INFORMATION OWNERSHIP, VALUATION, AND EXPLOITATION IN DIGITAL MARKETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zhouyu Wu

August 2021

INFORMATION OWNERSHIP, VALUATION, AND EXPLOITATION IN DIGITAL MARKETS

Zhouyu Wu, Ph.D.
Cornell University 2021

This dissertation consists of empirical works examining individuals' information-sharing behavior, especially when that information is potentially valuable to the individuals' information recipient(s). I explore how rights assignments can influence one's perception of the value of an information commodity, such as an algorithm that can be exploited for profits. I also examine individuals' personal data-sharing behavior, and whether it is motivated the ability by of others to privately benefit in the usage of one's personal data. Finally, I also explore how stated beliefs about information-sharing—including its access and usage by others—compares to actual privacy-seeking behavior.

BIOGRAPHICAL SKETCH

Wu Zhou Yu was born in Fujian Provence in China on January 11, 1990. Her early years were spent in universities, including as a toddler in Xiamen University where her parents attended and lectured. In 1992, she relocated to Montreal, Canada where her parents pursued their graduate studies in mathematics and statistics at McGill University. She was a quiet child, so she was often allowed to sit-in on lectures with her parents and wander around the department alone and without fuss. However, in one linear regression class—and, apparently, to the delight of other students in the room—she uncharacteristically exclaimed "I can do that!" after witnessing the lecturer draw a straight line through a smattering of dots. This act, of course, was Wu's claim to her eventual career as an adult.

In 1997, she immigrated with her parents to the United States, where she lived in the surburbs of Chicago and later adopted "Joy" as her nickname. She attended the University of Chicago for college and decided to major in economics after her first lecture in microeconomics with Dr. Victor Lima, when she gained an immediate interest in the economic framework for understanding human decision-making. In her second year, she enrolled in an experimental economics course taught by Professor John List. In this course, she first encountered the power of randomized controlled trials in identifying causal claims and also discovered the complex challenge of designing human subjects studies. She became determined that, at some point in her future, she would adopt experimental techniques to further her understanding of economic behavior.

While her college experience fostered a desire to pursue graduate studies, she was missing a developed understanding of real-world phenomena and wanted to gain experiences and perspectives outside of the ivory tower. In one of her jobs after college, she was an economic consultant for U.S. patent infringement cases, where she learned a keen appreciation for the rights protection and commercial exploitation of valuable information, specifically in software and technology industries. Eventually, she applied to PhD programs with the intent to conduct applied economics research using experimental economics, which she ultimately pursued at Cornell University in the field of applied economics and management.

She currently resides in Munich, Germany with her partner Thomas. After completing her PhD, she will continue her postdoctoral work at Ludwig-Maximilians-Universität München.

This dissertation is dedicated to the cats that followed me home: Maximus from the courtyard garden of the old Theological Seminary in Hyde Park, and Ser Jorah from someone's backyard in Ithaca.

ACKNOWLEDGEMENTS

This dissertation was written under the supervision of Professor Aija Leiponen, Professor David Just, and Professor Vicki Bogan, all of whom were exceptionally generous with their time, advice, and encouragement. My committee chair, Professor Leiponen, had and continues to have a tremendous positive impact on my research career and academic life, which cannot be overstated. My passion for and path towards studying the digital economy is entirely inspired by Professor Leiponen. I am also infinitely grateful for her—as a firm strategy scholar—tolerance, curiosity, and enjoyment of my dogmatic interest in running experiments on individuals.

Many other faculty have also contributed greatly to my experience as a PhD student. The second chapter of this dissertation was born from my second year project under the mentorship of Professor Just, Professor Leiponen, and with great support from the second year paper seminar led by Professor Shanjun Li. The third and fourth chapters are the result of incredible support—in ideation, resources, execution, and dissemination—from my advisor Professor Leiponen. The third chapter is also a product of the critical, early feedback from Ted O'Donoghue and the members of Cornell's Behavioral Economics Research Group.

I am grateful for the community of graduate students in the fields of Applied Econmics and Management (AEM), Policy Analysis and Management, and Economics that made my experience at Cornell especially memorable. I also treasure my time spent actively involved with the AEM Graduate Student Association, which fostered a great community in our department. Thank you Professor Bogan for telling me, with utter confidence, that I am well suited for an academic career. Thank you also to Professor Arnab Basu and Professor Just for supporting me during my PhD years in their terms as Director of Graduate Studies. Thanks to all the various graduate students with whom I shared Warren 434 with over the years, which made my day-to-day especially fun. I am also grateful for the pleasant exchanges, education, and fruitful experiences I had during my visitng periods at Aalto University and ETH Zurich.

In my family life, I have the peculiar privilege of orbiting around a group of pure and applied mathematicians, so I hope they enjoy this part of my acknowledgments. I thank my mom, Linyun Zhou, for teaching me to judge excellence, innovate for efficiency, and to pave a path of finite steps towards seemingly impossible outcomes. I thank my dad, Chunzhang Wu, for teaching me the nobility of a life spent in a continuous approach—however meandering—towards unreachable boundaries, and to indulge in the art of pattern-finding along the way. I thank my partner, Thomas Bååth, the most important person I met during graduate school, for teaching me to how to cultivate my moral compass, the joy of experiencing the present-life, and for showing me how the world before my eyes has so many funny combinations I never thought possible.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Digitization has transformed individuals' economic relationship with information. Understanding how digitization shifts our understanding of the decision-making process of buying and selling information does not require new economic theory. However, there are a lot of unanswered questions about how decisions are made in information markets, which had little inspiration from real-world phenomena prior to digitization. In tangible goods markets, information is often framed as a means to an end: a parameter in a trade that can reduce frictions between parties. However, like any other commodity, information is created by costly inputs and sold to those willing to pay to consume it. Digitization has facilitated new ways to trade intangibles, leading to new contexts in which information is experienced and exploited as a final or intermediate good. Moreover, the rise of online platform ecosystems has enabled individuals to sell information directly to potential buyers. Questions related to individuals' decisions—especially in digital markets—over information goods they create, generate, or control are largely unexplored.

Questions related to information "ownership" is a challenge to behavioral and microeconomics studies on information commodities. A strict definition of ownership might require that consuming the owned item excludes others from similarly consuming it. Without excludability and non-rivalry, information cannot be owned in the traditional sense. Of course, there are legal and business mechanisms that can provide property rights over valuable information, such as patents, copyrights, trade secrets, non-disclosure agreements, and so on. When an economic actor experiences or exploits a protected information, these mechanisms guarantee a rights owner the means to accrue royalties from a licensee or pursue penalties from an infringer. I am interested in understanding whether these rights assignments impact the valuation of the associated information, through an ownership-related channel. Here, the loss aversion theory from behavioral economics is not present to inflate valuations—for example, you do not lose your experience, memory, and overall benefits from consuming an information the moment that another person consumes it as well.

In many other cases, there are neither "legal" mechanisms that assign rights to information access nor sophisticated technologies that trace and extract rents from its usage—for example: data. Defining access and usage rights over data is challenging, and research on measuring the value of data is still in its infancy. However, understanding what motivates data-sharing behavior is foundational to digital economics research. In particular, personal data has this potential inalienability feature that invokes psychological (i.e., perceived) ownership. Even when information is shared with another party, the person it originates from can still feel linked to or control over that data. I am interested in exploring whether one's data valuation is influenced by how one's data can benefit others in downstream markets, despite the fact that a potential buyer's private value of a commodity is assumed—in neoclassical economics—to be orthogonal to the seller's reservation price.

This dissertation consists of three empirical studies examining information-sharing and, in particular, how individuals behave in response to information that can be exploited by others. Throughout my writing, I define information "ownership" loosely and broadly: this can be in reference to a person's rights assignments associated with, perceived ownership of, or inalienability to an information commodity. I explore whether rights assignments can interact with one's perception of an information commodity's—such as an algorithm's—value. I also examine the valuation of personal data—or, individuals' privacy-seeking behavior—and whether it can be influenced by an awareness of how others can privately benefit from monetizing one's data in a secondary market. Throughout my work, I employ experimental economic methods to elicit individuals' incentive-compatible decisions over valuing and sharing information. I empirically identify causal relationships using random, exogenous manipulations of framing and information provisions for decision-makers.

Beyond the ability to identify causal claims, experimental methods are appropriate and advantageous in researching ownership rights and digital markets. When studying rights assignments using observational (i.e., non-experimental) data on individuals' choices, factors related to a person's decisions to create or own information cannot be disentangled from the influence of their ownership experience. In personal data markets, data often have zero-value explicit prices and users, instead, have implicit prices for data-sharing. Therefore, researchers using observational data cannot make conclusive statements about the economic trade-offs between privacy costs and data-sharing benefits. Moreover, the external validity of experimental methods are particularly favorable for studying digital and online behavior. Users' interaction with the digital economy is, nearly uniformly, conducted through a personal computer device. The features and experiences of trading intangible goods in the real world can be more easily simulated in a controlled setting than compared to tangible goods markets.

