. Importantly, the personal identifiers of the registrants (e.g., names and emails) are maintained, made available to the researcher prior to data collection, and piped into the subjects' psychometric data generated during the survey. The personalized survey is distributed to each subject via a personalized email. More generally, the laboratory manages verification processes and monitors those who click through a survey without engaging with the content or those with a history of incomplete studies.

This study was advertised with the title "How well do you know yourself? An economic decision study" to avoid priming potential subjects with the idea that the study is meant to examine privacy preferences. In fact, the words "privacy" and "security" are not used for the entirety of the study until the exit survey questions related to privacy attitudes are asked.[13] Additionally, subjects earn a minimum of $2 in Amazon gift cards for completing the 15-minute survey, with the possibility of earning more based on the survey taker's decisions made within the study.

4.2. Data Generation Task

The first stage of the survey involves a 50-item five-factor questionnaire about the respondents' attitudes and habits. These items are taken from the standard sample of Likert-type assessment statements from the International Personality Item Pool, which is widely used in psychology research. All of the items are shown in Online Appendix Figures B.7 and B.8. Responses to these self-assessments generate personality scores across the following five traits: extraversion, agreeableness, conscientiousness, emotional stability, and intellect.[14] After completing all 50 items in the self-assessment, subjects view their personal identifiers, their personality scores, and information about how to interpret their scores (see Figure 1 and Online Appendix Figure B.1). Each score is an integer in the range from 10 to 50. A high score in extraversion, for example, indicates high extraversion and low introversion.

Similar to previous works measuring privacy valuations using the personal data gathered within a study, untruthful information can be contained in the responses. It is not clear how lower data quality (i.e., imperfect measures of a person's personality) ultimately influences the interpretation of these results (a longer discussion and analysis of psychometric data quality is presented in Section 5.3.2). However, importantly, the IPIP items are designed as self-assessments rather than external evaluations of an individual's personality. IPIP items, in particular, have been shown to be successful at eliciting and discriminating personality measures that can predict the real behaviors and traits of individuals (Matz et al. 2017), particularly those in Western and educated populations (Laajaj et al. 2019).

Considerations are made in the experiment to remove any contamination risk to the data-sharing choices in the second stage of the experiment. The self-assessments are elicited at the start of the study without revealing to subjects that the study intends to elicit their willingness to share this psychometric data with other persons in the study. No scores are more economically valuable than other scores are, and this indifference was made explicit to subjects prior to both self-assessment and data-sharing decisions.[15]

4.3. Eliciting Data-Sharing Choices

To elicit honest valuations of the privacy that subjects retain by *not* sharing data, the Becker–DeGroot–Marschak (BDM) (Becker et al. 1964) incentive-compatible method for eliciting willingness to accept is adopted. The BDM method is commonly used in experimental economics to reveal an individual's reservation price for a good, even for commodities without established market prices (List and Shogren 2002). One characteristic of the BDM method is its ability to reveal valuations that are not strategic prices. This feature is critical to this study, where the primary information treatment includes signals that can update subjects' beliefs about the value of data to recipients. The dominant strategy for a subject under the BDM method is to reveal the minimum acceptable price (reservation price) that can correctly reflect the consumption value of privacy from nondisclosure instead of a strategic response to new information about how valuable her data may be to a recipient. If the consumption value of privacy increases because of the secondary monetization ability of a potential recipient, then this is an increase in utility that a subject feels regarding the quality of the privacy commodity.

The implementation of a BDM price elicitation is not without its challenges. The instructions for selecting a minimum price, given an unknown final price drawn from a random distribution, can be unusual and confusing. It is easy to conflate the *minimum* acceptable price with *any* acceptable price. Moreover, data are not sold for explicit prices in the real world but are, rather, implicitly priced with bartered digital services. Thus, it is reasonably challenging for individuals to estimate their reservation prices without prior experience with data prices.

To remedy these challenges, the survey uses a short and coarse list of only five prices ($0.01, $0.49, $0.99, $1.99, and $2.99). Subjects do not need to consider a continuum of prices or estimate the minimum acceptable amount. Rather, they consider a few potential price outcomes and either accept or reject them.[16] The survey taker can, therefore, easily consider each price in isolation and imagine whether she would "take it or leave it" if that price were the final price. Imagining whether one can accept $2.99 in exchange for data

Figure 1. Example Subject's Psychometric Data

First Name	Last Name	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Intellect
Jane	**Doe**	**19**	**31**	**21**	**42**	**48**

sharing is much easier than forming a point estimate of one's minimum price.[17] Online Appendix Figure B.3 displays the price elicitation.

4.4. Data-Sharing Conditions

Privacy choices are elicited under more or less salient signals regarding how personal data can be utilized as an asset rather than simply experienced as a product by a data recipient. Each subject faces four different conditions in randomly assigned order for making data-sharing choices with no prior knowledge of the terms of each condition.[18] One of these conditions is randomly selected, and the subject's decision for this condition is implemented in actuality.

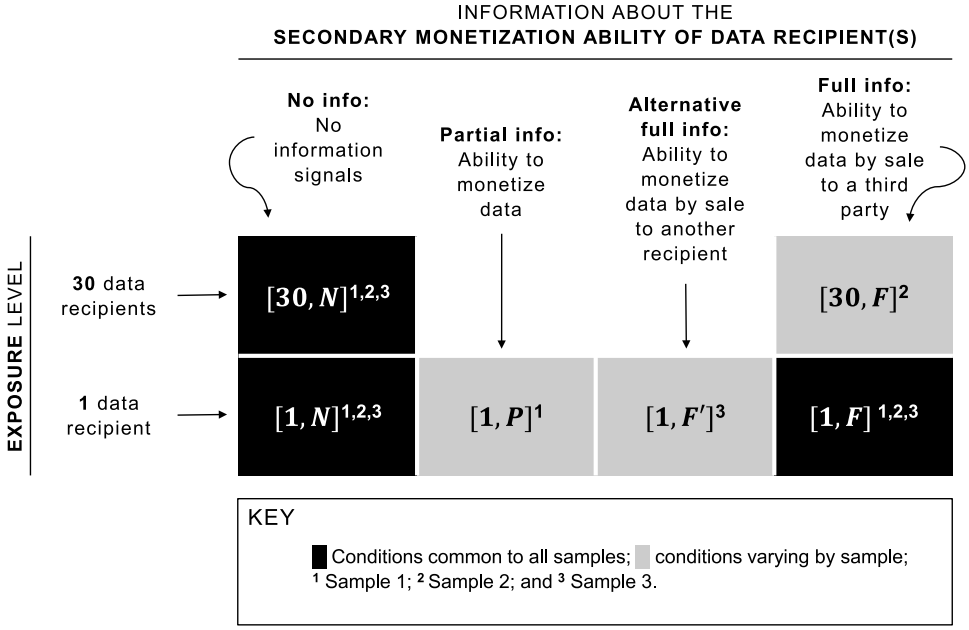
The choices for the conditions $[1,N]$, $[30,N]$, and $[1,F]$ are elicited for *all* subjects in the study. The baseline condition $[1,N]$ is the choice to share data with one recipient, with no signals about secondary monetization. In the main treatment condition $[1,F]$, full information is provided about the recipient's ability to monetize any data that they receive by sale to a third party. Finally, condition $[30,N]$ elicits the value of releasing data to 30 recipients and no provisions about their secondary monetization abilities. Individuals' disutility arising from a recipient's secondary monetization activity can be tested by comparing $[1,N]$ and $[1,F]$. In addition, aversion to secondary monetization can be compared with sensitivity to secrecy concerns (i.e., 30 recipients $[30,N]$ versus 1 recipient seeing their data $[1,N]$) to examine whether indirect exposure to an additional party explains a reluctance to share data under $[1,F]$.

In addition, all of the subjects make one additional data-sharing decision under one of three possible conditions that vary according to the experimental sample: $[1,P]$, $[30,F]$, or $[1,F']$. Each of these sample-varying conditions is designed to test the robustness of conclusions resulting from differences in the privacy choices from the three conditions common to all samples. Figure 2 summarizes the six different data-sharing conditions, and the survey text is provided in Online Appendix Table B.1.

