# Secondary Market Monetization and Willingness to Share Personal Data*

Joy Wu
Institute for Strategy, Technology and Organization
Ludwig-Maximilians-Universität München (LMU Munich)
joy.wu@lmu.de

November 5, 2022

Individuals who generate user data are often unaware of how their data can be used as valuable inputs into economic activities in secondary data markets. Little is understood about whether secondary monetization is a determinant of users' data privacy preferences. I examine whether individuals' willingness to share data—both in their likelihood of participating in the data market and the prices demanded for such participation—are motivated by recipients' ability to benefit from trading user data with a third party. A large, online lab experiment involving users' personally identifiable psychometric data was implemented with real data-sharing consequences and monetary benefits. I find that individuals decrease their willingness to share data when their recipient's ability to monetize data in a secondary trade is salient. Strategic responses to updated beliefs about a recipient's gain from the trade are ruled out with the chosen price elicitation. I test and find that increased data exposure (to more recipients) does not explain the significant, revealed disutility from secondary monetization. Moreover, these findings are robust to controlling for risk exposure differences between primary and secondary market recipients.

*Key words*: data privacy; personal data; data monetization; user-generated data; data markets; economic experiment

A 2021 version of this paper was titled "Privacy-Seeking Behavior in the Personal Data Market."

---

*This project was funded by a seed grant from Cornell University's Institute for the Social Sciences. The study received IRB approval from Cornell University and was pre-registered at the American Economic Association's registry for randomized controlled trials (AEARCTR-0004005). Special thanks to Aija Leiponen, Tobias Kretschmer, Ted O'Donoghue, Chris Forman, David Just, and Vicki Bogan. I also thank seminar participants in the Cornell Behavioral Economics Research Group; Applied Economics and Management at Cornell; Consortium on Competitiveness and Cooperation Doctoral Conference; the Technology and Innovation Management Group at ETH Zürich; and the Institute for Strategy, Technology and Organization at LMU Munich. All errors are my own.

## 1. Introduction

Recent events in the digital economy have raised questions about whether individuals' data-sharing choices can account for their preferences over activities in secondary data markets. In 2018, the *Cambridge Analytica* scandal involved a personality survey hosted on the Facebook platform that was used to curate users' psychometric data (Graham-Harrison and Cadwalladr 2018). The information contained in these data predicted marginal voters in the 2016 U.S. presidential campaign and even targeted those voters with persuasive information to induce shifts in their voting behavior. In 2020, a new firm, *Clearview AI*, harvested user-published facial images to fuel a successful facial recognition technology (Hill 2020). By 2022, the U.S. Federal Bureau of Investigation entered into a license with *Clearview AI* (*USAspending.gov* 2022). Furthermore, data can even be utilized *without* enabling direct access—by running aggregate statistics queries on consumer datasets harvested by data brokers, such as *Acxiom*. These examples motivate discourse on whether users would demand privacy regulations over the use of their data as inputs into commercial activities.

In today's data economy, secondary monetization of user data is a nearly ubiquitous operational feature of data markets. After individuals make an initial decision to share data with a data recipient (i.e., the primary market), their data can be traded and monetized by such a recipient with third-party actors (i.e., the secondary market). Once personal data enter secondary markets, the average internet user cannot partake in or control the economic activities involving his or her data. Since personal data are linked to a person's identity and contain information about that person, users may have preferences over how others utilize their data. For instance, it is possible that a user may feel continued ownership over the profile data she shares with a social media platform and that she is owed—what she perceives as—a fair share of a platform's profits accrued from her data. In another example, a user might believe it is fundamentally amoral for someone other than himself to economically benefit from his facial imaging data; as a result, he is less willing to share such data or demand more compensation if these ideals are violated.

To capture whether secondary markets negatively affect users, it is important to capture the "consumption value" of privacy afforded by users when they do *not* trade their data with a recipient in the primary market.[1] If a recipient's ability to economically benefit from user data through secondary trades raises a user's value of data privacy and, equivalently, increases the privacy cost from data-sharing, then this should be reflected in the user being less willing to share data with such a recipient. In this paper, I isolate and examine whether the willingness of users to share their data is influenced by the data recipient's secondary monetization ability.

[1] This privacy choice is formalized in Acquisti et al. (2013), where there is a tradeoff between the user's utility from consuming wealth and privacy.

Using an economic experiment, I test whether increasing the salience of a data recipient's ability to monetize data in a secondary market causes individuals to be less likely to share data and demand greater benefits in exchange for data-sharing. After individuals in the study generate personally identifiable psychometric data,[2] they face decisions in the form of a primary monetization choice to share data with a recipient. The study reveals individuals' relative privacy preferences over various data-sharing conditions with differential signals about the consequences of primary monetization. The data-sharing and secondary monetization consequences are real: data recipients are anonymous and randomly selected from individuals recruited for the study to earn profits from the data they obtain. I elicit choices that are honest valuations of the privacy afforded by *not* sharing data.

The results of the experiment show consistent evidence that a data recipient's ability to profit by trading data with a third party decreases users' willingness to share these data and, equivalently, increases users' value of keeping data from being released in a primary monetization choice. This revealed aversion to secondary monetization is present in both extensive and intensive margins: a decreased likelihood to participate in the data market (by not sharing data at all) and, among those willing to share, an increase in the reservation price of data-sharing. Further, the experiment explores and finds evidence that concerns about greater data exposure—e.g., a disutility due to data being made available to more people—do not explain the decrease in willingness to share data due to secondary monetization. In addition, I test and find that any differential risk of releasing data to parties in the primary market versus secondary markets—e.g., an aversion to access by "third parties"—also does not explain users' reluctance to share data with recipients capable of secondary monetization.

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, experimental research on privacy is primarily focused on users' preferences over allowing others to *access* their data, especially in their primary transactions with a data recipient (John et al. 2010, Tsai et al. 2011, Acquisti et al. 2013, Athey et al. 2017, Adjerid et al. 2019). Recent work also explores whether sharing data with a recipient is influenced by the risk of data being made available for third parties to access (Buckman et al. 2019). By predominantly focusing on access concerns, previous research has not fully captured the externalities of secondary market activities and their impact on users' willingness to share and sell personal data.

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

This work also empirically challenges privacy models that equate privacy to the value of secrecy—an outdated premise, particularly for personal data markets. Anti-surveillance sentiments being the core of individuals' privacy preferences (as described by Westin 1967) has been theoretically contested. Nissenbaum (2009) stipulates that what people care about most is *not* secrecy and restricting information flows but rather "ensuring [information] flows appropriately." My experiment contributes to empirical works that support more expansive models of privacy. Brandimarte et al. (2012), in their exploration of preferences for control, frames their study around the notion that "access" and "use" of personal information are two different dimensions of privacy considerations, in which privacy costs depend both on whether people have access to the data and, if so, what they can do with such information. Other works identify various underlying components of privacy preferences, including intrinsic versus instrumental factors by Lin (2020) or a larger combination of nonnormative factors in Acquisti et al. (2015). In this study, I put forward a framework for separating privacy preferences over data being experienced as a final good (e.g., for others to "see" or "know") versus utilized as an intermediate good (e.g., for others to use or trade as a resource in other activities).

This experiment also assumes, relies upon, and finds evidence that users have imperfectly informed choices, which contributes to the stream of work demonstrating the information disadvantage people have when making valuations about their data. Consistent with prior behavioral research, my results imply that individuals are vulnerable to sub-optimal data-sharing choices in online settings. Studies have shown that individuals' valuations of personal data are highly uncertain and unstable (Acquisti et al. 2016, Tomaino et al. 2021). Recently, Collis et al. (2021) find that real-world information about the market value of their data motivates people to revise the prices they demand in return for selling their data, with differential consequences for more or less economically vulnerable populations. My study also contributes to improving the methodology of experiments that elicit valuations of personal data. I accommodate for pricing-out behavior common in real-world data-sharing choices, where benefits are usually small and all-or-nothing decisions. My experiment implements a repeated-measure design that balances out carry-over effects, revealing relative privacy preferences that account for the idiosyncratic nature of personalized data and lessen individuals' contextual challenges when forming valuations.

In summary, this study empirically isolates a privacy cost due to data recipients' secondary market monetization capabilities, disentangled from disutility due to data access risks. This work extends empirical privacy research beyond individuals' secrecy concerns and into notions about whether they find it appropriate for a recipient to benefit from user data in secondary markets inaccessible to the user. I contribute to the notion that preferences over sharing data with features of an intermediate good can differ from sharing data as a final good. The research design of this

study motivates improvements to future works that elicit privacy preferences using prices as the instrument. Finally, the impact of secondary market activities on users can be easily overlooked, since users often lack the information to make privacy decisions that account for their preferences related to the operational features of these markets.

## 2. Background

### 2.1. Inalienability, Non-Rivalry, and Intermediate Goods

Individual data are characterized by their *inalienability* (coined by Koutroumpis et al. 2019): personal information that permanently refers to a specific individual. For user-generated data, the individual contributes to digitizing information related to his or her attitudes, attributes, tastes, and behaviors. These contributions can be more passive, such as exhaust data collected as a byproduct of other activities (e.g., internet browsing behavior and location tracking). More active data-generation activities include user-generated content on information platforms (e.g., comments and likes) or responses to online surveys (e.g., personality tests and consumer surveys). All these data are typically considered personal data, in the sense that they are connected to or identifiable[3] to the originator and inalienable data creator.[4] Due to inalienability, users may always feel perceived or psychological ownership[5] over their data, even after they trade them away in return for goods and services.

Another critical aspect of data is their *non-rivalrous* nature—e.g., utilizing some data does not prevent another entity from doing the same—enabling secondary markets for user data with increasing returns (Jones and Tonetti 2020).[6] This feature prevents users from maintaining control over their data in secondary markets, and users' agency over their data often starts and ends with generating and releasing that data to a recipient. In practice, users may face a series of cascaded choices from an "upstream" choice to join a platform—or select their privacy settings upon entering—before performing "downstream" data generation for a recipient (Adjerid et al. 2019). However, individuals are usually excluded from economic participation in further downstream transfers of their data: they cannot influence the manner in which their data flow in these secondary markets. Due to this non-rivalrous nature of data allowing for—theoretically—infinite transfers,

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

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

[5] A recent discussion by Morewedge et al. (2021) calls for more empirical research about the role of psychological ownership in the digital economy.

[6] Data produced in an economy "feeds back" and makes all firms more productive in a virtuous cycle of productivity and data.

users' preferences for whether and how their data are accessed and utilized in various commercial markets should be examined.

Finally, data are not often final goods (i.e., information "consumed" and experienced by a recipient through access alone) but rather *intermediate goods* (i.e., inputs into a larger commercial activity). Once released to a recipient, data can be traded repeatedly; however, the originator of the data no longer receives the surplus accrued from their data in secondary transactions. This situation leads to a valid debate over how and whether consumers should receive "data dividends" (Arrieta-Ibarra et al. 2018) when they supply those personal data monetized in secondary markets.

Seminal privacy research in information systems has documented an association between individuals' privacy attitudes and the secondary use of information. According to Culnan (1993), those less concerned about secondary use are associated with less concern 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. Moreover, recent experiments feature data with final and intermediate goods characteristics. Athey et al. (2017) measures the willingness of people to avoid surveillance—i.e., framing personal data as a final good—and largely finding weak revealed preferences (contrary to their stated concerns). Buckman et al. (2019) frames personal data as an intermediate good, examining whether being informed about the intended secondary use of data (by distributing them to third parties) influences privacy choices but finds no evidence of changes in data-sharing behavior. The findings from these studies provide some clues about users' preferences in relation to the nature of data markets.