In the chapter *The Licensing Behavior of Owners and Creators of Algorithms*, I explore valuation and decision-making in licensing algorithms: a type of information good that contains a finite-step, computer-implementable procedure to resolve a well-defined type of problem. In *Algorithms*, I study how individuals value algorithms attempting to solve the canonical "balance puzzle." In three experimental studies, I observe behavior indicative of psychological ownership in creators and non-creator owners of potential solutions. Licensors assigned to create and/or own these algorithms reveal higher reservation prices for licensing than licensees, who do not own but can use the algorithm to gain profits. When licensors and licensees are treated with meta-information about an algorithm's likelihood of being successful, I find a lack of evidence for these disparities in valuation among higher quality algorithms. However, among low quality algorithms, valuation gaps persist, suggesting a resistance to external signals for owners of poor quality algorithms.

In the chapters *Privacy-Seeking Behavior in the Personal Data Market* and *A Comparison of Stated and Revealed Privacy Preferences*, I examine a particular form of information ownership: individuals and their personal data. Firms are looking to commercialize, trade, and monetize the personal data they collect and receive from consumers. Internet users regularly choose to disclose and share their personal data in return for goods and services. In *Privacy-Seeking Behavior*, I experimentally examine whether a data recipient's ability to exploit data in a secondary market can motivate users' privacy behavior. An online experiment elicited individuals' willingness and reservation prices for sharing their personally-identifiable psychometric data when faced with real privacy consequences. I found that individuals' information disclosure behaviors were misaligned with their willingness to allow data recipients to monetize their data and trade with a third party. Individuals behaved more privately—by refusing to share data or by demanding greater benefits in exchange for privacy losses—when they became more aware of a data recipient's ability to sell their data for money. Moreover, when individuals considered allowing access and exposing their data to many recipients, the privacy responses were weaker than the responses to just *one* recipient's exploitation abilities.

In *Comparison of Stated and Revealed*, I study the attitudinal data of individuals compared to their actual privacy behavior from the experimental study in *Privacy-Seeking Behavior.* This survey work contributes to our understanding of how normative beliefs about data-sharing policies correlate with revealed behavior (i.e., one's valuation of personal data). While this chapter contains only exploratory, correlational conclusions, the data set is novel in its contribution to understanding the (presumed) empirical gaps between stated and revealed preferences over data privacy. The purpose of this chapter is to caution more casual observers of privacy research from falsely concluding a pervasive pattern of paradoxical attitudes and actual choices.

It is my hope that this dissertation is an exemplification for studying privacy, algorithms, and data through a price theory lens, by treating information as commodities bought and sold by users. This is a model of information-sharing behavior that is easy to replicate for extensions of this research. Moreover, when adding non-normative influences to neoclassical assumptions about economic decision-making, the behavioral economics literature has a vast set of frameworks and tools that can be referenced and employed in an applied setting.

Throughout these chapters, there are some key behavioral insights that can be made about digital and information markets. First, researchers should consider the consequences of individuals' perceived ownership over information.

Whether rights assignments are granted or not, individuals' relationship with that information may influence their economic behavior. Second, there is a pervasive salience problem about the consequences of information-sharing, which sheds light on the growing relevance of information acquisition costs for economic decision-making. In algorithms markets, this can be meta-information on the quality of an algorithm, which is reasonably costly to discern for the average non-expert. In personal data markets, information about consequences that have influence over individuals' privacy trade-off can be obfuscated and opaque due to limited capacities for attention. Moving forward, regulators and managers should be equipped with empirically tested prescriptions to facilitate the decisions of agents in digital markets, which may be vulnerable to framing and inattention.

CHAPTER 2

THE LICENSING BEHAVIOR OF CREATORS AND OWNERS OF ALGORITHMS

2.1 Introduction

Many forms of information can be assigned property rights, so that the trade and licensing of information can be facilitated in commercial markets. One category of information that can be transacted in licensing arrangements is contained in software. Bessen and Hunt (2007) construct a definition of software which involves a "logic algorithm" that is not "hard-wired." Algorithms are applications of mental processes, containing a finite set of instructions for obtaining a result and can be computer-implementable. Prior to the 1980s, algorithms could not be owned *per se*, and software patents could only be awarded if it resided in a system that was not exclusively software-based.[1] Since then, what constitutes patentable software has broadened, and software patenting—and, by our definition, ownership assignments over logic algorithms—have surged.

Recently, the "algorithm economy" is set to be the next rising tide in the era of digitization, directly proceeding the previous rise in big data (Sondergaard, 2015). While the collection of vast quantities of have been equated to the new oil of the digital economy, data is in many ways a credence good. Like oil, sophisticated technologies are what catapult this resource into industrial markets. Data technologies—including proprietary algorithms—are used to refine and exploit the information contained in data, ultimately transform data from an unintelligent resource into commercial actions. The most valuable algorithms are closely

[1] *Gottschalk v. Benson* 409 U.S. 63 (1972); *Parker v. Flook* 437 U.S. 584 (1978).

guarded secrets within firms, implemented into commercial markets in the form of proprietary search, high frequency trading, driver-less car functions, and others. Aside from these more mature examples, individual users can publish and earn royalties on their own algorithms from licensing to developers on emerging platforms like *Algorithmia* (Blankenship, 2019; Chowdhry, 2018). However, little research has been done to understand individuals' licensing behavior in this new algorithm economy.

Owners and potential buyers of an algorithm can form valuations of its contents based on its potential to realize commercial profits. This valuation, which determines licensing rates and prices, is challenging, technical, and—in the case of major software patents or proprietary algorithms—often negotiated privately between parties. For the individual algorithm creator, determining a royalty amount (e.g., on a platform like *Algorithmia*) can be even more challenging and vulnerable to non-normative influences outside of maximizing expected profits. Arrow's (1962) information paradox already presents a classical meta-information asymmetry problem to the market for ideas: an idea and its qualities cannot be fully revealed to potential buyers, because knowing would deflate the idea's value. However, little research has explored whether under-trading of ideas is caused by valuation asymmetries from non-normative influences unrelated to meta-information asymmetry. These include behavioral phenomena related to rights assignments of ideas that skew from classical assumptions on the independence of ownership allocation on market outcomes (Coase, 1937, 1960). Generally, economic theories underlying the study of technology licensing activities assume that decisions are guided by maximizing expected profits. This project explores and documents how ownership has a value-enhancing effect on algorithms, leading to valuation asymmetry and under-trading between licensors and potential licensees of these information commodities.

The economic experiments in this paper contain a stylized but incentive-compatible market for licensing algorithms, where ownership rights are salient and well-defined. I examine whether an exogenously assigned experience of owning algorithms can inflate valuations—relative to those who do not own them—of their information contents, leading to under-trading of intellectual property. This question is also separately examined for creator- and non-creator-ownership. I also test for the usefulness of "shutting down" valuation asymmetry by providing third-party, meta-information on an algorithm's potential for success (i.e., the likelihood it "solves" its intended problem). This form of intervention that corrects for ownership biases has clear applications in real-world software licensing activities, such as a third-party agency that evaluates, certifies, or reviews the quality of the algorithm for non-experts.

The results from these studies confirm that the market for information such as algorithms is vulnerable to under-trading. There are valuation asymmetries between (1) those who have rights-to-exclude and (2) those who can pay-to-exploit algorithms in licensing transactions. Licensors demand higher royalties than licensees are willing to pay, resulting in a fewer successful transactions where both parties can benefit. These results support the influence of affects and heuristics related to the possession and ownership of algorithms. I also find that ownership valuation-inflation exists both in self-created and non-self-created algorithms. This is striking, since non-creator owners and potential buyers lack meta-information asymmetry about the quality of an algorithm in a controlled experimental setting. This study also finds that minimizing choice uncertainty over how likely an algorithm will realize profits can mitigate valuation asymmetry in most cases; however, owners seem less responsive than buyers towards these meta-information signals.

2.2 Related Literature

Behavioral ownership-related biases have been widely studied in many other economic contexts, especially in documenting the empirical *endowment effect*. However, when studying the licensing of algorithms, a critical deviation from that literature is that loss aversion—the most popular mechanism explaining inflated valuations in owners—cannot apply in a context where information cannot be "lost" to the owner (i.e., ideas cannot be immediately or easily forgotten, for instance) when it has been experienced by or licensed away to another. While the loss aversion mechanism is uniquely fitted to commodities that can transfer its exclusive consumption from one party to the next, it lacks insight into information and experience goods that are non-rival. Most of the endowment effect literature, therefore, studies tangible objects, although there are some empirical examinations into intangible goods like the environment (List and Shogren, 2002). Nevertheless, non-rivalry is a key characteristic that separates information commodities—like the algorithms studied in this paper—from these other intangible goods.

An alternative and understudied phenomena that does not require a gain-loss framework for valuation asymmetries between owners and non-owners, but also predictions predicts the equivalent empirical effect is *psychological ownership* (i.e., the feeling that something is "mine") (Morewedge et al., 2009; Furby, 1991). In fact, in a series of experiments that manipulate the degree to which persons feel psychological ownership, Shu and Peck (2011) find that psychological ownership is an independent construct that mediates the empirical endowment effect. Beyond tangible and material goods, psychological ownership is theorized to apply to intangible commodities such as ideas (Belk, 1988) and even to a management context of collective ownership, membership, and belonging in organizations (Pierce and Jussila, 2011). Isaacs (1933) observed how children showed feelings of ownership for rhymes and songs if they had heard them first and, subsequently, excluded others from the use or experience of these intangible goods. Despite some motivation supporting the possibility that ownership feelings are not independent from the valuation of intangibles, ownership effects in information valuation and licensing are not well understood. My paper contributes to this open inquiry into the independent role that psychological ownership plays in ownership-enhanced valuation of information commodities.