Sample 1 includes condition $[1,P]$, which is a weaker secondary monetization treatment than $[1,F]$ by removing any information about data transfer to a third party and keeping only information about the monetization benefits for the recipient. This $[1,P]$ condition keeps indirect exposure risk as opaque as $[1,N]$ to examine whether an effect under $[1,F]$ is largely explained by concerns about third-party access or whether aversion to secondary monetization persists under $[1,P]$.

Sample 2 includes condition $[30,F]$, which releases data to 30 recipients who can *each* monetize data through trade with a third party. This condition interacts high exposure with secondary monetization signals. If subjects are primarily concerned about others experiencing their personal data as a product through direct or indirect access, then sharing with 30 data recipients with secondary monetization capabilities should significantly reduce subjects' willingness to share. This condition also provides an alternative test of whether salient secondary monetization decreases the willingness to share data by comparing $[30,N]$ and $[30,F]$.

Figure 2. Experimental Conditions



Sample 3 includes condition $[1, F']$, which explicitly defines the third party as another person in the study. This condition serves to control for differences in characteristics between recipients and third parties by drawing the third party from the pool of potential data recipients. Through this condition, the main effects under $[1, F]$ can be disentangled from the explanation that subjects can perceive third parties as riskier forms of exposure than the data recipients that they directly transact with.

4.5. Randomization Groups

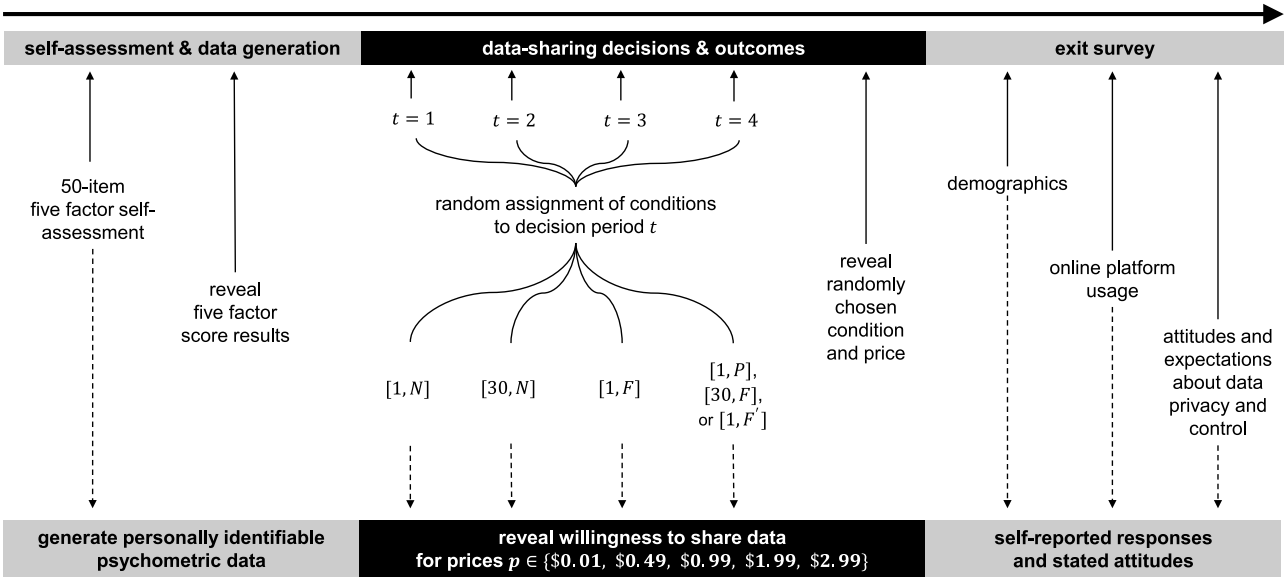
One advantage of a within-subject approach versus a between-subjects design (i.e., randomizing treatments across individual clusters) is that it removes the effect of extraneous subject-level characteristics on the outcome variable. This design is particularly attractive for eliciting people's personal data valuations. As described in Section 2.4, willingness to share data can be highly dependent on the qualities of each person's idiosyncratic data. This naturally leads to multimodal distributions for the value of privacy that are because of differences in the objective value of each person's privacy commodity[19] rather than heterogeneity in the intrinsic tastes for privacy.

Another advantage of a repeated-measures approach is that people's personal data valuations likely rely on *relative* choices (also described in Section 2.4) because most people do not have any prior experience with prices for their personal data. In fact, Birnbaum (1999) warned of a lack of context in between-subjects designs, creating a larger data interpretation issue than the context effects that are naturally present in within-subject designs.[20] Repeated measures in within-subject designs can capture the subject's relative preference rankings between conditions (e.g., "I value my data privacy more in this condition than in the other").

The challenge in implementing a within-subject design is the need to counterbalance order effects. For example, if $[1, F]$ is always the last and most experienced choice, then the resulting effects cannot be orthogonal to the spillovers from previous decisions. To resolve this issue in the experiment, block randomization is performed with a Latin square design to ensure that each condition appears at each ordinal position of a decision in a balanced fashion (see Online Appendix Table B.2). This approach is a standard technique for counterbalancing spillover effects by rendering the outcomes from one condition independent of the order in which they are presented to the respondent. This design also randomizes anchoring effects: for example, whether subjects experience $[1, F]$ before $[1, N]$ or $[1, N]$ before $[1, F]$. In the statistical interpretation of the results, variables related to this exogenously assigned order—including both anchoring and experience—can be controlled for and are independent of the experimental treatments.

Subjects' ex ante knowledge of the repeated-measures design is intentionally vague. Prior to any data-sharing decisions, the subjects are informed that there would be "several" scenarios with different conditions for releasing their data, where one is randomly selected and made real. Then, the realized outcome is revealed directly in the survey after all of the data-sharing choices are made. A summary of the survey chronology from the subjects' perspective is shown in Figure 3. In addition, payments and outcomes are delivered via email one week following survey submission, and subjects are aware of this prior to their data-sharing choices.[21]

Figure 3. Survey Chronology



4.6. Data Collection

The data are collected across three subsamples with multiple waves, with a combined total of 1,188 subjects from 2019 to 2020 (see the details in Online Appendix Table B.3). The survey also includes voluntary questions regarding the respondents' demographic information and social media usage. Approximately 79% of individuals self-report using Facebook with a nonanonymous account and accessing the platform at least once per week. Following this same definition of platform usage, 50.8% and 24.3% of the subjects self-report as users of LinkedIn and Twitter, respectively.[22] Finally, a series of privacy and data ownership attitudes are elicited using a Likert-type assessment of statements related to privacy concerns and data usage by firms (see Online Appendix Table A.9).

5. Results

5.1. Descriptive Summary

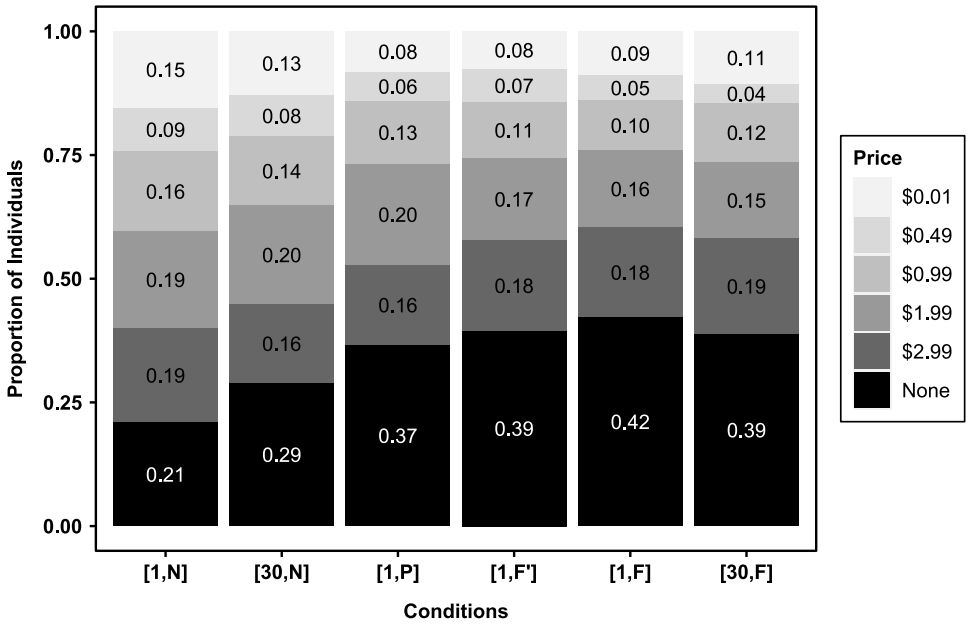
In the baseline condition $[1,N]$, 21% of individuals do not participate in the market for releasing data to one recipient given the available prices and with no information provisioned about secondary market activities of the data recipient. Conditions $[1,F']$, $[1,F]$, and $[30,F]$, which include salient information about the secondary monetization of personal data, approximately double the nonparticipation rate. Between 39% and 42% of individuals refuse to release their data under conditions where they are most aware of a recipient's ability to conduct secondary monetization. Pricing-out behavior is also nearly as prevalent (at 37%) under the weaker information treatment $[1,P]$. Moreover, the sharing of data with 30 recipients under no secondary monetization (condition $[30,N]$) reveals the second weakest privacy response, with 29% not participating in the data market.