Weak revealed preferences are found in studies focused on exposure concerns (or to how many parties the data are "flowing"), such as users' value of preventing surveillance or their tolerance for data being distributed to 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 2020)—making it difficult to isolate whether the operational features of the secondary data market are important for privacy concerns. Therefore, little is known about whether users may find certain *forms* of data transfer, such as primary versus secondary monetization, more appropriate than other forms.

Personal data's inalienability to the originating user, their non-rivalry, and intermediate goods nature bolster the importance of broadening empirical privacy research to understand the role of users as the suppliers of data inputs into digital economies. This study contributes to the economics of data markets by examining users' privacy preferences over their inalienable data as an intermediate good, which can be utilized for profit by their recipient in secondary trades. This work extends privacy choices beyond data-availability transactions (i.e., in allowing others' access) and considers the influence of those secondary transactions that benefit their data recipient.

### 2.2. Unique Privacy Concerns for Secondary Markets for User-Generated Data

There is growing theoretical and empirical grounding that privacy preferences are components of various, more primitive ideals. This is most salient when privacy economics research examines privacy as a commodity that can be consumed by individuals, with various attributes that can be valued more or less than others. These attributes can make privacy seem like a final good (i.e., treating privacy as an asset to be traded away) or intermediate good (i.e., privacy is the freedom to decide whether to trade data). For example, revealed privacy choices of individuals (e.g., the decision not to share data) may be determined by preferences for control over information-sharing and usage rights (Westin 1967, Brandimarte et al. 2012, Acquisti et al. 2016). Preferences can also be determined by an intrinsic taste for privacy separate from the instrumental taste (Lin 2020).

A user's aversion to his or her recipient's ability to conduct secondary data monetization (i.e., treating data as an intermediate good) has elements of treating privacy as both an intermediate and final commodity. First, a loss of control and self-determination can explain an aversion to secondary transfers of their data by others. Second, the loss of privacy is treated as an asset that can be bought and sold, and individuals may feel it is unfair if the trading of their privacy is profitable to others rather than only to themselves. Perhaps people have a perceived ownership over their data despite prior choices to trade them away, believing they deserve a fair share of the surplus others gain from their data. Or, perhaps, people care about procedural fairness and believe secondary monetization is amoral or untrustworthy (Culnan and Armstrong 1999). Preferences over these primitive ideals can explain users' revealed preferences when faced with salient secondary monetization capabilities of data recipients.

Overall, studies on more expansive privacy models motivate empirical work examining data-sharing patterns related to the presence of secondary markets, where data are shared as intermediate goods rather than experienced as final goods. While my study empirically elicits a revealed behavior of privacy (i.e., willingness to share data), these outcomes result from the unobserved preferences over components of a privacy commodity. This study examines the relationship between data-sharing and secondary data market monetization to support theories of privacy preferences beyond exposure concerns and into activities related to secondary markets, including but not limited to notions of self-determination, fairness ideals, and perceived ownership.

Finally, data markets are an important context for studying more expansive models of privacy and considering preferences that are perhaps not unique to privacy but can certainly influence privacy choice-making. External data-sharing is now a common and profitable digital business strategy among firms. Data-based consumer analytics have evolved privacy issues beyond the nuisance costs of unwanted solicitation theorized in Varian (1996) toward more 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. New entrants into data-based businesses, such as *Clearview AI*, have profited from harvesting user-published digital goods. As consumers become more aware of how the data economy operates, their privacy concerns may evolve beyond the simple aversions to spying and intrusion from an anti-surveillance perspective.

User-generated data that are "appendages" to personal identifiers can be highly valuable in commercial data markets. From the perspective of parties interested in utilizing data internally or externally, the value of consumer data is more than just a function of the ability to identify individuals and is not necessarily correlated with the degree of secrecy or (social) sensitivity of the content.[7] Psychometric data based on the five-factor model, for example, have been shown in psychology to have the ability to understand, predict, and discriminate attitudes and behaviors among individuals (Goldberg 1992, McCrae and John 1992, Junglas et al. 2008, Matz et al. 2017, Li et al. 2019).[8] Miller and Tucker (2018) make a similar characterization about the predictive power of a person's genetic data for 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.

### 2.3. Unstable and Uninformed Revealed Privacy Preferences

Following the "privacy calculus" framework of Laufer and Wolfe (1977), the privacy literature identifies a wide range of concerns that enter into the cost-benefit analysis of each disclosure decision (Culnan and Armstrong 1999, Dinev and Hart 2006). Within the realm of privacy economics, this tradeoff is described as an individual's utility over wealth and privacy: $u(w, p)$ (see Acquisti et al. 2013, Buckman et al. 2019). This model describes how the consumer with $p^+$ amounts of privacy is 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 due to various inattention, environmental, or nonnormative factors. This study, therefore, *never* assumes individuals to have perfectly informed choices. Instead, I assume individuals are motivated to improve their estimate of the benefits they demand, $\hat{r}$, if they become more informed.

[7] 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) finds that personal (or unique) identification is a required feature of the ability of shared user data to invoke privacy concerns, despite the decreasing commercial value of identification.

[8] The public-domain International Personality Item Pool (IPIP) for administering the five-factor personality measurement has been 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).

Users' lack of awareness is a uniquely challenging issue to address. Consequences may result from failed-to-imagine scenarios and incorrect presumptions about data property and exclusionary rights. Moreover, the complexity and novelty of data markets, especially confusion over users' "rights" hidden in terms and conditions, exacerbate the salience challenges to individuals' privacy decisions. Behavioral research has already documented instability in the valuations people place on their privacy (Adjerid et al. 2013, Acquisti et al. 2015). Most of this instability is attributed to a lack of awareness and incomplete information users have about disclosure outcomes (Acquisti and Grossklags 2005a).

Many other behavioral affects and heuristics may come into play regarding privacy choices (Acquisti and Grossklags 2008). For instance, individuals often face tangible, immediate benefits in return for generating data. Behavioral economics research has popularized individuals' tendencies to over-weigh payoffs closer to the present time (O'Donoghue and Rabin 1999). Suppose individuals myopically focus on the immediate benefits of information-sharing while ignoring the less vivid downstream outcomes of data trading. In that case, they do not accurately reveal their willingness to accept for all—and especially more opaque—privacy consequences.

Users' uninformed privacy choices are consequential for consumer welfare, motivating research on information interventions that can help users make better (i.e., utility-enhancing) privacy choices. In most data markets, individuals cannot reappropriate the information they have shared.[9] Therefore, their most significant privacy choice is often an initial decision on whether to disclose non-digital information that generates data. To correct for individuals' lack of awareness about data-sharing consequences, policy-oriented research study the effects of notice-and-consent policies (Athey et al. 2017, Tsai et al. 2011). However, the challenge here is identifying which opaque features of data markets can motivate users' privacy behavior. Therefore, my contribution to notice-and-consent policies is the evidence of a feature users value and are inattentive towards: the secondary monetization capabilities of their recipient.

### 2.4. Interpreting the Monetary Values of Personal Data

As used throughout economic experiments, monetary prices are an appropriate and compatible medium to value commodities, even when real data markets resemble a bartering economy (i.e., data in return for goods and services). Empirical privacy work has found monetary rewards to be effective in obtaining private information (Hui et al. 2007, Xu et al. 2010). In contrast, real commodities used in exchange for data have severe interpretation limits and biased estimates for the value of privacy (Tomaino et al. 2021).

[9] Even under the European GDPR provisions on individual control rights, the non-excludable nature of data makes the enforcement of erasure rights to personal data difficult and costly for both individuals and firms.

Prices can be modeled into all forms of data-sharing as an economic tradeoff, where—even when there are no explicit monetary amounts—there is always an underlying, implicit price representing the loss of the user's consumption value of privacy. Money is comprehensively used as the "numeraire" in a vast majority of real-world transactions, and therefore, easy for an individual to estimate their preferences for 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 data valuations across a population of individuals can be challenging. Preferences for sharing personal data are highly idiosyncratic across individuals, which can lead to uninformative average valuations. Idiosyncrasy is not an issue when there are only differences in preferences over some commodity; however, it becomes challenging when the commodities themselves differ from one person to the next: individual data are, after all, personalized. Some individuals, for instance, may be less averse than others toward sharing their psychometric data depending on whether there are better or worse forms of psychometric data.

Another complication to data-sharing valuations is that individuals might have relative rather than absolute valuations for the privacy they retain by not participating in the data market. For instance, they may know that they prefer more privacy under some condition versus another condition; however, they may have very imprecise estimates of the benefits they are willing to accept under either condition in isolation. Experimental evidence from Adjerid et al. (2018) has already shown that reference dependence is important and present in privacy decision-making, particularly in actual choice contexts.

To circumvent these challenges to empirical privacy research, it is helpful to allow individuals to make *relative* privacy choices—in essence, observing individuals' *changes* to their willingness to share data in response to a common environmental factor. To understand how some external, common factor influences personal data-sharing, observing repeated measures of individuals' privacy behavior can capture more informative privacy responses, which condition for idiosyncratic data and support reference-dependent decision-making. A discussion of the methods for implementing a within-subject design for 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 his or her consumption value of not sharing data, his or her $\hat{r}$—the reservation price demanded to compensate for privacy loss from data-sharing—can change when there is an improvement in his or her awareness of the relevant consequences (perhaps due to more salient environmental signals). Holding fixed the available market prices for data, I can capture

relative preferences using the differential inattention users have towards certain aspects of data markets (see more in Section 3.1).

## 3. Framework

This study focuses on a specific aspect of privacy decisions: whether a user's willingness to share her data—and, equivalently, her valuation of the state of privacy afforded by *not* sharing data—is concerned with the secondary monetization ability of the data recipient. Like other privacy economics experiments, this study starts with a basic decision related to the economic transaction of personal data. The consumption value of privacy (revealed through a person's willingness to share data) is the utility gained or lost by a user when she chooses to share data with a set of recipients.

Many different preference components related to the consequences of data-sharing can determine how much the individual gains or loses in utility. I organize two possible preference components. Suppose that there is some possible amount of disutility that the individual suffers due to data-sharing; that is, $V = v(e) + o(e)$, where $e \geq 0$ is the exposure or number of data recipients.[10] The first component $v$ is the utility loss from sharing data to be experienced as a *final good* by data recipients. The second component $o$ is the individual's disutility from sharing her data to be utilized as an *intermediate good* by recipients. Now, consider two forms of information-sharing—primary and secondary monetization—that can shape a user's privacy losses across these two preference components.

Primary monetization likely triggers an individual to lose utility related to a recipient accessing and experiencing her information as a final good. How much disutility she suffers—or even, does not suffer—depends on how much this aspect influences her value of privacy. This situation, reasonably, assumes that individuals are informed that they are sharing data with some set of recipients, $e$, when they actively choose to trade away data under primary monetization.[11] On the other hand, primary monetization can also trigger privacy concerns related to data as intermediate goods, especially for a user that is more conscious and aware of how her data can be utilized in secondary markets.

Secondary monetization likely triggers utility loss due to the recipient's ability to utilize data as an intermediate good. Recall that the non-rival nature of data enables and exacerbates the ability

[10] As described by Brandimarte et al. (2012), the action to share data with some set of recipients is a necessary precondition for the "access, use, and potential misuse" of data (pp. 341). Note that this simple framework does not restrict the user from gaining positive utility from sharing her user-generated data.