There is recent, growing interest in examining psychological ownership's role in digital, experiential, and sharing economies (Morewedge et al., 2021). In many of these new technology markets, the ownership of information is non-existent except through intellectual property rights, which bring about new questions related to the influence of the underlying mechanisms of psychological ownership. Digital and information economies are particularly fruitful for studying psychological ownership, because we know its value-enhancing effects on a target object can be divorced from the loss-framing common in tangible goods.[2] Licensing an algorithm, for example, does not entail a loss of legal ownership over the algorithm due to the non-rival nature of information. More-

[2]Here, loss aversion refers to the enhanced valuation of an object under the decision to trade-away (rather than acquire) its ownership. This is distinct from the feeling a threat of future loss (or impermanence) over objects.

over, psychological ownership can change, as motivations for decision-making may change, when comparing "solid" markets that transfer of tangible goods to "liquid" markets that sell access to experience goods (Morewedge et al., 2021).

More generally, framing effects have been found to influence the valuation of ideas. In the behavioral literature on entrepreneurship—where we can think of entrepreneurial ideas as one category of the ideas market—has recognized the role of biases and heuristics in the valuation of opportunities (Hooshangi and Loewenstein, 2018). Moreover, empirical evidence of the asymmetry in revealed prices between agents who are selling versus buying goods has been replicated in many experiments over past few decades, even for intangible public goods (List and Shogren, 2002). Understanding how the framing of decisions impacts valuation tasks is helpful in our predictions about decision-making asymmetries in algorithm licensing.

In addition to studying the decision-making of owners who may have been endowed a bundle of rights to license ideas, there is another type of agent in algorithm licensing: the creator-owner. Creator-ship's role psychological ownership is one channel that has surprisingly little research, despite the variety of possible applications. Studies have found that creator effects, such as the "IKEA effect" (Norton et al., 2012), increases one's valuation of self-assembled material objects through an experience interacting labor and love. However, the extension of such biases remains to be documented for intangible commodities like information.

2.3 Hypotheses Tested

This project examines the value-enhancing effect of information ownership in licensing transactions—where the "legal" rights (to exclude others from exploiting information) are not being transferred away. I posit that there exists a value-enhancing effect due to the rights assignments given to licensors of algorithms, leading to an empirically testable valuation asymmetry between licensors and licensees of algorithms.

There are three antecedents of psychological ownership: perceived control, self-investment, and knowledge. Control has been suggested to be critical in determining feelings of possession (Furby, 1978; Tuan, 1980, 1984). Psychological ownership can satisfy individuals motives to control and express identity to external environments (Belk, 1988). When traits associated with the self and positive self-associations are transferred to a good, this increases emotional attachment and enhances its value (Beggan 1992, Weiss and Johar 2016). Finally, a development of history and intimate knowledge between a person and a good can enhance the perceived value of that good.

It follows that any licensor's experience of control through their experience of have a "bundle of rights" over their algorithm can deepen psychological ownership, and enhance one's valuation of the owned object. In addition, if the owner is the creator of their algorithm, the experience of creating experience intensifies the self-investment and knowledge antecedents of psychological ownership. For non-creator owners, their feelings of psychological ownership can remain, despite the weaker role of self-investment and knowledge channels. For these reasons, my study expects ot find evidence consistent with how licensors have inflated valuations over the information, especially if the information is their own creation.

Property rights assignments are not a necessary requirement for a decision-maker to “feel” psychological ownership (Reb and Connolly, 2007). Therefore, it is necessary to also explore the potential for licensees to feel psychological ownership over algorithms they choose to use and exploit. However, the algorithm economy contains features that threaten the prevalence of psychological ownership among those paying for access. Not only do licensees acquire access for restricted ways to use or exploit the information, but their “consumption” of information does not prevent others from access to the information. This implies a shared ownership—where property can in theory be simultaneously experienced and used by many agents—that threatens the presence of psychological ownership feelings for the licensee (Haase and Kleinaltenkamp, 2011). Pay-for-access transactions also imply an impermanence and lack of control of the object, where the expectation to have control and continued possession of an object are threatened (Bagga et al., 2019; Bardhi and Eckhardt, 2017). The expectation to own an object in the future is also critical for maintaining a reference point for evaluating a good from a position of gain rather than something to be lost (Morewedge et al., 2021)—which fundamentally maintains the licensee’s relationship with the good as a “buyer” and never a future “owner.” For these reasons, my study expects and find evidence consistent with how licensees have weaker psychological ownership (than licensors) over information they experience, even if they intend to pay-to-use.

The first test examines whether creator-owners of algorithms make licensing decisions with perceptions enhancing their algorithm’s value:

Hypothesis 2.1 *In the absence of selection into creating algorithms, there is valuation asymmetry between the creator-licensor and potential licensee of an algorithm.*[3]

In order to proceed in finding empirical evidence, the confounding selection problem needs to be removed. There may be qualities of the individual who chooses to create an algorithm (in a natural setting) that correlates and predicts a higher valuation of his or her idea that do not exist in potential licensees of that idea. In a controlled laboratory setting, this confound can be removed with an exogenous assignment of who creates an algorithm.

Second, I posit that decision-makers also have enhanced values of their algorithms—even if they are non-creator licensors:

Hypothesis 2.2 *In the absence of selection into ownership and meta-information asymmetry, there is valuation asymmetry between the non-creator-licensor and potential licensee of an algorithm.*

Meta-information asymmetry between owners and potential buyers of ideas can form valuation asymmetries unrelated to the value-enhancing effects of psychological ownership. This is a typical feature and challenge to knowledge transfers. Knowledge sellers have an incentive to keep most of the information hidden from potential buyers to maintain the information's value for buyers; however, buyers then have only a subset of the meta-information to deduce an idea's quality (i.e., Arrow's Paradox, 1962). In a laboratory setting, meta-information symmetry can imposed on individuals who are randomly and exogenously assigned the role of licensor or licensee. Licensees in a controlled setting cannot

[3] "Valuation asymmetry" refers to a reservation price gap—between licensors and licensees—that leads to under-trading (i.e., licensors' willingness-to-accept is greater than licensees' willingness-to-pay).

solve the problem without entering a licensing agreement.

Finally, choice uncertainty in the quality of a good is one aspect that has been found to interact and even counteract the valuation asymmetries between sellers and buyers goods. Tools that reduce choice uncertainty can potentially alleviate the under-trading of algorithms, should we find the presence of value-enhancing ownership effects. While little theoretical research has supported this phenomenon, empirical evidence in willingness-to-pay versus willingness-to-accept experiments suggests this trend. Moreover, a common extension of valuation-asymmetry findings is to study more experienced decision-makers, who may be less vulnerable to ownership biases. For example, in Korting and Otto (2019), individuals who were less uncertain and experienced in the taste quality of chocolates revealed less value-enhancing effects of ownership. Whether a setting of minimal choice uncertainty can resolve ownership effects in valuations remains to be seen.

Hypothesis 2.3 *In the absence of choice uncertainty (in the quality of an algorithm), there is valuation asymmetry between the non-creator-licensor and potential licensee of an algorithm.*

While experiencing quality is accounted for in all my experiments (i.e., all studies allow participants to examine and review the content of the algorithm being valued), information valuation is nonetheless challenging. In my experimental setting, I provide meta-information on quality signals that directly map to the potential profits gained by the participant, which should minimize or even eliminate choice uncertainty.

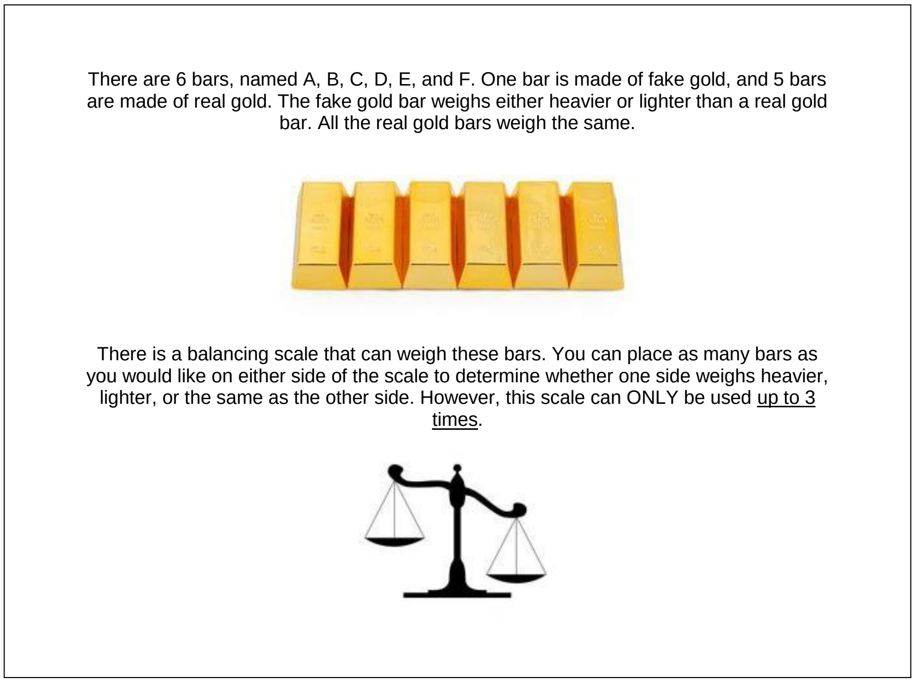
2.4 Overview of the Experiments

In the studies that follow, I demonstrate the value-enhancing effect predicted by the psychological ownership of algorithms. Individuals were assigned as either licensors and licensees, who evaluate, value, and trade solutions to a "six gold bar" logic problem (see Figure 2.1). However, instead of requiring a computer programming language, the logical steps of the algorithm were written in English. This allows the experiment to capture the essential, logic components upon which an algorithm is built from, but it does not necessitate highly specialized expertise to evaluate.