This descriptive evidence of the extensive margin not only highlights the importance of secondary monetization in privacy choices but also, highlights that exposure concerns may be an unlikely explanation for the lower instances of data sharing under conditions with salient secondary monetization. Figure 4 summarizes the proportion of individuals who choose to price out of each condition; in addition, reservation prices chosen by participants in each condition are displayed.

By leveraging the within-subject design, changes in data market participation can be observed at the individual level across conditions. The largest share of exits (i.e., a subject participating in the data market under one condition but not another) occurs when subjects are treated with information about secondary monetization. As shown in Figure 5, 23% of individuals are willing to share personal data with *one* recipient ($[1,N]$) but unwilling to share data when treated with information about one recipient's secondary monetization abilities ($[1,F]$).

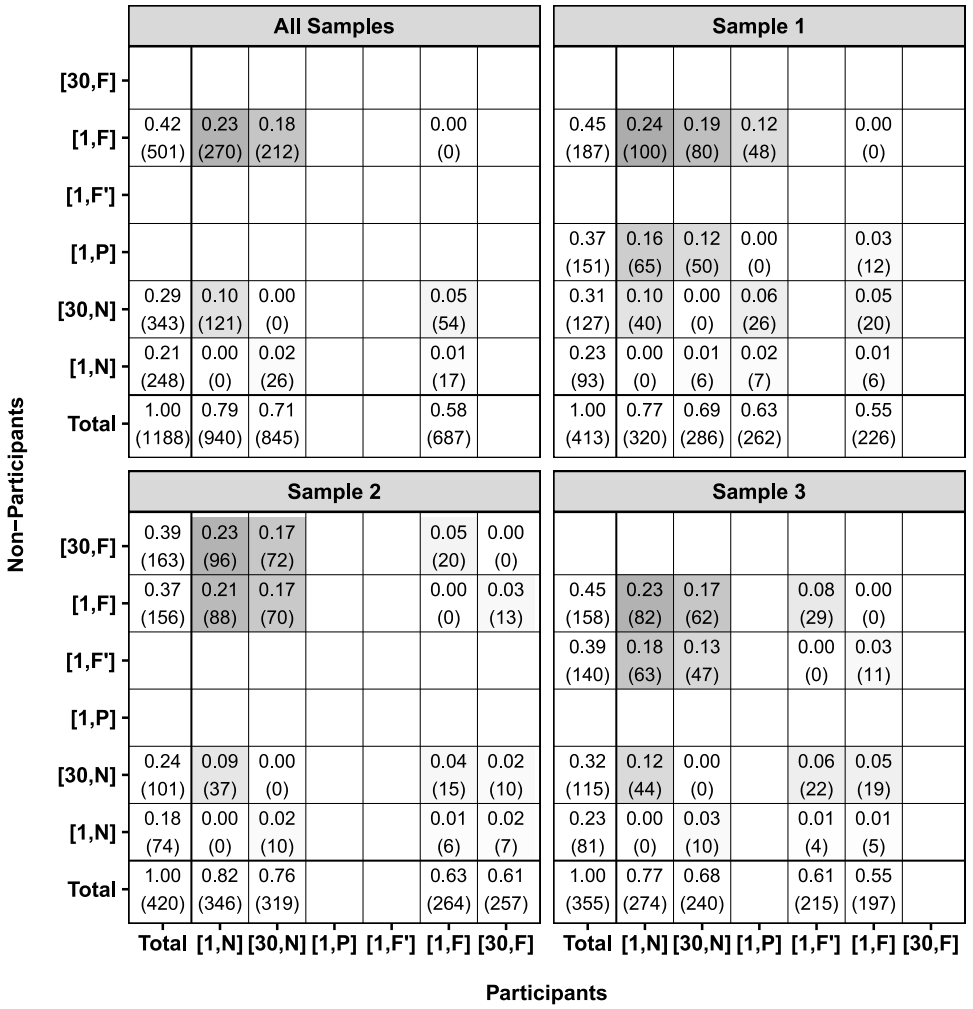
According to these descriptive results of exit behavior, the response to $[1,F]$ is unlikely to be explained by increased exposure or differential risk. When the exposure increases from 1 to 30 recipients (i.e., $[1,N]$ to $[30,N]$), 10% of the subjects exit the data market.

Figure 4. Price Choices by Data-Sharing Condition



Notes. Prices are the minimum willingness-to-accept choices. Conditions $[1,N]$, $[30,N]$, and $[1,F]$ display proportions out of all samples (1,188 individuals). Conditions $[1,P]$, $[30,F]$, and $[1,F']$ display proportions out of their respective samples (413, 420, and 355 individuals, respectively).

Figure 5. Data Market Participation and Nonparticipation by Condition



Non−Participants	All Samples							Sample 1						
	Total	[1,N]	[30,N]	[1,P]	[1,F']	[1,F]	[30,F]	Total	[1,N]	[30,N]	[1,P]	[1,F']	[1,F]	[30,F]
[30,F]														
[1,F]	0.42 (501)	0.23 (270)	0.18 (212)			0.00 (0)		0.45 (187)	0.24 (100)	0.19 (80)	0.12 (48)		0.00 (0)	
[1,F']														
[1,P]								0.37 (151)	0.16 (65)	0.12 (50)	0.00 (0)		0.03 (12)	
[30,N]	0.29 (343)	0.10 (121)	0.00 (0)			0.05 (54)		0.31 (127)	0.10 (40)	0.00 (0)	0.06 (26)		0.05 (20)	
[1,N]	0.21 (248)	0.00 (0)	0.02 (26)			0.01 (17)		0.23 (93)	0.00 (0)	0.01 (6)	0.02 (7)		0.01 (6)	
Total	1.00 (1188)	0.79 (940)	0.71 (845)			0.58 (687)		1.00 (413)	0.77 (320)	0.69 (286)	0.63 (262)		0.55 (226)	

Non−Participants	Sample 2							Sample 3						
	Total	[1,N]	[30,N]	[1,P]	[1,F']	[1,F]	[30,F]	Total	[1,N]	[30,N]	[1,P]	[1,F']	[1,F]	[30,F]
[30,F]	0.39 (163)	0.23 (96)	0.17 (72)			0.05 (20)	0.00 (0)							
[1,F]	0.37 (156)	0.21 (88)	0.17 (70)			0.00 (0)	0.03 (13)	0.45 (158)	0.23 (82)	0.17 (62)		0.08 (29)	0.00 (0)	
[1,F']								0.39 (140)	0.18 (63)	0.13 (47)		0.00 (0)	0.03 (11)	
[1,P]														
[30,N]	0.24 (101)	0.09 (37)	0.00 (0)			0.04 (15)	0.02 (10)	0.32 (115)	0.12 (44)	0.00 (0)		0.06 (22)	0.05 (19)	
[1,N]	0.18 (74)	0.00 (0)	0.02 (10)			0.01 (6)	0.02 (7)	0.23 (81)	0.00 (0)	0.03 (10)		0.01 (4)	0.01 (5)	
Total	1.00 (420)	0.82 (346)	0.76 (319)			0.63 (264)	0.61 (257)	1.00 (355)	0.77 (274)	0.68 (240)		0.61 (215)	0.55 (197)	

Notes. This figure displays the proportions of individuals participating and not participating in each condition based on whether any price is chosen. The numbers of individuals are displayed in parentheses.