[11] There are, of course, cases in the real world where users are imperfectly informed about whether they are consenting to release data in a primary monetization choice. In this study, the tradeoff is perfectly salient to subjects: they must release their data to a recipient in return for monetary compensation.

of others to use personal data as intermediate goods. The potential for users to suffer privacy costs due to this factor is propelled by the inalienability feature of personal data: even after the user trades data away, it is still "personal" to its originating individual. On the other hand, a user more concerned about the secrecy of their data might be concerned about the exposure consequences from secondary monetization, triggering disutility due to data being experienced as a final good.

In addition, disutility due to secondary monetization depends on how salient this attribute is to the individual. Suppose the user is aware of and suffers disutility from the secondary monetization of her data. In that case, she will reveal this utility loss by increasing the monetary returns demanded or refusing to participate in the primary data market. On the other hand, individuals may only be aware of the secondary monetization potential of their recipient if this information is explicit. If individuals are inattentive and lack such information, then they will not reveal a disutility due to secondary monetization in their privacy choices.

### 3.1. Predictions

It is challenging to observationally disentangle preferences over sharing data as a final or intermediate good given the nonexcludable nature of data markets. A decision to release data includes both preferences regarding whether to tolerate data being made available to be experienced *and* utilized by others. The experimental solution proposed by this study is to exploit the inattention users might have towards the intermediate goods nature of data—by manipulating its salience and, thus, influencing their immediate attention toward the consequences of that component in their decision.

#### 3.1.1. Manipulating Salience to Identify Disutility from Secondary Monetization

To organize and demonstrate how privacy behavior can change in response to information signals about secondary monetization specifically, I adopt the framework for inattention from DellaVigna (2009, p. 349). Let us suppose that the individual is inattentive over features related to data as intermediate goods, the $o$ component (e.g., how her recipients can monetize her data in a secondary market). Then, she perceives that

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

where $\theta$ is the inattention parameter as a function of salience $s \in [0, 1]$ of $o$. Assuming $\theta'(s) < 0$, $\theta(1) = 0$ is full awareness with a fully salient signal, and $\theta(0) = 1$ is complete blindness with no salient signal (which follows psychological theory that information attention is nondecreasing in salient signals). Therefore, varying the salience of information about the recipient's secondary monetization abilities should influence the preference component related to others utilizing her

data as an intermediate good and *not* her taste for the recipient experiencing her data as a final good.

This study intends to capture a revealed disutility to share data with a recipient who can profit in a secondary data market by utilizing data as an intermediate good. By using salient signals to manipulate users' awareness of recipients' ability to benefit through trade with third parties, an experiment can test whether individuals reduce their likelihood to participate and increase the prices they demand for their user-generated data in the data market.

The main prediction of this experiment is that users decrease their willingness to share data when they become more aware that their recipient can monetize data in a secondary market. Conversely, there are consequences and attributes entangled with a recipient's secondary monetization ability, which can potentially increase other privacy concerns related to the $v$ component, and thus decrease the subject's willingness to share data. These factors include increased exposure and differential exposure risks. Additional design considerations and secondary tests can be conducted to examine the importance of these attributes in users' privacy choices.

**3.1.2. 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 a dislike by the individual for having additional parties experiencing the user data as a final good—rather than any disutility specific to data being utilized as an intermediate good. If anti-surveillance desires are the main privacy concern of individuals, then this can dominate any dislike for the secondary monetization ability of recipients.

To control for this factor, the main intervention can be explicit in having a singular third party with which the recipient can transact, limiting the exposure level to only two parties (the recipient and third party). Then, a comparison condition can be designed to test for how users respond to much higher exposure levels in the absence of secondary monetization signals, e.g., many data recipients. If users are more likely to share data at high exposure levels (with no explicit secondary transaction ability) than at low exposure levels (with an explicit potential transaction between the recipient and a third party), then concerns about data exposure are unlikely to explain the decreased willingness to share under a secondary monetization treatment.

On the other hand, users may not strictly perceive that their recipient's transaction with a singular third party to be equivalent to one additional recipient. For example, there may be preconceptions that "a third party" means "any third party" and thus potentially many other individuals receiving one's data. Additional conditions can further disentangle these expectations about data exposure.

First, a partial information condition with explicit signals about the monetization abilities of the recipient, but *no* explicit signals about the secondary market can be explored. If users decrease

their willingness to share data under a recipient's monetization potential alone, compared to one recipient or even thirty recipients without monetization abilities, then increased exposure is unlikely to be the driving force behind the increase in the revealed value of privacy under secondary market monetization.

Second, an additional condition can test for high exposure to many data recipients, in which *each* of these data recipients can engage in secondary monetization. Then, users' willingness to share data can be compared with the main treatment, in which a singular data recipient has the same secondary monetization ability. If exposure to additional parties is the primary driver of lower willingness to share due under secondary monetization conditions, then we would observe a significant decrease in willingness to share with secondary monetization by many recipients. If such a situation does not occur, then it is unlikely that exposure concerns explain the revealed disutility to a recipient's secondary monetization ability.

**3.1.3. Differential Risk between Primary and Secondary Market Recipients** Secondary market activities can be perceived as riskier depending on the type of recipient to which the data are exposed. A differential risk from allowing data access by third parties versus data recipients can explain users' decrease in willingness to share with a recipient who can trade in secondary markets. To explore this possibility, different conditions can be compared.

First, the aforementioned partial information condition is informative. Here, signals are delivered to a subject with no information about secondary markets or third parties, but the condition tests for monetization signals alone to reveal users' response to the recipient privately gaining from the utilization of their data. This condition keeps signals about third-party access opaque, while maintaining information about the intermediate goods nature of data.

Second, conditions can be designed to more specifically rule out any differences between markets. The "third party" can be, explicitly, another data recipient from the primary data market. This design tests whether individuals are less willing to share their data when secondary transactions can occur between two indistinguishable persons, essentially creating a secondary market from the same pool of primary-market recipients. If users still decrease their willingness to share data when the profile of recipients in the secondary market is controlled to be identical to those in the primary market, then this suggests that exposure to differential risks in the secondary market is not driving the revealed disutility from secondary monetization.

Finally, the previous condition is also an especially useful comparison for data-sharing with many recipients *without* signals about their secondary monetization abilities versus two recipients *with* signals about their secondary market monetization abilities, as all recipients are drawn from the same anonymous and indistinguishable pool of potential recipients. This approach effectively

controls for both exposure level and type. If users reveal a higher valuation for not sharing data when there are two recipients who can benefit through trade with each other—versus sharing data with thirty recipients and no signals about secondary transactions—then concerns related to differential exposure risks are unlikely to be the driving force behind any revealed aversion to secondary monetization.

## 4. Experimental Design

The online experiment was designed to simulate a personal data market where users generated their psychometric data and faced real decisions to share those data with others in return for benefits. The study design can be broadly categorized into three stages: data generation, data sharing, and data earnings. In the data-generation stage, individuals generate psychometric data that are personally identifiable. In the data-sharing stage, users consider a primary monetization choice to participate in a data market by sharing their data with data recipients. In the final stage, real outcomes from data-sharing are realized by both the users and their data recipients, where the users receive monetary earnings from any data-sharing and the data recipients earn profits from user data through secondary monetization.

### 4.1. Subjects

The study was conducted in a U.S. university's business school lab (hereafter the "Lab"). The Lab maintains an Institutional Review Board (IRB)-approved subject pool for online studies. The Lab regularly hosts studies on economic games, where subjects partake in real trades with one another. A benefit of running a study in this form of research lab is the ability to tightly control the setting and implement real trades between subjects. There are credible consequences in this lab environment, as this experiment is similar in fashion to those of other studies hosted at the Lab with real buyers and sellers, randomly and anonymously paired to conduct real transactions.

Another reason that the Lab was well positioned for this study its high-quality maintenance of a subject pool. Importantly, the personal identifiers of subjects (e.g., names and emails) were available to the researcher prior to the launch of the survey and then attached to survey takers' psychometric data during the experiment. More generally, the Lab monitors and removes users who register under aliases, click through a survey without engaging with the content, or have 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 participants with the idea that the study was meant to examine privacy preferences. In fact, the words "privacy" and "security" were not used for the entirety of the study until the exit survey questions related to privacy attitudes were asked.[12] Additionally, a

[12] Adjerid et al. (2019) showed that individuals had a higher propensity toward privacy outcomes when prompted to make decisions about their "privacy" settings versus their "survey" or "app" settings.

minimum subject fee of $2 in Amazon gift cards for a fifteen-minute study, with a possibility to earn more based on the survey taker's decisions within the study, was offered.

### 4.2. Data-Generation Task

In the first stage of the survey, responses to a 50-item five-factor questionnaire about the respondents' attitudes, personality, and habits were collected. This survey was taken from the standard sample of Likert-type assessment statements from IPIP, which is widely used in psychology research. A sample of items are shown in Appendix Figure 13. Responses to these self-assessments generated each person's five-factor personality scores across the following traits: extraversion, agreeableness, conscientiousness, emotional stability, and intellect.[13]

After completing all 50 items in the self-assessment, respondents were presented with their scores and personal identifiers (see the example in Figure 1) as well as information about how to interpret their scores (see Appendix Figure 11). Each score is an integer in the range of 10 to 50. A high score in extraversion, for example, indicates high extraversion and low introversion.

**Figure 1 Example User-Generated Psychometric Data.**

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

Similar to previous work measuring privacy valuations over personal data gathered within the study, untruthful responses could have occurred in the information people disclosed. 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, it is important to note that 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 in eliciting and discriminating personality measures that can predict the real behaviors and traits of individuals (Matz et al. 2017), particularly in Western and educated populations (Laajaj et al. 2019).

Importantly, considerations were made in the experiment to remove contamination risk with data-sharing choices in the second phase of the experiment. The self-assessments were elicited at the start of the study, without revealing to participants that the study intended to elicit their willingness to share their data with other persons in the study. Certain scores were not more economically valuable than other scores, and this indifference was made explicit to subjects prior to both self-assessment and data-sharing decisions.[14]

[13] All self-assessment statements required a response from the subject. Skipped questions prevented the survey from continuing onto the next page and the later stages of the survey. Only 11 registered participants who started the survey did not continue through to completion, and of those, 5 did not complete the self-assessment stage.

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

### 4.3. Eliciting Data-Sharing Choices

To elicit honest valuations about the state of privacy afforded by *not* sharing data, the BDM (Becker et al. 1964) incentive-compatible method for eliciting willingness to accept was used. The BDM mechanism is commonly used in behavioral and experimental economics to reveal an individual's reservation price of a good, even for commodities without established market prices (List and Shogren 2002).

An important characteristic of the BDM method is its ability to reveal valuations that are not strategic prices. This advantage is critical in this study, where the primary information treatment includes signals that can update user beliefs about the value of data to their potential recipient. The dominant strategy in this auction is to reveal the minimum acceptable price (reservation price) that can correctly reflect the "consumption value" of privacy that the individual has for keeping data from being shared with a recipient—rather than a rational response to new information about how valuable her data may be to a potential buyer. Should the consumption value go up or down due to the ability of a potential buyer to gain more in secondary monetization, then this is directly related to the utility gained or lost by the user from aspects concerning privacy to which they were previously inattentive.

The implementation of a BDM price elicitation is not without its challenges. The instructions for how to choose a minimum price given an unknown final price drawn from a random distribution can be unusual and confusing. It is easy to conflate choosing the *minimum* acceptable price with *any* acceptable price. Moreover, it is important to recognize that data are not sold for explicit prices in the real world but, rather, are implicitly priced with bartered digital services. Thus, it is reasonably challenging for the average decision-maker to deduce her reservation price without prior experience with market prices for data.