All participants were provisioned with information about their role (e.g., owner or non-owner) in their licensing transaction with an anonymous study participant. Moreover, explicit information about the total potential earnings, how the algorithm's success or failure determines the outcome of the earnings, and how royalties are paid to the owner were explained in detail prior to any decisions made by the participant (e.g., see Figure 2.2). Respondents provided their reservation prices—i.e., the minimum amount owners were willing to accept and the maximum amount non-owners were willing to pay—for licensing. All participants understood that the actual, final royalty amount is arbitrary and hidden from both parties, following Becker et al. (1964) method for eliciting true reservation prices. It was also explained to participants that it was in their best interest to answer honestly given the unknown, final royalty amount.

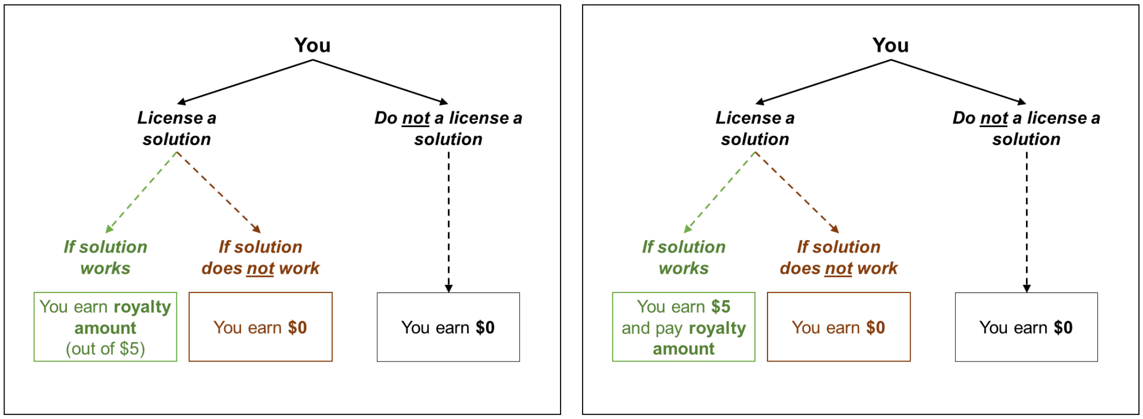
Study 1 demonstrates the main effect of value-asymmetry between owners and non-owners. Participants who create *and* do not create their algorithm reveal a higher reservation price for licensing it than compared to a potential li-

Figure 2.1: Logic Problem to be Solved with Algorithms in Experiments



censee. Moreover, stated beliefs about the quality of algorithms support wide disparities in valuation between licensors and licensees. Creator-owners have the greatest magnitudes of value-enhancement relative to licensees, in both revealed and stated valuations. Most intriguing is that the same effects exist among non-creator owners, who are simply endowed—and did not put effort into creating—their algorithms. Study 2 examines whether this effect can replicate in a setting with explicit, probability signals about the likelihood an algorithm will succeed. Valuation gaps under these quality-certain conditions seem to disappear for high and medium quality algorithms. However, there remains an asymmetry for the low quality algorithm. To understand the heterogeneous responses to meta-information across different quality algorithms, Study 3 explores the effect of this meta-information treatment for owners. Results show that owners are even less willing to widen their range of prices in revealed val-

Figure 2.2: Licensing and Earnings Directions for Owners (Left) and Non-Owners (Right)



Note: These directions were provided to Study 2 and Study 3 online participants. In Study 1, the total earnings was $15 instead of $5, and these licensing directions were provided in a lecture format (i.e., on a white board) by the experiment monitor.

uations when there is quality uncertainty.

2.5 Study 1

In this first and main experiment, I establish the value-enhancing effect of rights assignments in licensing decisions. I compare both the stated and revealed valuations of ideas between licensors and licensees who were engaged in trading a potential solution to the logic problem (i.e., an algorithm). In these transactions, *only* licensees (non-owners) had the ability to exploit the information for profits. Licensors are either creators or non-creater owners of the algorithm. However, if exploitation is successful, the royalties are paid back to the owner of the algorithm.

This experiment finds that in a controlled setting where participants are exogenously assigned (rather than self-selecting) into creating algorithms, the experience of creating supports the value-enhancing effects of psychological ownership. These results are consistent with how "creating-what-you-own" preserves psychological ownership—through self-identification and self-expression—despite the intangible nature of information. This experiment also finds that non-creator owners of algorithms preserve psychological ownership. In addition to exogenous assignment of information rights, non-creator licensors and licensees have identical meta-information about algorithms—neither group has systematically more or less information about the algorithm than the other. Despite this, it seems that the control antecedent impacted by rights assignments can preserve psychological ownership.

2.5.1 Method

At a university in northeastern United States, 167 total participants were recruited to participate in an "Intellectual Property Study" at a university lab that regularly conducts human-subjects studies.[4] Study participants were paid $5 for participating in the study and each had the opportunity to earn up to an additional $15. The algorithms (i.e., solutions to the logic problem) were created by participants who registered and attended the first two sessions of the study.[5] Each of these participants attempted to write a set of directions for solving the same logic problem shown in Figure 2.1.[6] In the proceeding laboratory sessions, 122 participants were assigned to each evaluate a randomly selected

[4]This Lab regularly recruits participants for economic studies with minimum wages of $25 per hour, paid in cash denominations. Since the lab requires phyical attendance in a room of 24 computer seats, most participants are affiliated with the university.

[5]Other than the sessions occuring in an earlier week than the subsequent sessions, no other factors in the advertisement distinguished these early sessions as being any different in content from sessions in subsequent weeks.

[6]Each individual had up to 30 minutes to complete their solutions and type them into a plain text online form; however, more time was given for those who wanted to stay a few minutes longer than the scheduled lab time. Scratch paper was provided. Access to the internet or external resources was not allowed, and independent work was enforced by the lab monitor.

solution out of the 42 created. From this process, the study was able to match 40 algorithms with four participants: A creator-licensor, an owner-licensor, and two licensees.[7]

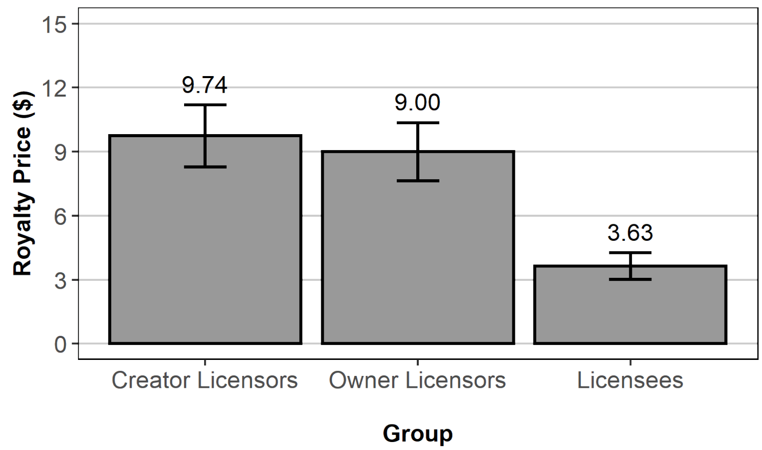
The survey measures three outcome variables of interest for each participant: (1) a stated confidence in algorithm success, (2) a stated impression of algorithm quality, and (3) a revealed offer price the participant was willing to accept or pay in royalties should the algorithm earn money for the licensee. Eliciting the participant's stated confidence proxied a measure of the probabilistic belief in the solution's success, rated on a scale from 1 to 10. Similarly, the participant's rating of the solution's overall quality—not only of the algorithm's perceived accuracy but also the clarity of its directions—was elicited as an impressionistic number between 1 and 10. The incentive compatible outcome of interest is the offer price participants were willing to accept (or willing to pay) to license the algorithm, where only successful algorithms realized profits of $15. Participants made binary choices over prices between $0 and $15, in $0.50 increments. The final, realized price is unknown to the participant, so that it was in their best interest to reveal their true valuations.