However, 18% of individuals are willing to enter the data market under $[30,N]$ but not under $[1,F]$. In contrast, only 5% of the subjects exhibit the reverse behavior (i.e., enter in $[1,F]$ and exit in $[30,N]$). Evidence can also be found in choices under sample-varying conditions. When a recipient's monetization ability is salient and third-party access is opaque, 16% of the individuals exit (i.e., $[1,N]$ to $[1,P]$ in sample 1). Under high-exposure conditions, salient secondary monetization also induces 17% of individuals to exit (i.e., $[30,N]$ to $[30,F]$ in sample 2). When the differential risk between recipients and third parties is controlled for, 18% of individuals exit (i.e., $[1,N]$ to $[1,F']$ in sample 3).

5.2. Estimated Participation and Prices

Given the nature of the price elicitation, acceptable prices are chosen simultaneously with whether to accept *any* price in the available range. Rejecting all available prices allows a subject to exit the data market. Rather than assuming a natural censoring by the price list, the chosen analysis method measures observed behavior in two parts: (1) the odds or likelihood of data market participation by individuals and (2) the price demanded among those who participate.

Unlike many market participation questions in economics that focus on intensive margin changes to a price variable, the extensive margin decision to participate in this data market is the more relevant and unbiased outcome of interest. In reality, data markets ask consumers to make all-or-nothing disclosures in exchange for a set menu of goods and services while rarely asking consumers how much they would be willing to accept in such an exchange. Although measuring price changes is useful for inferences internal to the experiment, it is difficult to map these magnitudes to external contexts, especially where the benefits from data sharing are not monetary amounts and when the

pricing decision is based on the individuals who select into the data market.[23]

First, this approach estimates how the explanatory variables impact data market participation. Whether individual i chooses to participate in the data market in decision t is indicated by

$$Participation_{it} = \begin{cases} 1, & y^*_{it} \le 2.99, \\ 0, & y^*_{it} > 2.99, \end{cases} \quad (1)$$

where y^*_{it} is the true, unobserved minimum acceptable price. Second, for those who participate, the observed minimum acceptable price when $Participation_{it} = 1$ is

$$Price_{it} = \begin{cases} 0.01, & y^*_{it} \le 0.01, \\ 0.49, & 0.01 < y^*_{it} \le 0.49, \\ 0.99, & 0.49 < y^*_{it} \le 0.99, \\ 1.99, & 0.99 < y^*_{it} \le 1.99, \\ 2.99, & 1.99 < y^*_{it} \le 2.99. \end{cases} \quad (2)$$

Because each individual i makes four data-sharing decisions across t decision periods, I use a panel random effects model that corrects for the nonindependence of multiple responses from a single individual (Liang and Zeger 1986):

$$[Participation_{it}, Price_{it}] = \alpha + \boldsymbol{\beta} \cdot \mathbf{Condition}_{it} + \boldsymbol{\delta} \cdot \mathbf{T}_t + \boldsymbol{\gamma} \cdot \mathbf{Y}_i + \theta_i + u_{it},$$

where $\mathbf{Condition}_{it}$ is a set of categorical variables indicating the conditions of interest to be compared with the omitted condition. A set of decision-period controls is denoted as $\mathbf{T}_t$. These are included to capture effects specific to each decision period that can presumably affect all individuals uniformly. The vector $\mathbf{Y}_i$ is a set of individual-specific characteristics. Controls for anchoring effects are included, such as whether the individual is randomly assigned to a no information (N) condition in the first two decision periods and secondary monetization information (P, F, or F') conditions in the last two decision periods.[24]

A set of optional characteristics for examining privacy behavior across different types of individuals is included, such as psychometric scores to explore whether a low or high score in each trait is correlated with relative privacy preferences. Studies find that five-factor scores have predictive power for online behavior (Junglas et al. 2008, Matz et al. 2017, Li et al. 2019), especially for people in Western, democratic, and educated populations (Laajaj et al. 2019). The individual-specific random effect is denoted by θ_i, and u_{it} is the error term. Because all individuals experience conditions that are randomly assigned to a decision period, the estimates are uncorrelated with the observed ($\mathbf{Y}_i$) and unobserved individual differences or error term (θ_i and u_{it}, respectively).

The results tables present the likelihood of participation and the prices demanded in the data market relative to the omitted condition $[1, N]$. There are three specifications for each part. First, the baseline model includes the condition under which a privacy choice is made, controlling for sample and order effects. The second specification extends the first by including a control for anchoring effects. The third specification includes controls for individual-specific characteristics.[25] All errors are clustered at the individual level, and regression specifications are compared using a Wald test to determine whether the inclusion of regressors has meaningful explanatory power. Table 1 presents the estimated impact of the $[1, F]$ and $[30, N]$ conditions on the odds of individuals participating in the data market and the prices demanded.[26]

Table 1. Participation and Prices for Sharing Personal Data (All Samples)

	Logit: *Participation*			OLS: *Price* ($) | *Participation* = 1		
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	1.289***	1.455***	2.089***	1.499***	1.389***	1.140***
	(0.113)	(0.138)	(0.385)	(0.061)	(0.079)	(0.212)
$[30, N]$: 30 recipients, no information	−0.433***	−0.433***	−0.441***	0.102***	0.102***	0.102***
	(0.054)	(0.054)	(0.055)	(0.024)	(0.024)	(0.024)
$[1, F]$: 1 recipient, full information	−1.018***	−1.016***	−1.039***	0.390***	0.393***	0.392***
	(0.064)	(0.064)	(0.065)	(0.033)	(0.033)	(0.033)
Sample controls	×	×	×	×	×	×
Order controls	×	×	×	×	×	×
Anchoring controls		×	×		×	×
Psychometric controls			×			×
Demographic controls			×			×
Individual clusters	1,188	1,188	1,188	975	975	975
Observations	3,564	3,564	3,564	2,472	2,472	2,472
Wald test	(1a), (2a)	(2a), (3a)	(1a), (3a)	(1b), (2b)	(2b), (3b)	(1b), (3b)
$Pr(>\chi^2)$	0.003	0.000	0.000	0.020	0.186	0.041

Notes. The omitted category is $[1, N]$: one recipient, no information. Clustered robust standard errors are in parentheses.
***$p < 0.001$.

I find that secondary monetization increases the disutility of sharing data. Individuals are more likely to not participate and increase prices when information is provided about their recipient's ability to monetize data through sale to a third party. The odds of not entering the data market under condition $[1,F]$ are 2.83 times greater than those under condition $[1,N]$ ($p < 0.001$) (column (3a) in Table 1). A similar pattern is reflected in the intensive margin, where prices under $[1,F]$ are \$0.39 higher than those in $[1,N]$ ($p < 0.001$) (column (3b) in Table 1).

Notably, the magnitude of individuals' aversion to a recipient's ability to transact personal data with a third party ($[1,F]$) is greater than their aversion to directly sharing data with 30 recipients ($[30,N]$). Increasing the number of recipients from 1 to 30 does result in greater disutility from releasing data. The odds of not participating in the data market are 1.55 times greater, and the prices are \$0.10 higher under $[30,N]$ ($p < 0.001$) (columns (3a) and (3b) in Table 1). However, the willingness to share is significantly weaker when a single data recipient's ability to trade with a third party is salient. The odds of individuals not participating in the data market under $[1,F]$ are 1.82 times greater, and the prices are \$0.29 higher than those in $[30,N]$ ($p < 0.001$).

Evidence of a disutility specific to secondary monetization is also found in the various mechanism tests using the conditions varying by sample. The results shown in Table 2 display the conditions common to all samples ($[1,N]$, $[30,N]$, and $[1,F]$) plus a condition tested in each sample ($[1,P]$, $[30,F]$, or $[1,F']$). All of the other model specifications are the same as those in Table 1, aside from the sample controls in these sample-specific results.