To remedy this challenge, the survey used a short and coarse list of only five prices (i.e., the survey taker did not need to consider a continuum of prices), which was neither long nor granular. Decision-makers in the experiment considered five potential prices—\$0.01, \$0.49, \$0.99, \$1.99, and \$2.99—to either accept or reject.[15] The survey taker could easily consider each price in sequence and imagine whether she would "take it or leave it" if that price were the final price. It is much easier to imagine if one can accept \$2.99 in exchange for data-sharing, rather than form a point estimate of one's minimum price.[16] See Appendix Figure 14 for the price elicitation.

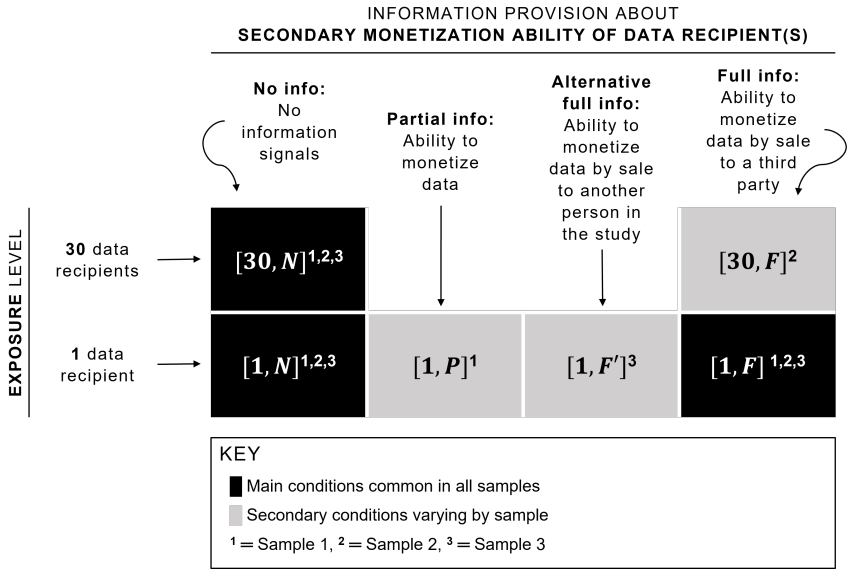
[15] The price range was chosen based on pilot surveys of this study, where participants were asked to share their data under a one-price choice. The \$0.49 and \$0.99 ending prices are commonly advertised format for prices, designed to evoke familiarity in subjects' minds with other digital goods and services. The \$0.01 was used in replacement of a \$0.00 choice to prevent any nonnormative factors from influencing this decision (i.e., there may have been a special reluctance to choose to give away something for free, even if a subject valued their privacy at a non-positive price).

[16] Price-reversing behavior in a multiple-price list can happen (e.g., accepting \$1.99 while rejecting \$2.99) either by mistake or due to uncertainty. However, survey mechanisms were not designed to prevent this to observe whether

### 4.4. Data-Sharing Conditions

Conditions were created to capture individuals' willingness to accept the primary monetization of their data with a recipient, with more or less salient signals about their recipient's secondary monetization ability. The baseline condition did not include information signals ($[1, N]$), and the main treatment condition provided full information to increase the salience of the recipient's ability to benefit through trade with a third party ($[1, F]$). The condition to elicit willingness to share under high exposure was operationalized as a choice to release data to thirty recipients ($[30, N]$), which helped test whether exposure concerns explained the effects found under $[1, F]$. All subjects made decisions under three conditions, $[1, N]$, $[30, N]$, and $[1, F]$; i.e., these conditions were replicated across all three samples.

**Figure 2 Six Experimental Conditions.**




Each subject was limited to four data-sharing decisions, one of which was to be randomly selected and implemented for the real data transaction. Decision rounds were limited to four in order to reduce inattention and survey fatigue. In replicating the experiment across three samples, one of the four conditions was alternated out for another condition. Specifically, each subject also had a fourth data-sharing decision—not necessarily the fourth-period decision—in which one of three secondary conditions were implemented. Figure 2 summarizes the six data-sharing conditions, and the exact survey text is provided in Appendix Table 6.

survey takers were inattentive to or indecisive in the price elicitation in post-experimental analyses. Approximately 6% of subjects exhibited some form of price-switching behavior, and they are included the analysis. The preregistration also did not specify the exclusion of these subjects.

The first sample included a partial information condition, $[1, P]$, that removed explicit signals about a recipient's ability to externally transact data with a third party, keeping only information about the monetization benefits to the recipient. This condition explored whether the monetization abilities of the recipient alone could decrease individuals' willingness to share data while serving to rule out exposure concerns as a potential underlying explanation of the $[1, F]$ effects.

The second sample included a high exposure, full information condition $[30, F]$, in which thirty data recipients could *each* monetize data through trade with a third party. This condition provided an interaction between high exposure and secondary monetization, assessing whether, if exposure concerns are primary to subjects, sharing with thirty data recipients and their third parties significantly reduced subjects' willingness to share compared to the case with one recipient and a third party. Furthermore, this condition provides an alternative context in which to replicate the test of whether salient secondary monetization decreases willingness to share data when comparing $[30, N]$ and $[30, F]$.

The third sample included an alternative description of the "third party" by explicitly defining it as another person in the study ($[1, F']$). This condition served to control for differences between recipients in the primary data market (i.e., another person in the study) versus "third parties" in the secondary market, rendering recipients and third parties indistinguishable by randomly selecting third parties from the same pool of potential data recipients. Through this condition, the main effects under $[1, F]$ can be disentangled from the explanation that by subjects perceived "third parties" as riskier forms of exposure than the exposure from releasing their data to recipients in the primary market.

### 4.5. Randomization Groups

The primary advantage of a within-subject approach, versus randomizing treatments across individual clusters (i.e., a between-subjects design) is that it removes the effect of extraneous subject-level characteristics on the outcome variable. This design is particularly attractive in a study eliciting people's personal data valuations. As described in Section 2.4, high or low willingness to share data can be highly dependent on the subject's personalized psychometric data. As an example, this is reflected in a recent privacy experiment by Collis et al. (2021), where the distribution of prices to share personal data is multimodal.

The second advantage of this 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.[17] A within-subject design is advantageous in terms of its ability to capture a 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 spillover effects from previous decisions. To resolve this in the experiment, block randomization was utilized with a Latin square design to ensure each condition appeared at each ordinal position of decisions in a balanced fashion (see Appendix Table 7). This randomization is a standard technique for counterbalancing spillover effects in a within-subject design, rendering the outcomes from one condition independent from the order in which it was presented to the respondent. When analyzing individual-level changes in data-sharing behavior, the experiment randomized whether subjects experienced, for example, $[1, F]$ 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 from the experimental treatments.

Prior to any data-sharing decisions, the survey explained to the subjects that they would be considering "several" scenarios for sharing their data with others in the study and that a randomly selected choice would ultimately be made real based on a randomly selected price. Each subject's outcome was revealed after all data-sharing choices were made. A summary of the survey chronology from the subjects' perspective is shown in Figure 3. One week following survey submission, payments and data were delivered via email to subjects.[18]

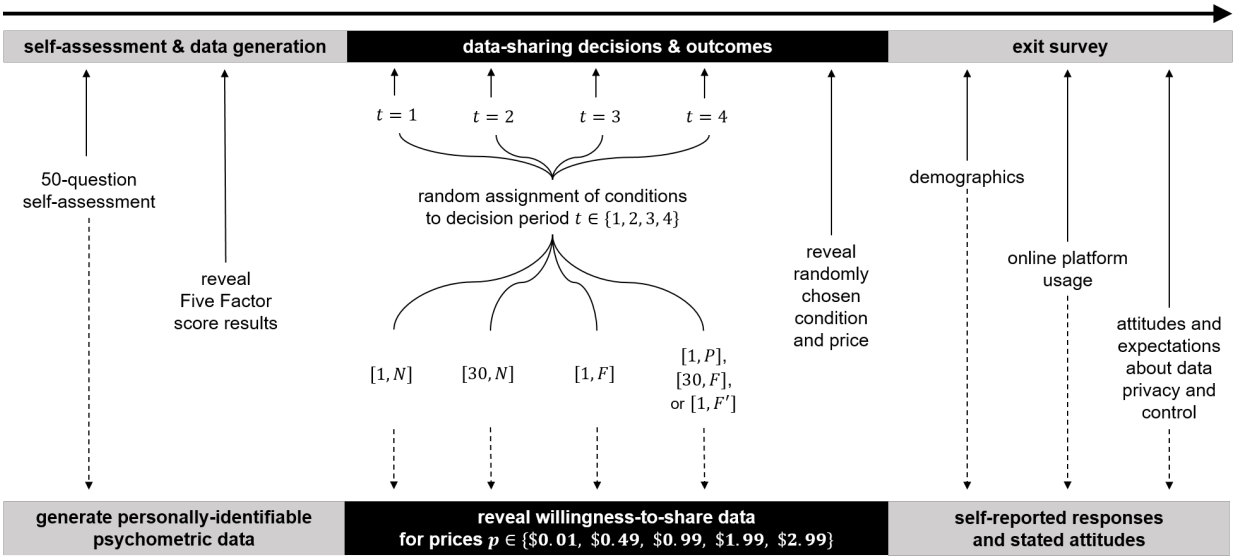
### 4.6. Data Collection

The data were collected across three sub-samples, with a combined total of 1,188 participants in Spring 2019, Fall 2019, and Summer 2020. The survey also included voluntary questions regarding the individuals' demographics and social media usage. Approximately 79% of individuals self-reported using Facebook with a non-anonymous account and accessing the platform at least once

[17] 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 goes away in a within-subject design, where 9 is rated as relatively smaller than 221.

[18] Less than one percent of subjects in the study were data recipients who could earn money from others' personal data. Subjects who were data recipients were not given any option to self-select and any choices other than to receive data and payments. The role of the data recipient was not advertised in the study and was imposed only on a randomly selected set of people who registered for the study. Profits were determined by a rate of return on the number of data recipients in their survey wave that ultimately shared their data with the recipient. See an example of this data delivery process in Appendix Figure EC.6.

**Figure 3 Survey Chronology from Subjects' Perspective.**




per week. Following this same definition of platform usage, 50.8% and 24.3% of participants self-reported as users of LinkedIn and Twitter, respectively.[19] Finally, a series of privacy and data ownership attitudes were elicited using a Likert-type assessment of statements related to privacy concerns and data usage by firms (see Appendix Table 4).