2.5.2 Results and Discussion

Overall, the experimental evidence supports a significant valuation asymmetry between licensors and licensees of algorithms. The gaps in stated valuations among licensors and licensees are aligned with the disparity in royalty prices (see Table 2.1). The average difference in offer price for algorithms among

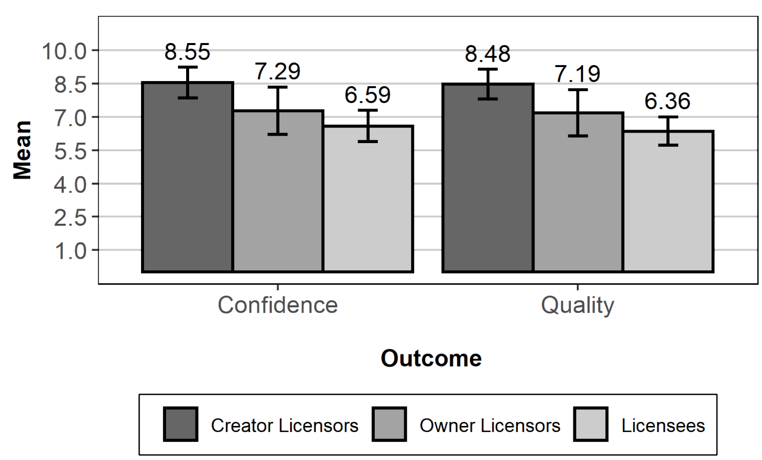
[7] Licensees were randomly and anonymously assigned to be paired with a creator or owner. There were 42 solutions in total (e.g., 42 creators), where 40 had complete matches with one owner and two licensees.

Figure 2.3: Average Price Valuations by Group



Note: These prices were elicited with a list of prices from \$0 to \$15, with \$0.50 increments. Text labels are means; error bars are 95% CIs.

Figure 2.4: Average Stated Valuations by Group



Note: These impressions were measured using a scale from 1 (lowest) to 10 (highest). Text labels are means; error bars are 95% CIs.

creator-licensors and licensees is \$6.35 (S.E. $= 0.80$), which is over 40% of earnings form a successful algorithm. Among non-creator-licensors and licensees there is an average difference of \$5.25 (S.E. $= 0.84$), which is approximately 30% of earnings from a successful algorithm. Notably, creating versus not creating one's algorithm had small (\$0.50) and statistically indistinguishable (S.E. $= 1.0$) influence on decisions.

Having created an idea is associated with higher confidence and quality impressions of that idea, relative to potential licensees of that idea. The average creator-licensor and licensee gap in stated confidence and quality ratings of algorithms is 2.45 (S.E. $= 0.56$) and 2.65 (S.E. $= 0.52$), respectively, on a scale from 1 (lowest) to 10 (highest). Interestingly, despite a gap in revealed price evaluations, there are statistically insignificant differences between stated valuations of non-creator-licensors and licensees. On the other hand, between creating and non-creating owners, there is a 1.33 ($p = 0.019$) gap in confidence and 1.35 ($p = 0.003$) gap in quality.

Table 2.1: Average Valuation Differences for Algorithms

Variable	**Mean of Valuation Differences**		
Pair	{Creator-Licensor, Licensee}	{Owner-Licensor, Licensee}	{Creator-Licensor, Owner-Licensor}
Price (\$0 to \$15)	6.350***	5.250***	0.500
	(0.803)	(0.844)	(1.033)
Confidence (1 to 10 scale)	2.450***	0.100	1.325*
	(0.559)	(0.563)	(0.567)
Quality (1 to 10 scale)	2.650***	0.175	1.350**
	(0.517)	(0.572)	(0.461)
Number of ideas	40	40	40

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Standard errors of means in parentheses. For each pair $\{A, B\}$, differences are valuations of A minus B.

An alternative description for how decision-makers reveal valuations is that their motives are influenced by fairness ideals. For example, a 50-50 split could be a heuristic that owners and non-owners use to decide on their reservation prices for licensing. From this perspective, there is an average difference in valuations among licensees and licensors that account for approximately 30%–40% of potential earnings (depending on if the owner is the original creator). For example, non-creator-owners on average reveal that nearly two-thirds of profits should be assigned to the licensor, whereas potential buyers reveal that licensors should have less than one-third of earnings.

While an analysis by actual algorithm success (hereafter referred to as the algorithm "grade") was not the original intent of Study 1, the findings of a posthoc analysis suggest that the algorithm grade (determined by an outside evaluator) may interact with valuation gaps among licensors and licensees, even with a small sample size. Among creator-buyer gaps, low quality algorithms correspond with greater difference in stated confidence but little evidence of correspondence with stated quality or revealed offer prices.[8] These findings may suggest high grade algorithms correspond to lower gaps in non-creator licensors and licensees. For the twenty five low quality algorithms, the average owner and buyer price difference is \$5.48 ($p < 0.0001$). For the fifteen high quality algorithms, the average owner and buyer price difference is \$4.87 ($p = 0.015$). However, given this small sample size and high variation in algorithms within each grade level, this difference valuation gaps among high and low quality grade algorithms is explored in Study 2.

There are several limitations to this study. Given that creators, owners, and

[8]Among low grade algorithms, the mean creater-buyer confidence difference is 2.88 ($p = 0.001$); quality difference is 2.72 ($p < 0.0001$); and price difference is \$6.34 ($p < 0.0001$). Among high grade algorithms, the mean of creator-buyer confidence difference is 1.73 ($p = 0.06$); quality differences is 2.53 ($p = 0.016$); and price difference is \$6.37 ($p < 0.0001$).

buyers were assigned by experimental session but not assigned within session, there may be unexplained, omitted variables that influence both the valuations and timing the sessions.[9] Furthermore, in the sessions for creators, each person was subject to the potential peer influence of others in the room who had the same task of creating potential algorithm for the logic problem. Perceptions of how one's effort compares to others (e.g., in relative amounts of time spent, scratch paper used, and keyboard typing activity) can influence valuations despite each individual partaking in their own, independent licensing transaction. Finally, the generalization of these algorithms cannot overcome the challenge that most ideas, software, or algorithms in the real world are unique and non-obvious. While successful algorithms are in the minority, all participants were reasonably aware of how the algorithm they were assigned to evaluate was only one of many potential algorithms present in the same room. Even though individual licensing transactions were designed to be independent of others, the licensing decisions can be non-independent due to peer influence in the laboratory room. Other boundary constraints on laboratory settings include a lack of expertise or experience in licensing activities and decisions. However, studying valuation asymmetries in the absence of selection bias—i.e., removing individual characteristics of creators that have value enhancing effects on information valuation—necessitates some costs to this generalizability issue.

[9]Creators who participated in the very first sessions, approximately one week prior to owner and buyer sessions.

2.6 Study 2

Given the presence of valuation asymmetry under meta-information symmetry (i.e., non-creator owners and potential buyers) from Experiment 1, this study examines whether this gap can be replicated in a situation with minimal choice uncertainty. This choice uncertainty is reduced across three channels. First, to reduce product quality uncertainty, meta-information signals about the algorithm's likelihood to succeed is provisioned to all participants. Second, in the absence of a physical a lab and experiment instructor, each participant is provided an on screen diagram of how their decisions impact their earnings in the event that the algorithm is successful or not. Finally, each participant evaluates three different algorithms of varying quality and experiences decisions (in a random order) to capture more experienced decisions in forming valuations of algorithms.

2.6.1 Method

At the same U.S. institution as Study 1, participants were recruited in an online (rather than in-person) study on "Evaluating Ideas" at another one of the university's human-subjects studies lab. Study participants were paid a minimum of $2 for completing the study and could earn bonuses of up to $5 depending on the outcomes of their decisions. Three ideas from Experiment 1 were chosen as algorithms to be evaluated by all participants, ranging from high quality (greater than 80% chance of success), to medium quality (a 50% chance of success), and to low quality (a less than 20% chance of success). Each algorithm contained information on the algorithm's likelihood of success, which was provided by an outside, expert reviewer. For each algorithm, participants indicated their reservation price for licensing, using a slider ranging between \$0 (i.e., any royalty is acceptable) and \$5 (i.e., the maximum possible profits). Finally, whether the participant's decision is framed in the perspective of an owner-licensor versus a potential licensee is randomly assigned.

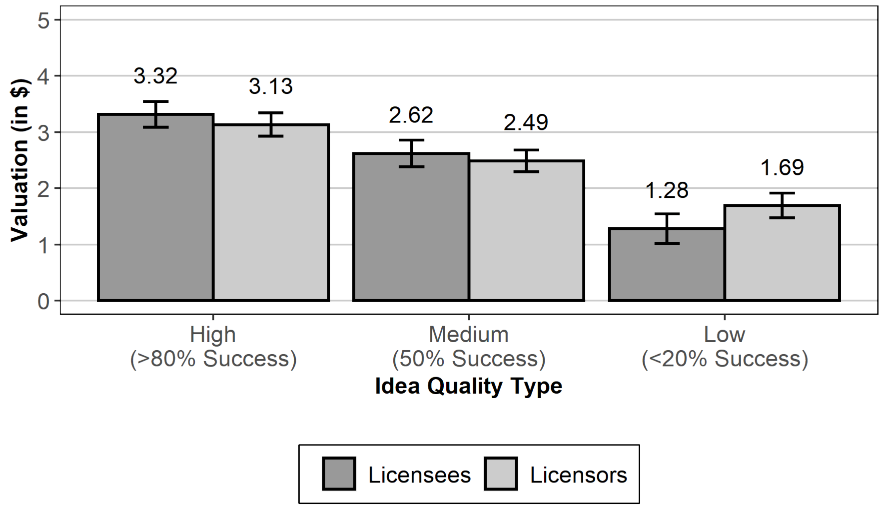
2.6.2 Results and Discussion

Out of 272 total participants, 135 were randomly assigned the role of licensee and 137 were randomly assigned the role of licensor. Licensees and licensors reveal a similar, responsive pattern of valuations among the high, medium, and low success algorithms. The average reservation prices are \$3.22 (S.E. $= 0.08$) for high success algorithms; \$2.55 (S.E. $= 0.08$) for medium success algorithms; and \$1.49 (S.E. $= 0.09$) for low success algorithms.