In sample 1, individuals are averse to a recipient's monetization ability without being treated with explicit

Table 2. Participation and Prices for Sharing Personal Data (by Sample)

	Logit: *Participation*			OLS: *Price* (\$) | *Participation* = 1		
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
Sample 1						
(Intercept)	1.317***	1.585***	1.176*	1.472***	1.341***	1.468***
	(0.138)	(0.188)	(0.580)	(0.070)	(0.101)	(0.330)
$[30,N]$: 30 recipients, no information	−0.425***	−0.428***	−0.441***	0.092*	0.091*	0.090*
	(0.083)	(0.084)	(0.086)	(0.039)	(0.039)	(0.039)
$[1,P]$: 1 recipient, partial information	−0.682***	−0.681***	−0.701***	0.252***	0.252***	0.252***
	(0.097)	(0.097)	(0.100)	(0.050)	(0.050)	(0.050)
$[1,F]$: 1 recipient, full information	−1.046***	−1.050***	−1.082***	0.428***	0.429***	0.427***
	(0.108)	(0.108)	(0.111)	(0.058)	(0.058)	(0.058)
Sample 2						
(Intercept)	1.554***	2.028***	4.034***	1.412***	1.310***	0.606*
	(0.149)	(0.204)	(0.659)	(0.064)	(0.104)	(0.301)
$[30,N]$: 30 recipients, no information	−0.395***	−0.407***	−0.426***	0.069†	0.069†	0.070†
	(0.098)	(0.102)	(0.106)	(0.036)	(0.036)	(0.036)
$[1,F]$: 1 recipient, full information	−1.007***	−1.021***	−1.081***	0.430***	0.432***	0.432***
	(0.115)	(0.115)	(0.120)	(0.052)	(0.052)	(0.052)
$[30,F]$: 30 recipients, full information	−1.079***	−1.099***	−1.166***	0.395***	0.397***	0.397***
	(0.118)	(0.120)	(0.125)	(0.053)	(0.053)	(0.052)
Sample 3						
(Intercept)	1.284***	1.406***	2.063**	1.551***	1.487***	1.439***
	(0.144)	(0.200)	(0.757)	(0.071)	(0.108)	(0.426)
$[30,N]$: 30 recipients, no information	−0.484***	−0.483***	−0.492***	0.168***	0.168***	0.168***
	(0.102)	(0.102)	(0.103)	(0.048)	(0.048)	(0.048)
$[1,F']$: 1 recipient, alt. full information	−0.789***	−0.785***	−0.800***	0.390***	0.392***	0.391***
	(0.103)	(0.102)	(0.104)	(0.057)	(0.057)	(0.057)
$[1,F]$: 1 recipient, full information	−0.998***	−0.996***	−1.016***	0.362***	0.364***	0.362***
	(0.113)	(0.113)	(0.114)	(0.062)	(0.062)	(0.062)
Order controls	×	×	×	×	×	×
Anchoring controls		×	×		×	×
Psychometric controls			×			×
Demographic controls			×			×

Notes. The omitted category is $[1,N]$: one recipient, no information. Clustered robust standard errors are in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; †$p < 0.1$.

information about third-party sales. As shown in Table 2, the odds of not sharing data under $[1,P]$ are 2.02 times greater, and the prices are \$0.25 higher than those under $[1,N]$ ($p < 0.001$) (columns (3a) and (3b) in sample 1 in Table 2). In fact, under this partial signal about a single recipient's secondary monetization ability, individuals' revealed disutility is still greater than their aversion to directly disclosing to 30 recipients. Participation odds are 1.3 times greater and the prices are \$0.16 higher under $[1,P]$ than those under $[30,N]$ ($p < 0.01$). Finally, exit behavior and prices also increase under $[1,F]$ relative to $[1,P]$, revealing an increase in disutility as the secondary market (or the external nature of the transaction) becomes more salient.

In sample 2, the privacy responses to secondary monetization signals are replicated under direct disclosure to 30 recipients. The odds of not participating are 2.09 times greater and the prices are \$0.33 higher under $[30,F]$ than under $[30,N]$ ($p < 0.001$) (columns (3a) and (3b) in sample 2 in Table 2). Moreover, individuals exhibit little change in privacy behavior under $[1,F]$ versus $[30,F]$, despite dramatically increasing their exposure to more data recipients and third parties. This challenges the notion that exposure concerns or differential risk can explain the magnitudes of the decrease in participation and increase in prices that occur when secondary monetization is salient.

Finally, in sample 3, the aversion to secondary monetization is replicated when controlling for the perception of differential risk between the recipient and the third party. When faced with a decision to share data with two persons in the study (who partake in secondary monetization with each other) versus disclosing to 30 people in the study (without secondary monetization), individuals significantly lower their willingness to share data in the former scenario—despite the third party being drawn from the same pool of potential recipients. The odds of exiting the data market are 1.36 times greater ($p = 0.002$) and the prices are \$0.22 higher ($p < 0.001$) under $[1,F']$ than those under $[30,N]$. There is a small increase in pricing-out behavior under $[1,F]$ than those under $[1,F']$ ($p = 0.004$) that suggests the existence of a privacy concern related to perceived risk differences between a recipient and a third party, but the price gap is statistically inconclusive.

5.3. Robustness Checks

5.3.1. Between-Subjects Comparison of First-Period Choices. Because of the block randomization design, subjects' first-period choices are randomly and evenly assigned across the conditions as a between-subjects experiment. However, first-period decisions have no prior context, so subjects cannot reveal their relative privacy preferences (as previously discussed in Sections 2.4 and 4.5), which can obscure the statistical precision and meaning of the estimates generated from these observations. Despite the expected challenges to first-period choices, the between-subjects comparison reveals a significant reluctance to participate in a data market with salient secondary monetization.

Table 3 shows that by limiting the analysis to only decisions made in the first period, individuals under $[1,F]$ are significantly less likely to participate in the market than are those under $[1,N]$ ($p < 0.05$) (column (3a) in Table 3).[27] However, the difference between the secondary monetization ability of 1 recipient versus releasing data directly with 30 recipients is not detectable during the first period.

The revealed preference differences between conditions become clearer in magnitude and precision with each decision (see Online Appendix Table A.5). Between the first and second decision periods, the block randomization design prevents spillovers from the injection or removal of secondary monetization signals.[28] The difference between $[30,N]$ and $[1,F]$ is significant by the second choice ($p < 0.10$). Therefore, the

Table 3. Between-Subjects Comparison of First-Period Participation and Prices for Sharing Personal Data

	Logit: *Participation*			OLS: *Price* (\$) | *Participation* = 1		
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	1.121***	1.108***	1.567**	1.560***	1.583***	1.533***
	(0.135)	(0.168)	(0.493)	(0.071)	(0.093)	(0.277)
[30, *N*]: 30 recipients, no information	−0.331†	−0.331†	−0.343†	−0.021	−0.020	−0.044
	(0.184)	(0.184)	(0.188)	(0.103)	(0.103)	(0.104)
[1, *F*]: 1 recipient, full information	−0.458*	−0.458*	−0.451*	0.030	0.030	0.011
	(0.182)	(0.182)	(0.185)	(0.104)	(0.104)	(0.105)
Sample controls		×	×		×	×
Psychometric controls			×			×
Demographic controls			×			×
Observations	892	892	892	625	625	625
Adjusted pseudo-R^2	0.001	−0.002	0.001	−0.003	−0.005	−0.009

Notes. The omitted category is $[1,N]$: one recipient, no information. Standard errors are in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; †$p < 0.1$.

disutility from secondary monetization can be disentangled from exposure concerns with more experience and without any spillovers for subjects with or without information about secondary monetization. By the most experienced decision, the between-subjects comparison reveals the largest decrease in individuals' odds of participating under $[1,F]$ relative to $[1,N]$ and $[1,F]$ versus $[30,N]$ ($p < 0.001$) (column (3a) in Table 3 in decision period $t = 4$).[29] This trend supports the notion that individuals make relative choices regarding their willingness to share personal data (Acquisti et al. 2012, Adjerid et al. 2018).

5.3.2. Quality of Psychometric Data. International Personality Item Pool scores based on the five-factor model are designed to be self-assessments; however, self-reporting can generate lower-quality data than those produced by an evaluator.[30] Although objective data quality does not confound the experimental treatments, it does challenge whether the findings of this study can be generalized to other forms of personal data, such as data that are more challenging to manipulate (e.g., credit scores). To examine this, IPIP responses are assessed for internal consistency.