# 5. Results

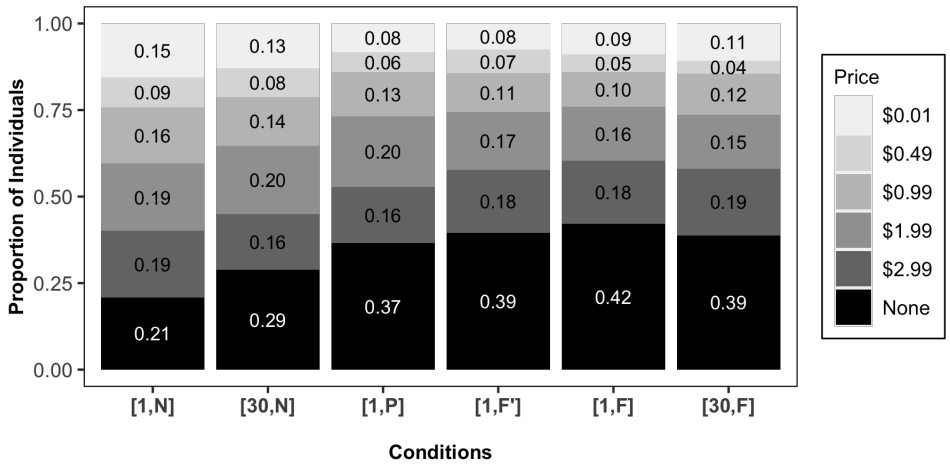
## 5.1. Descriptive Summary

Figure 4 summarizes the reservation prices for subjects by condition. Conditions where subjects encountered salient secondary monetization abilities of data recipients show the highest percentage of pricing-out behavior ($[1, F']$, $[1, F]$, and $[30, F]$), at around 39 to 42 percent. Pricing-out behavior is also nearly as prevalent (at 37%) under partial information ($[1, P]$), where only the monetization potential of the recipient is explicit. Finally, descriptive evidence points to exposure concerns (manipulated by $[30, N]$) as being less important for users' willingness to share data and an unlikely mechanism behind the lower instances of data-sharing due to secondary monetization.

Leveraging a within-subject design, changes in individuals' participation behavior can be captured between conditions—i.e., whether there is pricing-out behavior in one condition and pricing-in behavior in another condition at the individual level. The largest share of data market participation change occurs when salient signals are provided about the recipient's secondary monetization

[19] A minority of those surveyed reported high frequency engagement with these platforms. For example, 19% were Facebook users who posted or shared content on the platform at least once a week; 39.1% were LinkedIn users who responded to requests for connections within a week; and 12.3% were Twitter users who posted or shared content at least weekly. Similar to other studies, approximately 72% of participants self-identified as female. Their mean reported age was 23.8 years ($SD = 6.95$). Over 70% reported being students as opposed to being employed (full or part-time). This study's demographic make-up is not unusual compared to those of other studies conducted which this university's subject pools.

**Figure 4 Reservation Prices for Data-Sharing by Condition.**



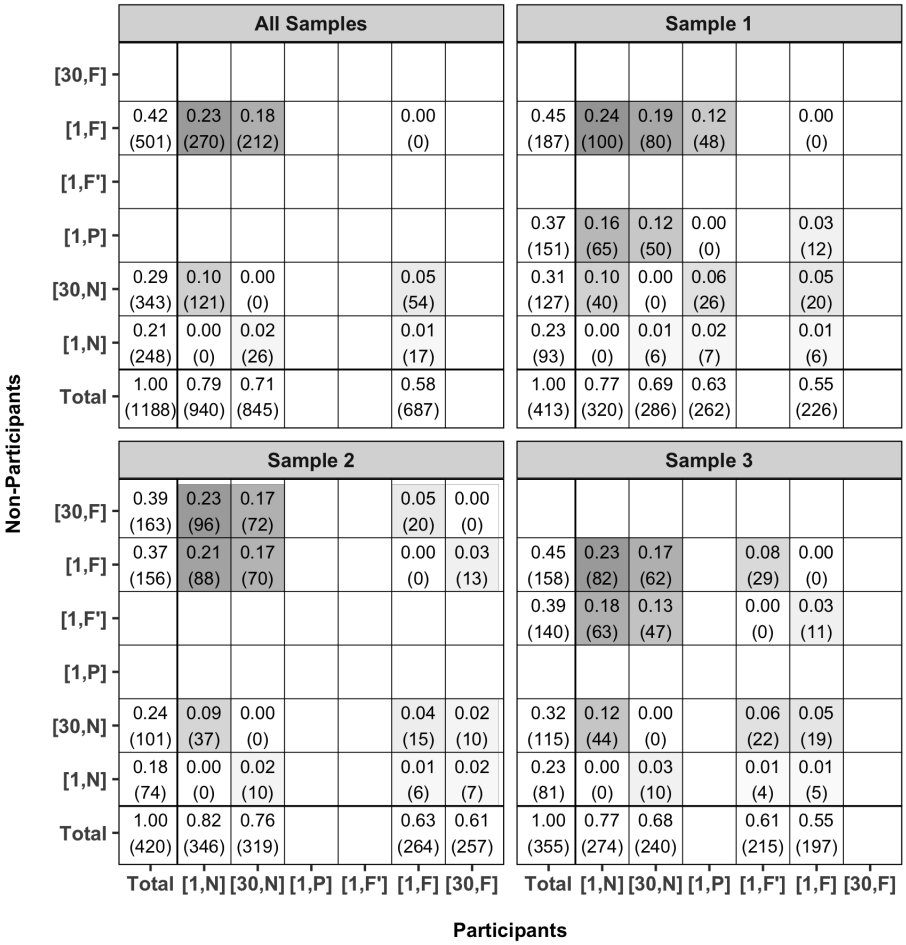

ability. A high share of exit behavior is also revealed when a recipient's monetization ability alone is salient (without explicit signals of third parties). To a lesser extent, there is also exit behavior when exposure increases from one to thirty recipients. Finally, even fewer changes are observed when comparing between secondary monetization conditions with differential risk.

As shown in Figure 5, 23% of subjects were willing to share personal data with *one* recipient ($[1, N]$) but unwilling to share data when treated with information about their recipient's secondary monetization abilities ($[1, F]$). The response to $[1, F]$ is unlikely explained by increased exposure. Data-sharing with many recipients ($[30, N]$) exhibits fewer changes in individuals' data market participation. A smaller proportion (10%) shared data under $[1, N]$ market but did not share when exposure was high in $[30, N]$. In fact, 18% of subjects were willing to enter the data market under $[30, N]$ but not under $[1, F]$ (in contrast, only 5% of subjects exhibited the reverse behavior).

The secondary conditions in each sub-sample also suggest that the proportion of people who exit the market under $[1, F]$ is unlikely to be explained by exposure concerns. Sixteen percent of subjects exited the data market in response to only a recipient's monetization abilities ($[1, P]$ in Sample 1), which was still greater than the number of those who exited under the high-exposure condition ($[30, N]$ in Sample 1). Second, when thirty participants *each* had the ability to trade with third parties ($[30, F]$ in Sample 2), only 5% of subjects who also participated under the $[1, F]$ condition, despite the increase in exposure, exited the data market. Finally, 18% of subjects exited the data market with two recipients and secondary monetization signals ($[1, F']$ in Sample 3), compared to 12% who exited with 30 recipients without secondary monetization signals ($[30, N]$ in Sample 3).

The descriptive results also suggest that differential exposure risks in the secondary market do not explain the large share of exit behavior under $[1, F]$. When differential risks are controlled for ($[1, F']$ in Sample 3), 18% of subjects still exited the data market in response to their recipient's secondary monetization abilities. Less than half the number of subjects (8%) changed their participation

**Figure 5 Data Market Participation and Non-Participation Across Conditions.**




*Note.* Proportions out of sample and number of individuals are in parentheses.

between $[1, F]$ and $[1, F']$ in response to the risk differential when the third party was not explicitly another person in the study.

## 5.2. Estimated Participation and Prices

Given the nature of the price elicitation, acceptable prices were chosen simultaneously with whether to accept *any* price in the available range. The selection of "I do not accept" for all prices allowed individuals to exit the experiment's data market. Rather than assuming a natural censoring by the price list, the chosen analysis method is to measure two parts: (1) the odds or likelihood of data market participation by individuals and (2) the price demanded among those who participated. Unlike many labor market participation questions in economics that focus on changes to the price variable, the extensive margin is both relevant and without selection bias. Real data markets often operate with "all or nothing" data-sharing decisions and rarely ask consumers "how much" they will accept in exchange for data. While measuring price changes is useful for inferences internal to the experiment, it is difficult to map these magnitudes to external contexts, especially where

benefits from data are not monetary amounts and when the pricing decision is estimated for the subset of subjects who select into the data market.[20]

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

$$Participation_{it} = \begin{cases} 1, & y^*_{it} \leq 2.99 \\ 0, & y^*_{it} > 2.99. \end{cases} \tag{1}$$

where $y^*_{it}$ is the true, unobserved minimum acceptable price. Second, for those who selected into participating, it estimates the average changes to the observed reservation price that they were willing to accept. The observed minimum acceptable price when $Participation_{it} = 1$ is

$$Price_{it} = \begin{cases} 0.01, & y^*_{it} \leq 0.01, \\ 0.49, & 0.01 < y^*_{it} \leq 0.49, \\ 0.99, & 0.49 < y^*_{it} \leq 0.99, \\ 1.99, & 0.99 < y^*_{it} \leq 1.99, \text{and} \\ 2.99, & 1.99 < y^*_{it} \leq 2.99. \end{cases} \tag{2}$$

Since each individual $i$ made 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):

$$[Participation_{it}, Price_{it}] = \alpha + \beta \cdot \mathbf{Condition}_{it} + \delta \cdot \mathbf{T}_t + \gamma \cdot \mathbf{Y}_i + \theta_i + u_{it} \tag{3}$$

where $\mathbf{Condition}_{it}$ is a set of categorical variables indicating the conditions of interest to be compared with a "leave-out" 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; i.e., this controls for a subject's experience and learning effects.

The variable $\mathbf{Y}_i$ is a set of participant-specific characteristics. Importantly, included in the estimation are controls for anchoring effects, such as whether the individual was randomly assigned to a no information ($N$) condition in the first two decision periods and any secondary monetization information ($P$, $F$, or $F'$) condition in the last two decision periods. Additionally, whether the individual experienced 30 recipients before or after one recipient is included.

A set of optional characteristics are included as an opportunity to control for variation in privacy behavior across different types of participants. Included are psychometric scores of individuals to

[20] While I recognize that there exist behavioral mechanisms that influence data market participation separately from those prices demanded (as opposed to natural censoring), identification of these mechanisms is beyond the scope of this study. A Heckman-style selection model is not my estimation method, given the lack of a valid exclusion restriction. Therefore, a two-part estimation—without a correction for selection bias—is the preferred style of inference for this study.

control for whether a low or high score in each trait is correlated with individuals' relative privacy preferences between conditions. Studies find that five-factor scores have predictive power over 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 participant-specific random effect is denoted by $\theta_i$, and $u_{it}$ is the error term. Since all individuals experienced conditions randomly assigned to a decision period, the estimates on the data-sharing conditions are uncorrelated with the observed ($\mathbf{Y}_i$) and unobserved individual differences and error term ($\theta_i$ and $u_{it}$, respectively). Estimates of individuals' participation and conditional prices include three specifications of random effects panel regressions. The results tables present the likelihood of participating and the prices demanded in the data market relative to the condition with one data recipient and no information provisions. There are three specifications for each part. The first specification and baseline model includes variables indicating the condition under which a privacy choice was made, controlling for sample and order effects. The second specification extends the first by including a control for anchoring effects—i.e., if $[1, F]$ or $[30, N]$ comes before or after $[1, N]$. The third specification includes controls for individual-specific characteristics.[21] 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 likelihood of subjects participating in the data market and minimum prices demanded.[22]

Salient secondary monetization decreased subjects' willingness to share data. Relative to condition $[1, N]$, the odds of not entering the data market under condition $[1, F]$ were 2.83 times greater. A similar pattern was reflected in the intensive margin, where prices under $[1, F]$ were approximately \$0.39 higher ($p < 0.0001$) relative to $[1, N]$. Overall, individuals were more likely to *not* participate and increased the prices demanded when information was provided about their recipient's ability to monetize data through sale to a third party.