Unlike the first study, participants were given *explicit* signals on the algorithm's success likelihood, which directly relates to whether the licensee can exploit the algorithm and earn \$5. Among high and medium quality algorithms, average royalties among licensees and licensors have an interval of available market clearing prices that allow for licensing. However, there is a valuation gap for low quality algorithms. This low success algorithm each participant sees is labeled with a less than 20% chance of success. Licensors average, minimum price (\$1.69) is higher than licensees average, maximum price (\$1.28), leaving a \$0.41 ($p = 0.0196$) gap in licensee and licensor reservation royalty prices.

Again, one might consider the heuristics decision-makers are using to decide on valuation. This can include some fairness ideals that split profits 50-50 be-

Figure 2.5: Average Licensees' and Licensors' Valuations by Type



Note: Text labels are means; error bars are 95% CIs.

tween parties. Notably, for both high and medium quality algorithms, licensees on average allocate a greater share of profits for algorithm owners. For high quality algorithms, both licensors and licensees have a preference for splitting a greater share of profits to the algorithm owner. On the other hand, low quality algorithms tend towards a minority share of profits for the owner.

Using a panel random effects model that corrects for the non-independence of multiple responses from a single individual, I find evidence that licensing (i.e., being the algorithm owner) interacted with low success algorithms has a positive and significant effect on valuations. Results are presented in Table 2.2. In model (3), optional individual characteristics are controlled for, but the effect remains unchanged. Interestingly, employed—as opposed to students, unemployed, or underemployed—individuals are also associated with higher valuations.

Table 2.2: Effects on Algorithm Valuations by Group and Type

Variables	**OLS**		
Dependent: Royalty Price	(1)	(2)	(3)
(Intercept)	3.206***	3.315***	3.215***
	(0.104)	(0.115)	(0.238)
Licensor	0.032	-0.185	-0.237
	(0.125)	(0.155)	(0.160)
Low Quality	-1.735***	-2.034***	-2.034***
	(0.107)	(0.154)	(0.154)
Med. Quality	-0.668***	-0.696***	-0.696***
	(0.065)	(0.110)	(0.110)
Licensor $\times$ *Low Quality*		0.595**	0.595**
		(0.211)	(0.211)
Licensor $\times$ *Med. Quality*		0.056	0.056
		(0.130)	(0.130)
$\log(Duration)$			0.019
			(0.047)
Female			0.182
			(0.162)
Single			-0.211
			(0.188)
In College			0.029
			(0.156)
Employed			0.388*
			(0.180)
Individual clusters	272	272	272
Observations	816	816	816
ANOVA: Wald Test	(1),(2)	(2),(3)	(1),(3)
$Pr(> \chi^2)$	0.010	0.062	0.006

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

These results suggest that explicit signals about an idea's probabilistic success do eliminate disparities in valuations among licensors and licensees, when those signals are not describing an idea with low chances of success. One explanation for this is that licensors may have "stickier" valuations of their own ideas, and are more resistant or even in denial of external signals about their algorithm's quality. On the other hand, licensees more dramatically vary valuations of ideas based on explicit quality signals. Conversely, this is indicates a stronger willingness to accept meta-information signals for algorithms they can exploit.

2.7 Study 3

Given the persistence of valuation asymmetry for some but not all types of algorithms, a follow-up experiment examines how meta-information signals influences the decision-making of licensors across the same three different algorithm types. Thus, licensors treated with meta-information were treated identically to licensors in Study 2. Licensors without meta-information were left to evaluate the three algorithms without explicit information about the algorithms' likelihood of success.

2.7.1 Method

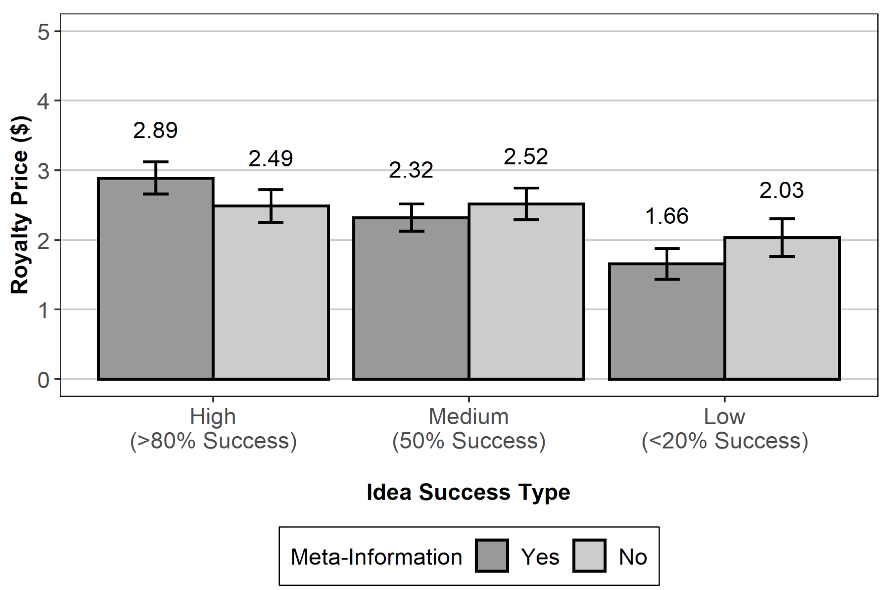
Participants were recruited again for the online study "Evaluating Ideas" at the same university lab in order to prevent duplicating participants from the prior experiment. This study is a replication of Study 2 except in the removal of licensees and the randomization of meta-information signals for licensors.

2.7.2 Results and Discussion

Out of 283 total participants, 141 were randomly assigned the role of licensor with explicit idea success information, and 142 were randomly assigned the role of licensor without explicit idea success information. Both groups reveal a similar, decreasing pattern of valuations among the high, medium, and low success ideas. However, licensors without meta-information demonstrate notably weaker variation across algorithm quality. Average prices for high and medium success algorithms with non-explicit success signals have statistically indistinguishable differences in valuations.

When licensors in Study 3 are treated to the meta-information—identical to the experience of licensors in Study 2—valuations for high quality algorithms increase and valuations for low quality algorithms decrease. Replicating the findings from Study 2, high quality algorithms with meta-information exceed the 50-50 split heuristic one might use as a benchmark for how royalty decisions are made. Also replicating findings from Study 2, are the lower share of profits licensors reveal for low-quality algorithms.

Figure 2.6: Average Licensor Valuations by Type and Meta-Information



Note: Text labels are means; error bars are 95% CIs.

Using the same empirical strategy as Study 2, I find confirmation that meta-information provisions have diverging effects on high quality versus lower quality algorithms. This suggests the prevalence of choice uncertainty in the absence of meta-information. However, whether choice uncertainty increases valuation asymmetries between buyers and sellers is unclear. The moderating effect of choice uncertainty on an ownership effect does not align with the experimental evidence on high quality algorithms,[10] where there is a possible opposing influence more consistent with a deflation of the value of high quality algorithms. However, this does not negate the findings from prior experiments, as these choice uncertainty effects on valuation are a separate category from effects related to psychological ownership.

The leading limitation to Study 3 is the inability to discern the impact of meta-information signals on licensees in addition to owners. Future studies into the this might conclude that meta-information corrects for valuation asymmetry not only on licensor inflation but also licensee deflation.

[10]For example, licensees value high quality algorithms \$0.40 more than licensors, allowing for a range of market clearing prices.

Table 2.3: Licensor Valuations by Meta-Info Treatment and Type

Variables	**OLS**		
Dependent: Royalty Price	(1)	(2)	(3)
(Intercept)	2.714***	2.486***	2.719***
	(0.109)	(0.118)	(0.270)
MetaInfo	-0.055	0.403*	0.400*
	(0.132)	(0.165)	(0.165)
Low Quality	-0.840***	-0.453**	-0.453**
	(0.103)	(0.146)	(0.146)
Med Quality	-0.269***	0.029	0.029
	(0.064)	(0.099)	(0.099)
MetaInfo × *Low Quality*		-0.776***	-0.776***
		(0.200)	(0.200)
MetaInfo × *Med Quality*		-0.596***	-0.596***
		(0.123)	(0.123)
$\log(Duration)$			-0.107†
			(0.061)
Female			0.208
			(0.155)
Single			-0.173
			(0.197)
In College			0.085
			(0.158)
Employed			-0.002
			(0.164)
Individual clusters	283	283	283
Observations	849	849	849
ANOVA: Wald Test	(1),(2)	(2),(3)	(1),(3)
$Pr(>\chi^2)$	0.000	0.330	0.000

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

2.8 Conclusion

In a series of experiments on algorithm licensing, I find that information rights assignments influence valuation asymmetries. I find evidence that the value-enhancing effects of ownership exist independent of potentially confounding influences often found in real-world settings. First, these studies remove selection into algorithm creation or ownership. Second, these experiments examine valuations in the absence of meta-information asymmetry between licensors and potential licensees. Moreover, in Study 1, valuation asymmetry—in the form of reservation royalty prices—is corroborated with asymmetry in stated valuations of these information goods. I also find support for one intervention that can correct under-trading due to these valuation gaps: meta-information provisions that reduce choice uncertainty about the underlying quality of an information commodity. Similarly, Koutroumpis et al. (2020) predict the necessity of meta-information to facilitate the trade of credence information goods, and algorithms certainly fall into this category. This can be realized in the form of a third-party certification or expert evaluators of algorithms and other information with uncertain commercial outcomes. However, valuation gaps may not be fully eliminated, as licensors demonstrate relatively more resistance to meta-information signals, especially for lower-quality algorithms.