Correlation analysis and factor analysis are used to test the quality of the data (see Online Appendix Figure A.1 and Table A.6). If individuals answer the survey inattentively or randomly, then the results would show a low association among the items intended to measure the same trait, and the underlying factors in a factor analysis would not show any correspondence to the five personality traits. The general pattern of the survey responses shows consistency and reliability, which rules out the prevalence of low-quality, randomly generated psychometric scores.

Understanding the degree to which data quality suffers from inattentive respondents does not identify whether there is a set of individuals who are intentionally curating a false persona. However, how falsified responses interact with disclosure behavior is unclear. First, the experiment's data-sharing opportunity is revealed only after the five-factor survey is completed. Second, inaccurate psychometric data can have both costs and benefits for the subject. For instance, the individual can incur disutility from being misrepresented to others. On the other hand, inaccurate information can obfuscate a person's true personality, which provides a form of secrecy that may be valuable to the individual. An interesting direction for future research is to examine these two possibilities by exogenously assigning objectively true or false data and assessing the unintended consequences of privacy-protecting tools, such as garbling. In addition, a more challenging research direction is to disentangle endogenous motivations for individuals who intentionally falsify their data.

5.4. Limitations

There are several limitations to the interpretation of these results. First, there is a natural difference between for-profit and nonprofit settings. The expectations of many individuals involved in a research study are disassociated from their perceptions of commercial markets. For instance, this study cannot conclude whether individuals are aware of the secondary monetization activities of online retailers. Replicating this study in the field, utilizing personal data traded in real markets, or measuring behavior through real commercial products are all ways to understand how privacy preferences generalize outside a controlled, laboratory market.

Second, there are context effects that come with price elicitation survey questions. Although the relative valuations in this study provide meaningful inferences about privacy preferences, the study's average prices should not be generalized to real market prices for persons' psychometric data. Although the multiple price list elicitation method in this study is intended to minimize this issue, the estimates of reservation prices can vary depending on the lower and upper bounds. One alternative approach is to directly ask subjects to declare a minimum acceptable price. However, as mentioned in the previous sections of this paper, point estimates of reservation prices for sharing data are unnatural, unstable, and difficult for a decision maker to compute.

Third, I cannot correct for selection bias in the valuation of personal data by those who opt in to the data market in this study. Therefore, I cannot discern between those who are naturally censored by the price range and those who exist on an unobserved distribution of prices. For example, \$6.99 could be enough to capture all, some, or none of those who reject \$2.99. However, wider ranges of price lists cannot easily correct this issue because of the aforementioned context effects. Furthermore, extremely high upper-bound prices can be less believable to decision makers, which effectively renders them opt-out choices and thus, difficult for researchers to interpret.

Fourth, a laboratory study cannot completely rule out the potential influence of Hawthorne and experimenter demand effects, despite employing the most important design choice: the inclusion of real outcomes and incentive compatibility. Even though an analysis of between-subjects outcomes confirms the intuition of the main findings, this strategy is not ideal as it rules out the importance of capturing relative privacy preferences over idiosyncratic data.

Finally, individuals in the experiment are not in a state of "digital resignation" (Draper and Turow 2019). As captured by their revealed preferences, subjects believe that the choice to share personal data in the experiment is consequential and that a recipient can neither obtain similar data nor easily conduct secondary

monetization in unsanctioned ways. In reality, individuals may be very aware that their personal data are available through third parties (e.g., data brokers) that have indirectly collected their personal data (James et al. 2017, Barocas and Levy 2020). The infeasibility of unsanctioned data trades is key for internal validity, and the unexpectedness of sanctioned data trades is key for efficacy. However, these design features limit the ecological validity for settings where consumers are fully aware of secondary data markets.

6. Discussion

I find evidence that individuals consistently lower their willingness to share data when it is salient that a data recipient has monetization abilities in secondary markets. This result begs the following question. What are the underlying concerns that drive this change in individuals' value of privacy? The experiment rules out three apparent mechanisms. First, concerns about increasing exposure do not explain the lower willingness to share data. Second, concerns about the greater risk from access by third parties do not explain the effect. These findings are broadly consistent with the idea that access risks yield weak privacy responses. For instance, Buckman et al. (2019) find that the risk of distributing data to a third party as a form of external secondary use yields no change in privacy behavior. Third, using the BDM method, I also rule out strategic responses to new information about the value of data to the recipient.

Although further research is needed to answer this question, the results suggest that secondary monetization of personal data violates some set of ideals that influence privacy preferences. One obvious nonnormative factor relates to fairness. Individuals can experience disutility if they do not obtain a fair share of the profits gained by a recipient from a secondary data exchange. Other factors may be more related to the notion of contextual integrity (Nissenbaum 2009). Individuals may feel that it is unethical for others to profit from information about them, leading individuals to demand greater compensation in exchange for data sharing. Finally, other mechanisms can be specific to the nature of personal data markets, such as data inalienability, which may lead to the perceived ownership of personal information.[31] Despite a decision to trade away data, individuals may still feel ownership over what they consider "their" data and demand data dividends from secondary transactions.

The research design of this study also offers guidance and insight for conducting future experiments on data sharing and privacy valuation. This study accommodates and confirms that all-or-nothing privacy responses are prevalent in data-sharing decisions. A price elicitation that does not include an opt-out choice can miss this distinction in people's intended responses to notice-and-choice regimes. By using prices as an instrument to reveal preferences in a repeated-measures design, researchers can also better capture relative preferences (i.e., preferring to preserve privacy under some conditions more than others). This method accommodates the contextual challenges that individuals have in making point estimates about the value of their data and controls for the idiosyncratic nature of psychometric data that could drive heterogeneous preferences.

The results of this study have important practical, policy, and theoretical implications for personal data markets. Unlike data markets that rely on user data to be collected "freely" by firms or from take-it-or-leave-it data transactions (where individuals forgo all control rights), data markets may find success in treating personal data as inalienable to individuals. For instance, the study provides empirical evidence for the success of using "data vaults" to facilitate data transactions, which has been supported by privacy architecture researchers (Mun et al. 2010) and organizations, such as Solid.[32] The findings also support policies that treat data as consumer possessions that primarily benefit their owners (Arrieta-Ibarra et al. 2018). The results are also consistent with the theoretical benefits of granting data control rights to individuals rather than to firms as individuals can balance their value of privacy with the economic gains from selling their data, whereas firms may choose to hoard and inefficiently use data (Jones and Tonetti 2020).

Overall, this study presents a promising direction for broadening and deepening the research on the privacy preferences that determine individuals' choices concerning their personal data. The findings prove that it would be naive to conclude that people are unconcerned about secondary data markets when studies focus only on access concerns and context-specific usage risks. Instead, how data are made available to the secondary market is important, especially when the method involves the individual's data being economically exploited by others. This study provides rich and consistent evidence demonstrating that privacy is more valuable when there is potential for secondary monetization. The findings are informative for theoretical privacy models and have implications for the design of data markets and privacy policy regimes.

Acknowledgments

The author is grateful to the anonymous associate editor and three anonymous reviewers for their support and contributions to the revisions of this work. For helpful comments and suggestions, the author thanks Vicki Bogan, Chris Forman, Ori Heffetz, David Just, Tobias Kretschmer, Aija Leiponen, Ted O'Donoghue, Marcel Preuss, Alex Rees-Jones, David Tan, and Axel Zeijen as well as seminar participants at the Cornell Behavioral Economics Research

Group; the Consortium on Competitiveness and Cooperation Doctoral Conference; the Innovation, Entrepreneurship, and Technology Brown Bag at Cornell; the Technology and Innovation Management group at Eidgenössische Technische Hochschule Zürich; the Institute for Strategy, Technology, and Organization at Ludwig-Maximilians-Universität München; the Max Planck Institute for Innovation & Competition; and the Workshop on Information Systems Economics.

Endnotes

[1] This privacy choice is formalized in Acquisti et al. (2013), where there is a trade-off between the utility from consuming wealth and privacy.

[2] These data are in the form of five-factor scores and are the same as those obtained by Cambridge Analytica from users' activities on Facebook.

[3] Personal identification can also occur through the reverse engineering of identifiers (Abowd and Schmutte 2019).