Notably, the magnitude of subjects' revealed aversion to the secondary monetization treatment, $[1, F]$, is larger than that to the high-exposure manipulation, $[30, N]$. Under $[30, N]$, the odds of subjects not participating in the data market were 1.55 times greater, and the prices demanded were approximately \$0.10 higher ($p < 0.0001$). While exposure manipulations significantly decreased

[21] Demographic controls included female, marital status, Facebook usage, and employment status. Psychometric controls included whether the individual scored above 30 (on a scale of 10 to 50) in each of the five-factor traits: extraversion, agreeableness, conscientiousness, emotional stability, and intellectual.

[22] The change in prices demanded was estimated among the sub-sample of individuals who participated in the data market under $[1, N]$ and the focal condition. If $C$ is the focal condition and $\bar{C}$ is the comparison condition, then the odds ratio of $C$ and $\bar{C}$ is $\exp\left[\hat{\beta}\right] = \exp\left[\log\left(\text{odds}C/\text{odds}\bar{C}\right)\right] = \text{odds}C/\text{odds}\bar{C}$, where $\hat{\beta}$ is the coefficient estimate of the condition of interest.

**Table 1 Individuals' Participation and Prices for Personal Data (All Samples).**

| **Variables** | **Logit** | | | **OLS** | | |
|---|---|---|---|---|---|---|
| Dependent: | *Participation* | | | *Price* ($) \| *Participation*= 1 | | |
| Comparison: $[1, N]$ | (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) |
| Salient secondary monetization $[1, F]$ | -1.018*** | -1.016*** | -1.039*** | 0.390*** | 0.393*** | 0.392*** |
| | (0.064) | (0.064) | (0.065) | (0.033) | (0.033) | (0.033) |
| High exposure $[30, N]$ | -0.433*** | -0.433*** | -0.441*** | 0.102*** | 0.102*** | 0.102*** |
| | (0.054) | (0.054) | (0.055) | (0.024) | (0.024) | (0.024) |
| Sample controls | × | × | × | × | × | × |
| Order controls | × | × | × | × | × | × |
| Anchoring controls | | × | × | | × | × |
| Psychometric controls | | | × | | | × |
| Demographic controls | | | × | | | × |
| Individual clusters | 1188 | 1188 | 1188 | 975 | 975 | 975 |
| Observations | 3564 | 3564 | 3564 | 2472 | 2472 | 2472 |
| **ANOVA: 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 |

*Note:* Clustered robust standard errors are in parentheses.
$^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; and $^{***}p < 0.001$.

subjects' willingness to share data, they decreased it significantly less than they did under salient secondary monetization. The odds of subjects not participating in the data market under $[1, F]$ were 1.82 times greater than $[30, N]$ ($p < 0.0001$), and the prices demanded approximately \$0.29 higher ($p < 0.0001$). Just *one* data recipient's trade with a third party resulted in a lower willingness to share data that compared to sharing with 30 recipients.

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

First, salient signals about a recipient's monetization alone—without explicit information about third-party sales—could motivate lower willingness to share data. As shown in the Sample 1 panel in Table 2, the odds of subjects not sharing data under this partial information condition $[1, P]$ were 2.02 times greater relative to $[1, N]$. The results of data market participants' reservation prices followed the same pattern: $[1, P]$ had approximately \$0.25 higher prices than in $[1, N]$ ($p < 0.0001$). In fact, the decrease in willingness to share data under $[1, P]$ versus the high-exposure manipulation $[30, N]$ was significantly larger: participation was 1.3 times greater relative to $[30, N]$ ($p = 0.0007$), and the prices were \$0.16 higher ($p = 0.001$). Note that there was also a significant increase in the exit behavior and prices under $[1, F]$ relative to $[1, P]$, revealing a disutility related to salient signals about secondary market recipients or the external nature of the transaction.

**Table 2 Participation and Prices for Personal Data (By Sample).**

| **Variables** | **Logit** | | | **OLS** | | |
|---|---|---|---|---|---|---|
| Dependent: | *Participation* | | | *Price* (\$) | *Participation*= 1 | |
| Comparison: $[1, N]$ | (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) |
| $[30, N]$ | -0.425*** | -0.428*** | -0.441*** | 0.092* | 0.091* | 0.090* |
| | (0.083) | (0.084) | (0.086) | (0.039) | (0.039) | (0.039) |
| $[1, P]$ | -0.682*** | -0.681*** | -0.701*** | 0.252*** | 0.252*** | 0.252*** |
| | (0.097) | (0.097) | (0.100) | (0.050) | (0.050) | (0.050) |
| $[1, F]$ | -1.046*** | -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) |
| $[30, N]$ | -0.395*** | -0.407*** | -0.426*** | 0.069† | 0.069† | 0.070† |
| | (0.098) | (0.102) | (0.106) | (0.036) | (0.036) | (0.036) |
| $[1, F]$ | -1.007*** | -1.021*** | -1.081*** | 0.430*** | 0.432*** | 0.432*** |
| | (0.115) | (0.115) | (0.120) | (0.052) | (0.052) | (0.052) |
| $[30, F]$ | -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) |
| $[30, N]$ | -0.484*** | -0.483*** | -0.492*** | 0.168*** | 0.168*** | 0.168*** |
| | (0.102) | (0.102) | (0.103) | (0.048) | (0.048) | (0.048) |
| $[1, F']$ | -0.789*** | -0.785*** | -0.800*** | 0.390*** | 0.392*** | 0.391*** |
| | (0.103) | (0.102) | (0.104) | (0.057) | (0.057) | (0.057) |
| $[1, F]$ | -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 | | | × | | | × |

*Note:* Clustered robust standard errors are in parentheses.
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$.

Second, privacy responses to data recipients' secondary monetization abilities also replicated under a thirty-recipient, high-exposure scenario. In Sample 2, the effect of secondary monetization signals persisted when individuals considered thirty data recipients—the odds of them not participating were 2.09 times greater ($p < 0.001$), and the prices were \$0.33 higher ($p < 0.001$) under $[30, F]$ than under $[30, N]$. Moreover, individuals exhibited little change in privacy behavior under $[1, F]$ versus $[30, F]$, which dramatically increased their exposure to more data recipients who *each* had the ability to make third-party sales. This situation provided additional evidence that exposure concerns did not explain the striking decrease in willingness to share data when secondary monetization is salient.

Finally, in Sample 3, information was provisioned to explicitly describe the "third party" as another person in the study ($[1, F']$). Despite controlling for the perceived risk differential between parties in the secondary versus primary market, there remained a decrease in willingness to share data under salient secondary monetization. Using this condition, the results were also robust to finding that exposure concerns did not explain the revealed disutility to secondary monetization. When faced with a decision to share data with two persons in the study (with secondary monetization) versus thirty people in the study (without secondary monetization), subjects had significantly lower willingness to share data in the former. The odds of exiting the data market were 1.36 times greater ($p = 0.002$), and the prices \$0.22 higher ($p = 0.0003$) compared to $[30, N]$. Notably, there was also an increase in pricing-out behavior under $[1, F]$ relative to $[1, F']$ ($p = 0.004$), suggesting the existence of a privacy concern related to risk manipulation, but the price changes were statistically inconclusive.

## 5.3. Robustness Checks

**5.3.1. Between-Subjects Comparison of First-Period Choices** First-period decisions have no context or experience, and subjects could not reveal their relative preferences (as previously discussed in Sections 2.4 and 4.5), which should challenge statistical precision and the meaning of the estimates. Despite the expected challenges to first-period choices, there was a detectable reluctance to select into data markets when secondary monetization was salient.

**Table 3 Between-Subjects Comparison of First-Period Participation and Prices.**

| **Variables** | **Logit** | | | **OLS** | | |
|---|---|---|---|---|---|---|
| Dependent: | *Participation* | | | *Price* (\$) \| *Participation*$=1$ | | |
| Comparison: $[1, N]$ | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) |
| (Intercept) | 1.121*** | 1.108*** | 1.567** | 1.540 | 1.562 | 1.489 |
| | (0.135) | (0.168) | (0.493) | (0.074) | (0.094) | (0.276) |
| High exposure $[30, N]$ | -0.331† | -0.331† | -0.343† | 0.021 | 0.020 | 0.044 |
| | (0.184) | (0.184) | (0.188) | (0.103) | (0.103) | (0.104) |
| Salient secondary monetization $[1, F]$ | -0.458* | -0.458* | -0.451* | 0.050 | 0.050 | 0.055 |
| | (0.182) | (0.182) | (0.185) | (0.106) | (0.107) | (0.107) |
| Sample controls | | × | × | | × | × |
| Psychometric controls | | | × | | | × |
| Demographic controls | | | × | | | × |
| Observations | 892 | 892 | 892 | 625 | 625 | 625 |
| Adj. pseudo-$R^2$ | 0.001 | -0.002 | 0.001 | -0.003 | -0.005 | -0.009 |

*Note:* Standard errors are in parentheses.
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$.

Based on the design of the experiment's randomization groups, subjects were randomly and evenly assigned across the three main conditions in the fashion of a between-subjects experiment: $[1, N]$, $[1, F]$, and $[30, N]$. Table 3 shows that by limiting the analysis to only decisions in the first

period, the between-subjects comparison finds that individuals faced with salient secondary monetization were significantly less likely to participate in the market than were those not faced with such signals ($p < 0.05$). However, differences between secondary monetization treatment subjects and high-exposure condition subjects are not detectable during the first period. Similarly, price differences are not detectable in the first period.

Revealed preferences among these three main conditions became clearer—both in magnitude and in precision—as subjects in the study became more experienced (see Appendix Table 5). The gap between $[30, N]$ and $[1, F]$ is significant by the second period ($p < 0.10$). In the fourth decision period, the between-subjects comparison reveals the largest decrease in subjects' likelihood to participate ($p < 0.001$) under secondary monetization provisions, including when the condition is compared to high-exposure-condition individuals ($p < 0.001$). The intensive margin choices (the prices demanded among those subjects who participated) also becomes precise and significantly different. These results support the notion that individuals make *relative* choices in their willingness to share personal data. Experience interacts with their data-sharing choices, and when faced with repeated choices, subjects' revealed preference for each condition becomes more precisely estimated.

**5.3.2. Quality of Psychometric Data** IPIP scores based on the five-factor model are designed to be self-reported. However, the quality of these user-generated data are undoubtedly a level-below that of a third-party expert evaluator. For example, a psychological analysis of each subject by an expert, perhaps in a field setting and over a long period of time, would curate a superior quality dataset of individuals' psychometric information. Therefore, it is plausible that individuals behave differently based on their perceived quality of their data to recipients and secondary markets. While this does not confound the experimental treatments, it does challenge whether the findings of this study can be generalized to data captured about a user from an outside evaluator (i.e., not user-generated data).

While the data in this study are user-generated, five-factor item responses can be assessed for consistency in responses. Appendix Figures 6, 7, 8, 9, and 10 provide correlation plots of individuals responses, categorized by items that measure the same personality trait. If individuals are answering the survey inattentively or randomly, then the results would show low correlation between subjects' responses for the same trait. The general pattern of survey responses shows consistency (i.e., positive correlation), which rules out the prevalence of low-quality, randomly generated psychometric scores. However, understanding the degree to which data quality suffers from inattentive respondents does not identify whether there is a subset of individuals who are intentionally curating a false persona, which cannot be discerned from the data of this study.