2.9 Acknowledgments

Thank you so much to my advisors, mentors, and peers who have helped and supported me and this project along the way. Thank you to my committee members Aija Leiponen, David Just, and Vicki Bogan for their continuing encouragement. I also thank the graduate students in Economics and Applied Economics and Management at Cornell for their support, especially in participating in early pilots of this study. The continued improvement of this work was also made possibly be the comments from discussants, anonymous reviewers, and participants of the Center for Advanced Studies seminar series at LMU Munich.

CHAPTER 3

PRIVACY-SEEKING BEHAVIOR IN THE PERSONAL DATA MARKET

3.1 Introduction

Digitization has revolutionized commerce in information markets, including the market for individual information. Much of the digital economy is financed and facilitated by user-generated data—often personally or uniquely identifiable—containing information related to individuals' behaviors, intentions, and attributes. Users of various digital technologies passively and actively generate personal data in return for goods and services. Digital footprints of consumers are collected, curated, stored, and verified at very low marginal costs (Goldfarb and Tucker, 2019). Previously, firms have been mostly focused on how to analyze these data and exploit them internally. More recently, firms are looking towards opportunities to externally exploit user data and trade them in digital ecosystems (Thomas and Leiponen, 2016). Average people are often excluded from this secondary data market.

There are some prominent, recent examples of personal data exploitation in secondary markets. In 2018, the *Cambridge Analytical* 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 was used to predict the marginal voters in the 2016 U.S. presidential campaign, and even targeted those voters with persuasive information to induce shifts in their voting behavior.[1] In 2020, a new firm *Clearview*

[1] The underlying psychometric information was based on the Five Factor model, and these data are effective at predicting and even influencing people's behavior (Matz et al., 2017).

AI had harvested user-published facial images to fuel their highly successful facial recognition technology (Hill, 2020). Both of these recent cases demonstrate how the exploitation of user-generated data can be commercially valuable, even motivating public discourse on regulating secondary data markets. Interestingly, both cases also involve data—while "personal"—previously disclosed or published by the individual and, therefore, exist beyond the boundary of what economists would normally consider as information in a *private* state. Whether individuals actually want privacy regimes that grant them control over who can access and use these categories of data is still an open question.

The idea that secondary market activities may incur privacy costs is not new. Varian (1996) explained 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 the individual and third-party interests. While the costs incurred from third-party data access can potentially motivate individuals' privacy behavior, this paper examines an adjacent motivator: the *second party*'s ability to benefit from an external trade with third parties. My paper is the first, to my knowledge, to identify individuals' privacy responses to a second party's data monetization potential in external markets.

Using an economic experiment, I test whether increasing the salience about a second party's data exploitation abilities causes individuals to be more privacy-seeking and demand greater benefits in exchange for privacy-losses. To capture relative privacy valuations, the experiment compares users' privacy-seeking behavior in response to various information provisions about the conditions for releasing their data to a second party. This study specifically evaluates willingness-to-share *personally-identifiable psychometric data* in the form of Five

Factor scores (i.e., the same data exploited by *Cambridge Analytica*) in return for monetary incentives. The primary treatment condition includes explicit information about the recipient's ability to monetize data by selling to third parties. As a qualitative benchmark for understanding the magnitude of these privacy responses, a second baseline condition stipulates that data would be released to many recipients. Various secondary treatments are tested to examine more fully these privacy responses to data exploitation. These include provisions that exclude information about sales to third parties; include many recipients who can exploit data; and defines the "third party" as another person recruited for the study.

Throughout my results, I find consistent evidence that information signals about the second party's ability to exploit data motivates privacy-seeking behavior—both in their willingness-to-participate in a data market (i.e., selecting out) and in greater minimum prices (i.e., demanding more benefits) for releasing their data. These privacy responses are also stronger than the condition where data would be released to many data recipients. There are several conclusions from these findings.

First, privacy preferences are not fully revealed in any one disclosure choice. Even after individuals disclose their information in a first-stage survey to generate their psychometric data, many exhibit privacy-seeking behavior when deciding whether to share that data with additional parties. In fact, 21% of individuals reject the maximum possible price \$2.99 and choose to *not* participate in the data market (absent any information provisions about that recipient's exploitation abilities).

Second, the individual's taste for privacy is not independent from a data recipient's ability benefit from the data transfer. Increased salience about a recipient's exploitation abilities decreases the odds of participating in the data market by 2.8 times. Decomposing this main information treatment, aversion to both the recipient's ability to make money and sale to third parties separately contributes to the individual's privacy-seeking behavior. Moreover, this privacy-seeking behavior is robust to replacing the "third party" with "another participant" in the study.

Third, in the category of psychometric data, individuals are relatively insensitive to increased exposure to many second party data recipients. Compared to data-sharing with *thirty* recipients, their odds of exiting the data market were 1.8 times greater when treated with information about just *one* data recipient's exploitation abilities.

These findings contribute to our understanding of data privacy preferences in three ways. First, this study demonstrates that information disclosure behavior can be an unreliable measure of downstream data-sharing preferences. The second finding provides a new perspective on the costs associated with secondary data market activities and third party data usage, by demonstrating evidence that individuals' privacy behavior responds to the second party's *ability to benefit* from data exploitation in external markets. Finally, the third finding offers a valuable comparison between tastes for exposure versus exploitation, and finds that an aversion to others experiencing one's personal information can be less motivating for privacy behavior a category of commercially valuable data.

3.2 Related Literature

Traditional models of privacy support a theory of individuals acting strategically in information markets and calculating all the costs and benefits to data-sharing, suggesting that privacy regulations are inefficient for markets (Stigler, 1980; Posner, 1981). Following the "privacy calculus" framework from Laufer and Wolfe (1977), the privacy literature has found a wide range of concerns that enter into the cost-benefit analysis of each disclosure decision (Culnan and Armstrong, 1999; Dinev and Hart, 2006).

In addition, a behavioral economic perspective on privacy decision-making has emerged, which finds that disclosure choices respond to non-normative, environmental factors, including choice architecture, framing, perceptions of control, and contextual cues (John et al., 2010; Brandimarte et al., 2012; Acquisti et al., 2013; Adjerid et al., 2019). This paper supports and extends these applications of behavioral economics in measuring privacy preferences. A privacy response to a data recipient's private gain is—although intuitively reasonable—not economically obvious. A rational economic model would assume that the private benefits to the potential buyer should be independent of the individual's minimum willingness-to-share their personal data.

Empirical research has documented an association between individuals' privacy-seeking attitudes and the secondary-use of information (Culnan, 1993; Angst and Agarwal, 2009; Sutanto et al., 2013).[2] This study will further our understanding of secondary data markets from these studies in two ways. First, I

[2] In Culnan (1993), those less concerned (or with positive attitudes) related to secondary-use were correlated with less concern about other privacy features, including control over personal information access and the nuisance of privacy invasions. Sutanto et al. (2013) found that an information technology solution that prevents third-party data-sharing reduced perceived privacy intrusions.

focus on and measure privacy costs associated with the exploitation abilities of the second party (i.e., the data recipient) in the secondary market, rather than specific risks associated with third party uses. Second, the economic experiment in this study will elicit revealed, incentive-compatible behavioral responses to real privacy consequences. As a field experimental study by Athey et al. (2017) has demonstrated, it is challenging to use survey and stated attitudes about data privacy, as these observations may be slanted from the privacy regime individuals may actually want.

Relying on disclosure behavior is also not without its challenges. Research has argued that there is instability to valuations people place on their privacy (Acquisti et al., 2015; Adjerid et al., 2013). Most of this instability is attributed to a lack of awareness and incomplete information users have about disclosure outcomes (Acquisti and Grossklags, 2005a). My study finds upstream disclosure behavior can be misaligned with willingness-to-participate in downstream data markets, which support this understanding that the estimated value—either implicit or explicit—people place on their personal information are uncertain, unstable, and prone to misdirection.

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; Acquisti et al., 2016; Tsai et al., 2011; Posner, 1981). However, researchers are still challenged with finding what precise features or consequences of secondary data markets can actually motivate users' privacy behavior. Perhaps the closest study related to mine is Buckman et al. (2019), where the authors experimentally elicited privacy valuations that were treated with more salient, negative risks and consequences of personal data disclosure related to third party access. They found statistically indistinguishable changes to prices subjects were willing-to-accept in return for privacy losses. Deviating from past notice-and-choice interventions in the literature, my study finds one useful provision related to the recipient's private benefit of users' data, which is not independent from users' reservation prices for sharing their data. My study deepens and extends our knowledge by finding that decision-makers do reveal privacy preferences that are sensitive to the exploitation abilities of data recipients in these secondary markets, and exhibit relatively weak privacy responses to data access by and exposure to more parties.