[4] Consistent with the data privacy literature across disciplines and fields (e.g., Nissenbaum 2009, Acquisti et al. 2016, Jones and Tonetti 2020, Koutroumpis et al. 2020) and the European General Data Protection Regulation (GDPR), the privacy choice studied in this work is related to the activities surrounding this definition of personal data.

[5] Data produced in an economy "[feed] back" and make all firms more productive in a virtuous cycle of productivity and data (p. 2832).

[6] Prior studies examine individuals' disclosure of identifying or sensitive content (e.g., email addresses and medical history) (John et al. 2010, Athey et al. 2017, Buckman et al. 2019). However, identification is becoming easier and less expensive. For example, Acquisti and Gross (2009) shows how Social Security numbers can be predicted from publicly available data. On the other hand, Glasgow and Butler (2017) find that personal (or unique) identification is a required feature of the ability of shared personal data to invoke privacy concerns, despite the decreasing commercial value of identification.

[7] The public domain International Personality Item Pool (IPIP), which is used for five-factor personality measurement, is pervasively used in Western, educated, industrialized, rich, and democratic (WEIRD) nations for various research and assessment purposes, contributing to its predictive and persuasive power over human behavior and attitudes for WEIRD populations (Goldberg et al. 2006, Laajaj et al. 2019).

[8] Sometimes, the literature refers to this as subjective (psychological) versus objective privacy harms. For example, Calo (2011) largely conceptualizes subjective privacy harm as the undesirable observation by others rather than as a psychological aversion to personal data being used by others. However, privacy can still be an intermediate good for other psychological benefits (e.g., a desire for control or self-determination).

[9] Even under the European GDPR provisions on the right to be forgotten, the nonexcludable nature of data makes the enforcement of erasure rights for personal data difficult and costly for both individuals and firms.

[10] This paper describes each component as a disutility or privacy loss (positive v and o). However, this framework does not restrict individuals from gaining *positive* utility from sharing their personal data. Both components generalize to nonnegative utility (i.e., an individual enjoys being observed (negative v) or having her personal data utilized in secondary markets (negative o)).

[11] The framework allows o to depend on the number of recipients that can benefit from data being utilized as an asset. However, this does not exclude the possibility that an individual primarily suffers from the existence of activities in secondary markets and remains insensitive to how many data recipients can benefit from secondary monetization.

[12] These practices include no deception and transparency in the terms and conditions of data transactions. The subjects are not able to access the list of study registrants, thus preventing unsanctioned data sharing with other subjects. Unsanctioned data monetization by those randomly selected to receive personal data (i.e., data recipients) is also unlikely given the lack of commercial opportunities for consumers to sell individual or small-scale personal data in secondary data markets.

[13] Adjerid et al. (2019) shows that individuals have a greater propensity toward privacy outcomes when prompted to make decisions about their "privacy" settings versus their "survey" or "app" settings.

[14] All self-assessment statements require a response. Skipped questions prevent the survey from continuing to the next page and the later stages of the survey.

[15] Subjects are told at the start of the self-assessment that the bonus payment amounts are unrelated to this stage in the survey (e.g., "Your responses for each statement will NOT determine your earnings in this study").

[16] The price range is chosen based on pilot surveys of this study. The prices ending in 49 and 99 cents are commonly advertised formats for prices and are designed to evoke familiarity in subjects' minds with other digital goods and services (e.g., paid mobile apps and online subscription services). The \$0.01 choice is used in place of a \$0.00 choice to prevent any nonnormative factors from influencing this decision (i.e., there may be a special reluctance to choose to give away something for free, even if a subject values her privacy at a nonpositive price).

[17] Price-reversing behavior in a multiple-price list does occur (e.g., accepting \$1.99 while rejecting \$2.99) either by mistake or because of uncertainty. However, the survey is not designed to prevent this behavior, allowing these forms of inattention, incomprehension, or indecision into the data. A small number of subjects exhibited some form of price-switching behavior, and they are included in the analysis based on the minimum price that they accept regardless of inconsistency with higher rejected prices. The preregistration also does not specify the exclusion of these subjects.

[18] The number of decision rounds is limited to four to reduce inattention and survey fatigue.

[19] This is empirically reflected, for example, in Collis et al. (2021).

[20] In Birnbaum (1999), respondents in a between-subjects study could rate the number 9 as having the same magnitude as the number 221, whereas this issue disappears in a within-subject design, where 9 is rated as relatively smaller than 221.

[21] Study registrants are randomly selected to be data recipients. Data recipients are not given any option to self-select into their roles, any additional choices, or any reasonable possibility of unsanctioned data monetization upon receiving the personal data of others. An example of the data delivery process is shown in Online Appendix Figure B.9. The role of the data recipient is not advertised in the study and is imposed only on a randomly selected set of people registered for the study. The profits from sanctioned secondary monetization are determined by an undisclosed rate of return on the number of subjects in their survey wave who released personal data.

[22] A large minority of respondents report high-frequency engagement with these platforms. For example, 19% are Facebook users who post or share content on the platform at least once a week, 39.1% are LinkedIn users who respond to requests for connections within a week, and 12.3% are Twitter users who post or share content at least weekly. Approximately 72% of participants self-identify as female, their mean reported age is 23.8 years (standard

deviation = 6.95), and more than 70% report being students as opposed to being employed (full or part time). This study's demographic makeup is similar to that of other studies conducted using this laboratory's subject pool.

[23] Although I recognize that there exist behavioral mechanisms that influence data market participation separately from those prices demanded (as opposed to natural censoring), the identification of these mechanisms is beyond the scope of this study. A Heckman-style selection model is not used given the lack of a valid exclusion restriction. Therefore, a two-part estimation—which recognizes but does not correct for selection bias—is the preferred style of inference for this study.

[24] Additionally, whether the individual experiences 30 recipients before or after 1 recipient is included.

[25] The demographic controls include gender, marital status, Facebook usage, and employment status. Psychometric controls include whether the individual scored above 30 (on a scale from 10 to 50) in each of the five-factor traits: extraversion, agreeableness, conscientiousness, emotional stability, and intellect.

[26] Suppose C is the focal condition and $\overline{C}$ is the omitted condition; then, the odds ratio of C and $\overline{C}$ is $\exp[\hat{\beta}] = \exp[\log(\text{odds}C/\text{odds}\overline{C})] = \text{odds}C/\text{odds}\overline{C}$, where $\hat{\beta}$ is the coefficient estimate of the condition of interest. The change in prices demanded is estimated among the individuals who participate in the data market under the omitted and focal conditions.

[27] The price differences between those who select into one condition and those who select into another condition are not detectable in the first period; however, differential selection into each condition confounds a meaningful between-subject comparison of the intensive margin.

[28] Subjects with $[1, F]$ at $t = 2$ experienced $[1, P]$, $[30, F]$, or $[1, F']$ at $t = 1$ and vice versa. Subjects with $[30, N]$ in $t = 2$ experienced $[1, N]$ in $t = 1$ and vice versa. See Online Appendix Table B.2.

[29] The intensive margin choices (for those who participate under each condition) also become precise and significantly different.

[30] For example, a psychological analysis of each subject by an expert, perhaps in a field setting or over a long period, would curate a superior-quality data set of individuals' psychometric information.

[31] A recent discussion by Morewedge et al. (2021) calls for empirical research about the role of psychological ownership in the digital economy. The self-identification antecedent to the psychological ownership of personal data is more specifically explored in Spiekermann et al. (2013).

[32] This project is led by Tim Berners-Lee (inventor of the World Wide Web) for individuals to store and control their data (https://solidproject.org/).

References

Abowd JM, Schmutte IM (2019) An economic analysis of privacy protection and statistical accuracy as social choices. *Amer. Econom. Rev.* 109(1):171–202.

Acquisti A, Gross R (2009) Predicting social security numbers from public data. *Proc. Natl. Acad. Sci. USA* 106(27):10975–10980.

Acquisti A, Grossklags J (2005) Privacy and rationality in individual decision making. *IEEE Security Privacy* 3(1):24–33.

Acquisti A, Grossklags J (2008) What can behavioral economics teach us about privacy. Acquisti A, Gritzalis S, Lambrinoudakis C, De Capitani di Vimercati S, eds. *Digital Privacy: Theory, Technologies, and Practices* (Taylor & Francis Group, New York), 363–374.