Despite this inability to identify how closely the data represent *true* personalities, it is unclear how information accuracy or quality interacts with data-sharing behavior in this context. Inaccurate psychometric data can incur both costs and benefits for the user who shares data. First, such inaccuracy can lead to disutility for the user due to being falsely represented to a data recipient. On the other hand, inaccurate information can obfuscate a user's true personality, which she may value being kept secret. An interesting direction for future research would be to disentangle these two mechanisms through exogenously assigning objectively true or false psychometrics, through privacy-protecting tools like garbling. Furthermore, a more challenging design would be to disentangle endogenous motivations for individuals who intentionally falsify their data, which is less easily controlled in an experimental setting.

### 5.4. Limitations

There are several limitations to the interpretation of my results. First, there is a natural difference between for-profit and non-profit studies (i.e., research studies). The expectations of many subjects coming into a research study are disassociated from for-profit data markets. This study cannot conclude whether individuals are aware of data monetization in these secondary markets in real-world data exchanges. However, replicating this study in the field would help increase the understanding of how these results can be generalized outside of a controlled environment.

Second, there are strong anchoring effects of price elicitation survey questions. While the relative valuations in this study provide meaningful inferences, 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 attempts to minimize this issue, the point 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 regarding the challenge of measuring privacy preferences, point estimates of reservation prices are unnatural, unstable, and difficult to compute for a decision-maker with no prior experience with explicit prices for personal data.

Third, I cannot correct for selection bias on the relative 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 versus those who would exist on an unobserved distribution of prices. For example, it could be the case that \$6.99 is enough to capture all or none of (or somewhere in between) those who reject \$2.99. Wider ranges of price lists cannot easily correct for this issue due to the aforementioned anchoring effects. Furthermore, extremely high upper-bound prices can be less believable to decision-makers, which effectively renders them opt-out choices and difficult to interpret for researchers.

Finally, a lab 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 behavior in only the first decision period confirms the intuition of the main findings, this strategy is not ideal, as it rules out the importance of capturing more experienced decisions, which are less prone to errors (i.e., the miscalculation of privacy tradeoffs).

## 6. Conclusion

This study finds 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 primitive concerns that drive this change in users' 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, differential characteristics of secondary markets involving "third parties" versus characteristics of data recipients in the primary market do not explain the effect. This finding is also broadly consistent with Buckman et al. (2019), in which the authors find that the consequence of distributing data to a third party as a form of external secondary use yields null results. Third, by design, the experiment also rules out strategic responses to market signaling. The BDM method elicits reservation prices reflecting a person's true minimum price in exchange for not sharing data, which is not dependent on new information regarding how more or less valuable data are to a potential set of recipients.

While further research is needed to disentangle this question, the results point to preferences that relate specifically to an aversion to secondary monetization. One obvious nonnormative factor is related to fairness. Individuals feel disutility if they do not split the surplus (according to their ideal fair share) gained by a recipient from the exchange. Individuals may also feel it is unethical for others to profit from information about them, leading the individual to demand greater compensation in exchange for data-sharing. Other mechanisms can be specific to personal data markets, such as data inalienability resulting in perceived ownership of personal information: despite a decision to trade away data, individuals may still feel ownership over what they still consider to be "their" data and demand data dividends for secondary economic transactions.

The research design of this study also offers guidance and insights for future experiments on data-sharing and 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 privacy responses to notice-and-choice regimes. While I use prices as the instrument to reveal preferences, my repeated-measure design captures relative

preferences (i.e., whether they prefer to preserve privacy under one condition versus another). This accommodates contextual challenges individuals have for making point estimates about the value of their data and controls for the idiosyncratic nature of psychometric data that could be driving heterogeneous preferences.

Overall, this study explores a promising avenue for broadening privacy-related preferences that impact users' data-sharing choices. The findings from this study prove that it would be naïve to conclude that people are unconcerned with secondary data markets if studies focus only on exposure concerns and context-specific usage risks. Instead, the form in which data are made available to the secondary market is important, particularly when one's "own" data can be economically exploited by others. This study provides rich and consistent evidence demonstrating how privacy choices are affected by secondary market monetization. The findings are informative for theoretical privacy models and have implications for the design of data markets and privacy policy regimes.

## References

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

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

Acquisti A, Gross R (2009) Predicting social security numbers from public data. *Proceedings of the National Academy of Sciences* 106(27):10975–10980.

Acquisti A, Grossklags J (2005a) Privacy and rationality in individual decision making. *IEEE Security and 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, Technlologies, and Practices*, 363–374 (New York and London: Taylor & Francis Group).

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

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

Adjerid I, Acquisti A, Brandimarte L, Loewenstein G (2013) Sleights of privacy: Framing, disclosures, and the limits of transparency. *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13 (New York, NY, USA: Association for Computing Machinery).

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

Adjerid I, Peer E, Acquisti A (2018) Beyond the privacy paradox: Objective versus relative risk in privacy decision making. *MIS Quarterly* 42:465–488, URL http://dx.doi.org/10.25300/MISQ/2018/14316.

Arrieta-Ibarra I, Goff L, Jiménez-Hernández D, Lanier J, Weyl EG (2018) Should we treat data as labor? moving beyond "free". *AEA Papers and Proceedings* 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.

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

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

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

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

Collis A, Moehring A, Sen A, Acquisti A (2021) Information frictions and heterogeneity in valuations of personal data. Available at ssrn: https://ssrn.com/abstract=3974826 or http://dx.doi.org/10.2139/ssrn.3974826.

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

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

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

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

*USAspendinggov* (2022) Contract to clearview ai, inc.

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

Goldberg LR (1992) The development of markers for the big-five factor structure. *Psychological 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. *Journal of Research in 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* URL 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. *The New York Times* URL 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. *Management Information Systems Quarterly* 31(1):19–33.

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

Jones CI, Tonetti C (2020) Nonrivalry and the economics of data. *American Economic Review* 110(9):2819–58.

Junglas I, Johnson N, Spitzmueller C (2008) Personality traits and concern for privacy: An empirical study in the context of location-based services. *EJIS* 17:387–402.

Koutroumpis P, Leiponen A, Thomas LDW (2019) The nature of data. *Innovation and Entrepreneurship Working Papers* (Imperial College Business School: London, UK).

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

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

Laufer R, Wolfe M (1977) Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social 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 in Psychology* 10.

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

Lin T (2020) Valuing intrinsic and instrumental preferences for privacy, working Paper.

List JA, Shogren JF (2002) Calibration of willingness-to-accept. *Journal of Environmental Economics and Management* 43(2):219–233.

Matz SC, Kosinski M, Nave G, Stillwell DJ (2017) Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences* 114(48):12714–12719.

McCrae RR, John OP (1992) The five-factor model: issues and applications. *Journal of Personality* 60(2):175–532.

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

Morewedge CK, Monga A, Palmatier RW, Shu SB, Small DA (2021) Evolution of Consumption : A Psychological Ownership Framework. *Journal of Marketing* 85(1):196–218.

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

O'Donoghue T, Rabin M (1999) Doing it now or later. *American Economic Review* 89(1):103–124.

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

Tomaino G, Wertenbroch K, Walters DJ (2021) Intransitivity of consumer preferences for privacy. INSEAD Working Paper No. 2021/50/MKT.

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

Varian H (1996) Economic aspects of personal privacy .

Westin AF (1967) *Privacy and Freedom* (Atheneum), 1st edition.

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

## Appendix A: Additional Tables & Figures

**Table 4 Attitudes on Data Privacy, Ownership, Access, and Usage**

| Label and Description | Agree | | | ... | | | Disagree | NAs |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| **Privacy** | | | | | | | | |
| "I am concerned about my data privacy when ... | | | | | | | | |
| ... I can be personally identified." | 55.1 | 25.7 | 11.4 | 2.9 | 2.4 | 1.3 | 1.2 | 0.0 |
| ... others can easily access my information (i.e. it is unsecure)." | 56.2 | 25.6 | 10.4 | 2.8 | 1.7 | 2.1 | 1.2 | 0.0 |
| ... others can use my information for their own purposes." | 46.5 | 26.6 | 13.9 | 5.6 | 4.0 | 2.1 | 1.3 | 0.0 |
| ... my information is highly sensitive." | 76.5 | 14.4 | 3.9 | 2.3 | 1.2 | 0.9 | 0.8 | 0.0 |
| **Ownership** | | | | | | | | |
| "I believe I should be able to ... | | | | | | | | |
| ... choose who I share my data with." | 65.5 | 24.9 | 5.5 | 2.2 | 0.8 | 0.5 | 0.5 | 0.2 |
| ... exclude others from accessing my data." | 60.3 | 26.3 | 7.6 | 2.8 | 1.8 | 0.7 | 0.5 | 0.1 |
| ... retract my data after I have shared it." | 48.1 | 24.4 | 11.6 | 6.8 | 5.1 | 2.5 | 1.2 | 0.2 |
| **Sharing** | | | | | | | | |
| "If I share my personal data with a third party, I believe they should be able to ... | | | | | | | | |
| ... use my data for their own purposes." | 9.3 | 24.4 | 21.9 | 9.8 | 9.8 | 10.8 | 13.7 | 0.3 |
| ... use my data to make money." | 4.9 | 10.6 | 15.2 | 9.9 | 14.2 | 16.9 | 28.1 | 0.1 |
| ... sell my data to another party." | 3.5 | 4.7 | 6.5 | 5.6 | 12.6 | 22.6 | 44.4 | 0.2 |
| **Expectations** | | | | | | | | |
| "When I share my personal data with a business, I expect that they will ... | | | | | | | | |
| ... use my data to understand me better as a customer." | 34.0 | 42.0 | 16.1 | 3.5 | 1.5 | 1.6 | 1.3 | 0.0 |
| ... use the information to provide better products and services to me." | 34.0 | 39.6 | 16.6 | 4.4 | 2.5 | 1.5 | 1.3 | 0.0 |
| ... sell my data to other businesses." | 11.4 | 14.3 | 15.1 | 8.1 | 11.5 | 17.5 | 22.1 | 0.0 |
| ... use my data to understand whether I would be willing to pay more for their products or services." | 23.4 | 37.1 | 19.3 | 7.7 | 5.6 | 3.9 | 3.1 | 0.0 |

*Note:* All values are percentages out of total sample. 1=Strongly agree, ..., 7=Strongly disagree.