Acquisti A, Brandimarte L, Loewenstein G (2015) Privacy and human behavior in the age of information. *Science* 347(6221):509–514.

Acquisti A, John L, Loewenstein G (2012) The impact of relative standards on the propensity to disclose. *J. Marketing Res.* 49(2):106–174.

Acquisti A, John L, Loewenstein G (2013) What is privacy worth? *J. Legal Stud.* 42(2):249–274.

Acquisti A, Taylor C, Wagman L (2016) The economics of privacy. *J. Econom. Literature* 54(2):442–492.

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

Adjerid I, Peer E, Acquisti A (2018) Beyond the privacy paradox: Objective vs. relative risk in privacy decision making. *MIS Quart.* 42(2):465–488.

Adjerid I, Acquisti A, Brandimarte L, Loewenstein G (2013) Sleights of privacy: Framing, disclosures, and the limits of transparency. *Proc. Ninth Sympos. Usable Privacy Security (SOUPS'13)* (Association for Computing Machinery, New York), 1–11.

Arrieta-Ibarra I, Goff L, Jiménez-Hernández D, Lanier J, Weyl EG (2018) Should we treat data as labor? Moving beyond "free." *AEA Papers Proc.* 108:38–42.

Athey S, Catalini C, Tucker C (2017) The digital privacy paradox: Small money, small costs, small talk. NBER Working Paper No. 23488, National Bureau of Economic Research, Cambridge, MA.

Barocas S, Levy K (2020) Privacy dependencies. *Washington Law Rev.* 95(2):555–616.

Becker G, DeGroot M, Marschak J (1964) Measuring utility by a single-response sequential method. *Behav. Sci.* 9(3):226–232.

Birnbaum M (1999) How to show that 9 is greater than 221: Collect judgments in a between-subjects design. *Psych. Methods* 4(3):243–249.

Brandimarte L, Acquisti A, Loewenstein G (2012) Misplaced confidences: Privacy and the control paradox. *Soc. Psych. Personality Sci.* 4(3):340–347.

Buckman JR, Bockstedt JC, Hashim MJ (2019) Relative privacy valuations under varying disclosure characteristics. *Inform. Systems Res.* 30(2):375–388.

Calo RM (2011) The boundaries of privacy harm. Preprint, submitted April 28, https://ssrn.com/abstract=1641487.

Collis A, Moehring A, Sen A, Acquisti A (2021) Information frictions and heterogeneity in valuations of personal data. Preprint, submitted December 2, https://ssrn.com/abstract=3974826.

Culnan MJ (1993) "How did they get my name?": An exploratory investigation of consumer attitudes toward secondary information use. *MIS Quart.* 17(3):341–361.

Culnan MJ, Armstrong PK (1999) Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organ. Sci.* 10(1):104–115.

DellaVigna S (2009) Psychology and economics: Evidence from the field. *J. Econom. Literature* 47(2):315–372.

Dinev T, Hart P (2006) An extended privacy calculus model for e-commerce transactions. *Inform. Systems Res.* 17(1):61–80.

Draper NA, Turow J (2019) The corporate cultivation of digital resignation. *New Media Soc.* 21(8):1824–1839.

Farrell J (2012) Can privacy be just another good? *J. Telecomm. High Tech. Law* 10(2):251–264.

Glasgow G, Butler S (2017) The value of non-personally identifiable information to consumers of online services: Evidence from a discrete choice experiment. *Appl. Econom. Lett.* 24(6):392–395.

Goldberg LR (1992) The development of markers for the big-five factor structure. *Psych. Assessment* 4(1):26–42.

Goldberg LR, Johnson JA, Eber HW, Hogan R, Ashton MC, Cloninger CR, Gough HG (2006) The International Personality Item Pool and the future of public-domain personality measures. *J. Res. Personality* 40(1):84–96.

Graham-Harrison E, Cadwalladr C (2018) Revealed: 50 million Facebook profiles harvested for *Cambridge Analytica* in major data breach. *The Guardian* (March 17), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Hill K (2020) The secretive company that might end privacy as we know it. *New York Times* (January 18), https://www.nytimes.

com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

Hui KL, Teo HH, Lee SYT (2007) The value of privacy assurance: An exploratory field experiment. *MIS Quart.* 31(1):19–33.

James TL, Wallace L, Warkentin M, Kim BC, Collignon SE (2017) Exposing others information on online social networks (OSNs). *Inform. Management* 54(7):851–865.

John LK, Acquisti A, Loewenstein G (2010) Strangers on a plane: Context-dependent willingness to divulge sensitive information. *J. Consumer Res.* 37(5):858–873.

Jones CI, Tonetti C (2020) Nonrivalry and the economics of data. *Amer. Econom. Rev.* 110(9):2819–2858.

Junglas IA, Johnson NA, Spitzmüller C (2008) Personality traits and concern for privacy: An empirical study in the context of location-based services. *Eur. J. Inform. Systems* 17(4):387–402.

Koutroumpis P, Leiponen A, Thomas LDW (2020) Markets for data. *Indust. Corporate Change* 29(3):645–660.

Laajaj R, Macours K, Pinzon Hernandez DA, Arias O, Gosling SD, Potter J, Rubio-Codina M, Vakis R (2019) Challenges to capture the big five personality traits in non-weird populations. *Sci. Adv.* 5(7):eaaw5226.

Laufer R, Wolfe M (1977) Privacy as a concept and a social issue: A multidimensional developmental theory. *J. Soc. Issues* 33(3):22–42.

Li Y, Huang Z, Wu YJ, Wang Z (2019) Exploring how personality affects privacy control behavior on social networking sites. *Frontiers Psych.* 10(1771):1–9.

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22.

Lin T (2022) Valuing intrinsic and instrumental preferences for privacy. *Marketing Sci.* 41(4):663–681.

List JA, Shogren JF (2002) Calibration of willingness-to-accept. *J. Environ. Econom. Management* 43(2):219–233.

Matz SC, Kosinski M, Nave G, Stillwell DJ (2017) Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl. Acad. Sci. USA* 114(48):12714–12719.

McCrae RR, John OP (1992) The five-factor model: Issues and applications. *J. Personality* 60(2):175–215.

Miller AR, Tucker C (2018) Privacy protection, personalized medicine, and genetic testing. *Management Sci.* 64(10):4648–4668.

Morewedge CK, Monga A, Palmatier RW, Shu SB, Small DA (2021) Evolution of consumption: A psychological ownership framework. *J. Marketing* 85(1):196–218.

Mun M, Hao S, Mishra N, Shilton K, Burke J, Estrin D, Hansen M, Govindan R (2010) Personal data vaults: A locus of control for personal data streams. *Proc. 6th Internat. Conf. (Co-NEXT '10)* (Association for Computing Machinery, New York), 1–12.

Nissenbaum H (2009) *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press, Stanford, CA).

O'Donoghue T, Rabin M (1999) Doing it now or later. *Amer. Econom. Rev.* 89(1):103–124.

Spiekermann S, Korunovska J, Bauer C (2013) Psychology of ownership and asset defense: Why people value their personal information beyond privacy. Preprint, submitted January 22, https://ssrn.com/abstract=2148886.

Sutanto J, Palme E, Tan CH (2013) Addressing the personalization-privacy paradox. *MIS Quart.* 37(4):1141–1164.

Tomaino G, Wertenbroch K, Walters DJ (2023) Intransitivity of consumer preferences for privacy. *J. Marketing Res.* 60(3): 489–507.

Tsai JY, Egelman S, Cranor L, Acquisti A (2011) The effect of online privacy information on purchasing behavior: An experimental study. *Inform. Systems Res.* 22(2):254–268.

USAspending (2023) Recipient profile: Clearview AI, Inc. Accessed August 31, 2023, https://www.usaspending.gov/recipient/8cbe9b49-79af-44f4-3eb5-87296ee4a591-C/latest.

Varian HR (1996) Economic aspects of personal privacy. Privacy and self-regulation in the information age. Lehr WH, Pupillo LM, eds. *National Telecommunications and Information Administration Report*. Reprinted in *Internet Policy and Economics: Challenges and Perspectives* (Springer, New York).

Xu H, Teo HH, Tan BC, Agarwal R (2010) The role of push-pull technology in privacy calculus: The case of location-based services. *J. Management Inform. Systems* 26(3):135–174.