**Table 5 Participation and Prices Between-Subjects Comparison by Decision Period**

| **Variables** | **Logit** | | | **OLS** | | |
|---|---|---|---|---|---|---|
| Dependent: | *Participation* | | | *Price* (\$) \| *Participation*= 1 | | |
| Comparison: $[1, N]$ | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) |
| | | | **Decision Period** $t = 1$ | | | |
| (Intercept) | 1.121*** | 1.108*** | 1.567** | 1.540 | 1.562 | 1.489 |
| | (0.135) | (0.168) | (0.493) | (0.074) | (0.094) | (0.276) |
| High Exposure $[30, N]$ | -0.331† | -0.331† | -0.343† | 0.021 | 0.020 | 0.044 |
| | (0.184) | (0.184) | (0.188) | (0.103) | (0.103) | (0.104) |
| Salient Secondary Monetization $[1, F]$ | -0.458* | -0.458* | -0.451* | 0.050 | 0.050 | 0.055 |
| | (0.182) | (0.182) | (0.185) | (0.106) | (0.107) | (0.107) |
| Observations | 892 | 892 | 892 | 625 | 625 | 625 |
| | | | **Decision Period** $t = 2$ | | | |
| (Intercept) | 1.399*** | 1.244*** | 1.687*** | 1.531 | 1.574 | 1.107 |
| | (0.145) | (0.175) | (0.508) | (0.069) | (0.092) | (0.273) |
| High Exposure $[30, N]$ | -0.501** | -0.505** | -0.489* | 0.094 | 0.093 | 0.108 |
| | (0.194) | (0.194) | (0.198) | (0.101) | (0.101) | (0.102) |
| Salient Secondary Monetization $[1, F]$ | -0.815*** | -0.824*** | -0.813*** | 0.119 | 0.127 | 0.110 |
| | (0.189) | (0.190) | (0.193) | (0.104) | (0.104) | (0.105) |
| Observations | 891 | 891 | 891 | 640 | 640 | 640 |
| | | | **Decision Period** $t = 3$ | | | |
| (Intercept) | 1.416*** | 1.339*** | 2.187*** | 1.421 | 1.444 | 1.320 |
| | (0.146) | (0.177) | (0.522) | (0.070) | (0.093) | (0.274) |
| High Exposure $[30, N]$ | -0.457* | -0.461* | -0.477* | -0.066 | -0.066 | -0.081 |
| | (0.196) | (0.196) | (0.200) | (0.103) | (0.103) | (0.103) |
| Salient Secondary Monetization $[1, F]$ | -1.369*** | -1.376*** | -1.456*** | 0.404 | 0.406 | 0.397 |
| | (0.187) | (0.187) | (0.192) | (0.113) | (0.113) | (0.114) |
| Observations | 890 | 890 | 890 | 605 | 605 | 605 |
| | | | **Decision Period** $t = 4$ | | | |
| (Intercept) | 1.412*** | 1.243*** | 1.679*** | 1.387 | 1.404 | 1.011 |
| | (0.146) | (0.176) | (0.498) | (0.072) | (0.097) | (0.292) |
| High Exposure $[30, N]$ | -0.448* | -0.449* | -0.443* | 0.026 | 0.026 | 0.035 |
| | (0.196) | (0.197) | (0.199) | (0.105) | (0.105) | (0.105) |
| Salient Secondary Monetization $[1, F]$ | -1.412*** | -1.427*** | -1.449*** | 0.361 | 0.362 | 0.387 |
| | (0.187) | (0.188) | (0.192) | (0.116) | (0.117) | (0.117) |
| Observations | 891 | 891 | 891 | 602 | 602 | 602 |
| Sample controls | | × | × | | × | × |
| Psychometric controls | | | × | | | × |
| Demographic controls | | | × | | | × |

*Note:* Standard errors in parentheses.
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$

**Figure 6 Item Score Correlations for Extraversion Trait (Factor I).**


**Figure 7 Item Score Correlations for Agreeableness Trait (Factor II).**

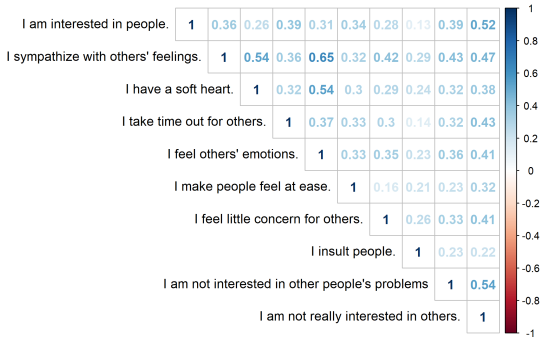



**Figure 8 Item Score Correlations for Conscientiousness Trait (Factor III).**

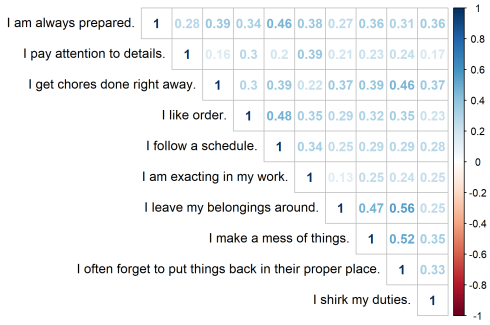


**Figure 9 Item Score Correlations for Emotional Stability Trait (Factor IV).**

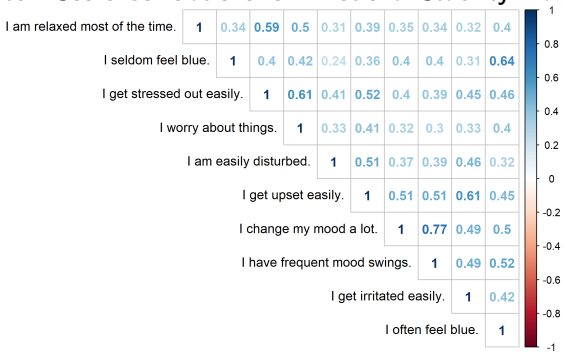
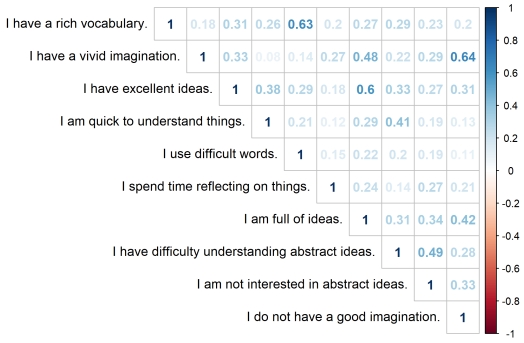



**Figure 10 Item Score Correlations for Intellect Trait (Factor V).**



## Appendix B: Experimental Design

**Figure 11 Survey Information on User-Generated Self-Assessment Scores**

**Thank you for completing your self-assessment!**
Based on your responses, your core personality has been measured across five factors: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect. Each factor score is measured on a scale from 10 to 50.

| First Name | Last Name | Extraversion | Agreeableness | Conscientiousness | Emotional Stability | Intellect |
|---|---|---|---|---|---|---|
| **Jane** | **Doe** | **19** | **30** | **31** | **42** | **48** |

Note: Each factor is measured on a scale from 10 to 50.

The 50-item questionnaire you completed is a widely used personality measure based on the Five Factor Model. Extensive research has been done to relate these questions to behavioral and psychological phenomena. The Five-Factors (or "Big 5") are set of essential traits fundamental to your core personality. Each trait is measured across a spectrum of extremes. For example, a low score on extraversion would mean high introversion.

• **Extraversion (or surgency):** Measures assertive, energetic, or outgoing behaviors. A high score indicates high extraversion, and a low score indicates low extraversion.
• **Agreeableness:** Measures empathy, sympathy, and kindness. A low score indicates low agreeableness, and a high score indicates high agreeableness.
• **Conscientiousness:** Measures your sense of responsibility, duty, and foresight. A low score indicates low conscientiousness, and a high score indicates high conscientiousness.
• **Emotional stability (or neuroticism):** Measures irritability and moodiness. High scores indicate high emotional stability (low neuroticism), low scores indicate low emotional stability (high neuroticism).
• **Intellect (or imagination):** Measures inquisitiveness, openness to new experience, thoughtfulness, and propensity for intellectually challenging tasks. High scores indicate high intellect.

**Figure 12 Instruction page for data-sharing choices.**

**You and all other participants in this study will be choosing whether to share personal data from this survey.** This data includes first name, last name, and self-assessment scores generated from the same quiz you just took (displayed in the table below).

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

Over the next few pages, **you will be shown different possible scenarios for sharing this data.** Each scenario asks whether or not you would accept certain amounts of money for releasing your data. It is in your best interest to answer honestly, as **one of these scenarios will be randomly selected and made real.**

Any data you and other participants release will be sent together in an email to the recipient(s) approximately 1 week after the survey's participation deadline.

When you are ready, please continue.

NEXT →

*Note.* Respondent's name and score from self-assessment populated in data table.

**Figure 13 Example Self-Assessment Questions and Responses.**

| | Very Inaccurate | Moderately Inaccurate | Neither Accurate Nor Inaccurate | Moderately Accurate | Very Accurate |
|---|---|---|---|---|---|
| I am relaxed most of the time. | ○ | ○ | ○ | ● | ○ |
| I leave my belongings around. | ○ | ○ | ○ | ○ | ● |
| I have difficulty understanding abstract ideas. | ○ | ● | ○ | ○ | ○ |
| I pay attention to details. | ○ | ○ | ○ | ● | ○ |
| I keep in the background. | ○ | ● | ○ | ○ | ○ |

**Table 6 Survey Text Used for Each Data Sharing Condition**

| Condition | Survey Text |
|---|---|
| $[1, N]$ | "One participant is randomly selected to receive personal data released from you and other participants." |
| $[30, N]$ | "Thirty participants are randomly selected to receive personal data released from you and other participants." |
| $[1, P]$ | "One participant is randomly selected to receive personal data released from you and other participants. If this participant has your data, they can use your data to make money. The more data they have from participants in this study, the more money they can make." |
| $[1, F]$ | "One participant is randomly selected to receive personal data released from you and other participants. If this participant has your data, they can use your data to make money by selling it to a third party. The more data they have from participants in this study, the more money they can make." |
| $[1, F']$ | "One participant is randomly selected to receive personal data released from you and other participants. If this participant has your data, they can use your data to make money by selling it to another participant. The more data they have from participants in this study, the more money they can make." |
| $[30, F]$ | "Thirty participants are randomly selected to receive personal data released from you and other participants. If these participants have your data, they can each use your data to make money by selling it to a third party. The more data they have from participants in this study, the more money they can make." |

**Table 7 Randomization Group and Condition Orders**

| Sample | Group | Conditions (in order) | | | |
|---|---|---|---|---|---|
| | | **t = 1** | **t = 2** | **t = 3** | **t = 4** |
| 1 | 1 | $[30, N]$ | $[1, N]$ | $[1, P]$ | $[1, F]$ |
| 1 | 2 | $[1, N]$ | $[30, N]$ | $[1, F]$ | $[1, P]$ |
| 1 | 3 | $[1, P]$ | $[1, F]$ | $[30, N]$ | $[1, N]$ |
| 1 | 4 | $[1, F]$ | $[1, P]$ | $[1, N]$ | $[30, N]$ |
| 2 | 5 | $[30, N]$ | $[1, N]$ | $[30, F]$ | $[1, F]$ |
| 2 | 6 | $[1, N]$ | $[30, N]$ | $[1, F]$ | $[30, F]$ |
| 2 | 7 | $[30, F]$ | $[1, F]$ | $[30, N]$ | $[1, N]$ |
| 2 | 8 | $[1, F]$ | $[30, F]$ | $[1, N]$ | $[30, N]$ |
| 3 | 9 | $[30, N]$ | $[1, N]$ | $[1, F']$ | $[1, F]$ |
| 3 | 10 | $[1, N]$ | $[30, N]$ | $[1, F]$ | $[1, F']$ |
| 3 | 11 | $[1, F']$ | $[1, F]$ | $[30, N]$ | $[1, N]$ |
| 3 | 12 | $[1, F]$ | $[1, F']$ | $[1, N]$ | $[30, N]$ |

**Figure 14 Survey Question for Eliciting Reservation Prices for Data-Sharing**

For each of the possible prices below, please indicate whether you would 'accept' and release your data, or 'not accept' and not release your data.

| | **I would accept** | **I would not accept** |
|---|---|---|
| \$0.01 | ○ | ○ |
| \$0.49 | ○ | ○ |
| \$0.99 | ○ | ○ |
| \$1.99 | ○ | ○ |
| \$2.99 | ○ | ○ |

If this scenario is made real, the computer will choose one of the prices. If you selected 'I would accept' at that price, then we will release your data (under this scenario's conditions), and you will earn the price. If you selected 'I would not accept' at that price, then we will not release your data, and you will not earn the price.