Finally, my work contributes in improving methodology for eliciting privacy valuations. Research has found monetary rewards to be effective in motivating information disclosure (Preibusch, 2015; Xu et al., 2010; Hui et al., 2007). However, observed values vary widely. Moreover, price elicitations can be challenging to implement in privacy choice settings, where individuals have no prior experience with data prices. In the same Buckman et al. (2019) experiment, a posthoc analysis discovered suggestive evidence of individuals seeking to "price out" of the market after perceiving a greater risk of third-party data access. In order to capture this "all-or-nothing" decision, my elicitation method allows for a choice to participate *in addition* to price data (conditional on participation). My price list is coarse and reflects typical amounts seen in digital Apps that trade "free goods" in return for consumer data. Most importantly, the coarse price list is in practice a small collection of "all-or-nothing" decisions, which removes the need for individuals to form sophisticated point estimates about the value of their data.

In summary, this paper this paper will contribute to this landscape of privacy decision-making research in several important ways. First, I offer a behavioral perspective on how valuations of one's personal are non-independent from the data exploitation abilities of the data recipient. My results show the importance of this factor in privacy choices, especially on the decision to enter or exit the data market. Second, I demonstrate how disclosure behavior is unreliable in approximating individuals' willingness to forgo privacy protections and control over their data in secondary markets. Similar to Buckman et al. (2019), I address the prior literature's limitations by measuring privacy valuations by adopting a price elicitation methodology from experimental economics. However, I extend the basic elicitation design, allowing for "all-or-nothing" privacy choices in addition to pricing decisions about personal data (details are provided in Section 3.4.4). Finally, I provide empirical measures of privacy preferences over user-generated psychometric data, which deviates from the more standard demographic and identifier categories of personal data in the prior literature (see a longer discussion in Section 3.2.1).

3.2.1 The Unique Privacy and Exploitation Concerns of Psychometric Data

Prior studies have examined individuals' disclosure behavior of identifying, tracking, or sensitive contents of personal information (e.g., email addresses, shopping behaviors, sensitive secrets, and even medical history) (John et al., 2010; Acquisti et al., 2013; Athey et al., 2017; Buckman et al., 2019). This study is unique in its focus on personally-identifiable and user-generated psychometric data. It is not immediately obvious that psychometric data is important for studying privacy concerns. On the spectrum of information secrecy, Five Factor personality information is rather neutral and, arguably, observable to others (i.e., close friends and colleagues can have good estimates of one's personality score). In fact, universities and workplace settings easily and often elicit responses to Five Factor surveys, simply because individuals benefit from assessments of their personality.

However, from the perspective of party that is interested in exploiting data internally or externally, the value of consumer data is (1) not only a function of the ability to identify individuals and (2) not necessarily correlated with the degree of secrecy or (social) sensitivity of the content.[3] While the inclusion of personal identifiers is characteristic of personal data, the ability to uniquely identify may be decreasingly valuable in commercial settings. As consumer data becomes more vast and interconnected, with more sophisticated search and verification technologies, identification and identifiers may become easier—and less costly—to obtain. For example, Acquisti and Gross (2009) show how social security numbers can be predicted from publicly available data. Moreover, recent studies have discussed the dependency of privacy within social groups, such as the capability of surveillance and targeting of individuals who may have little online presence but are affiliated socially with friends who do (Acemoglu et al., 2019).

For these reasons, this study examines individuals' responses to the market value of psychometric information, in addition to their identifiers. Psychome-

[3]Consider the example of a data set linking individuals' social security numbers with their names. In an identity theft, these social security numbers have high exploitation value. However, to a firm that seeks to forecast individual-level online shopping behaviors, the randomness of how social security numbers are generated holds little explanatory power. Personal identifiers contain information valuable for surveillance and communications targeting but hold fewer dimensions of analytical power than behavioral and attitudinal data that can be linked to those identifiers.

tric data have the potential to be extremely valuable in commercial settings; data based on the Five Factor model has been shown in psychology to have the ability to understand, predict, and discriminate attitudes and behaviors among individuals in real-life outcomes (Goldberg, 1992; McCrae and John, 1992; Matz et al., 2017). The public-domain International Personality Item Pool (IPIP) for administering the Five Factor personality measurement has been pervasively used across the world for various research and assessment purposes (Goldberg et al., 2006)—which contributes to its predictive and persuasive power over human behavior and attitudes. Similarly, Miller and Tucker (2018) make this important characterization about the predictive power of a person's genetic data on future health risks. These data (unlike cookie data or email addresses) contain information about other behaviors and traits that have consequences for long-term welfare. One of the most vivid, exploitative uses of psychometric data in recent memory was, of course, conducted by *Cambridge Analytica* in targeting and persuading the behavior of marginal voters in the 2016 United States presidential campaign (Graham-Harrison and Cadwalladr, 2018).

3.3 Hypotheses Tested

3.3.1 Disclosure Behavior is Unreliable in Revealing Privacy Preferences

Classical perspectives on privacy economics theorize that data markets can sufficiently trade-off between the economic benefits of using consumer data and the privacy concerns of consumers (Posner, 1981; Stigler, 1980). This trade-off is described in Acquisti et al. (2015) and then in Buckman et al. (2019) as an individual's utility over wealth and privacy: $u(w,p)$. Suppose the consumer with p^+ amounts of privacy is considering to enter a state with $p^- < p^+$ amounts of privacy. In order for the decision-maker to agree to this change, she needs to receive at least the minimum amount in benefits (i.e., reservation price) r, where $u(w+r, p^-) = u(w, p^+)$.

The individual's ability to determine what benefits they would accept in return for privacy losses requires an ability to account for all costs to one's disclosure behavior. This is an unrealistic expectation: People often have an incomplete and limited capacity for attention when it comes to understanding the consequences of sharing personal information and the reason behind why their information is being collected (Acquisti and Grossklags, 2005b). Asymmetric information is the most obvious challenge to privacy choices—hence, much of the empirical privacy research studies privacy responses under various "notice and choice" regimes.

Lack of awareness is a uniquely challenging issue for personal data markets. Consequences may result from failed-to-imagine scenarios and incorrect presumptions about data property and exclusionary rights. As an example, facial data shared by users from decades ago—when facial recognition technology was relatively unknown—are now being harvested for commercial use by *Clearview AI* (Hill, 2020). Moreover, the complexity and novelty of data markets, the abstractness of information goods property, and the inalienability of personal data to the individual are all considerations that exacerbate the salience challenges to individuals' privacy decisions (Koutroumpis et al., 2019, 2020).

In most data markets individuals cannot re-appropriate information they have shared.[4] Therefore, the user's most consequential privacy choice is often an initial decision to disclose non-digital information, which can be used to create data. When generating personal data, individuals are often faced with tangible, immediate benefits in return. Behavioral economics research has popularized individuals' tendencies to over-weight payoffs closer to the present time (O'Donoghue and Rabin, 1999). If individuals myopically focus on immediate benefits to information-sharing while ignoring the less vivid downstream outcomes of data trading, then they do not accurately reveal their willingness-to-accept for all—and especially more opaque—privacy consequences.

Given all these obstacles towards more informed choices, observing disclosure behaviors in isolation likely leads to poor measures of the individual's true willingness-to-accept for data-sharing in *all situations* (i.e., a full transfer or "loss" of data ownership), especially for downstream markets where privacy costs are opaque at the time of information disclosure.

Hypothesis 3.1 *Individuals' upstream information disclosure choices are misaligned with their willingness-to-share personal data in downstream markets.*

An additional challenge arises in choosing which consumer "notices" should be used to correct for information asymmetry. Even if it is possible to provision the entire universe of potentially relevant privacy information to the decision-maker, this only swings the pendulum back to salience issues related to limited capacities for attention. Therefore, privacy researchers are tasked with finding the most relevant information provisions to support individuals' privacy calculus—that is, information that can meaningfully inform and motivate their

[4]Even under the European General Data Protection Regulation (GDPR) provisions on individual control rights, the non-excludable nature of data makes the enforcement of erasure rights to personal data difficult and costly to both individuals and firms.

privacy choices.

3.3.2 Impact of the Second Party's Economic Activities on Privacy Behavior

This study presents a novel privacy choice intervention: increasing transparency about exploitation in secondary data markets. There are two conclusions to be made when such an information provision can be shown to increase privacy-seeking behavior. The first is that individuals lack full awareness about the value of their data to others, along with a general lack in understanding about why their data is being collected (Acquisti et al., 2013). The second is that individuals are averse to a data recipient's ability to monetize personal data, and—when data exploitation in secondary markets is not salient—their data-sharing behavior understates their value of privacy.

External data exploitation is now a widely considered and profitable digital business strategy among firms. Consumers in the modern, digital economy have evolved beyond aversions to spying and intrusion from an anti-surveillance era (Westin, 1967). Moreover, data-based analytics have evolved privacy issues beyond the nuissance costs of unwanted solicitation theorized in Varian (1996), towards 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 the harvesting of user-published, digital goods.

Hypothesis 3.2 *Privacy-seeking behavior over personal data is stronger when it be-*

comes more salient to the individual that their data recipient is able to benefit from secondary data market exploitation.

The representative decision-maker of this study operates in a world in which her data contains commercial value for others. What motivates her privacy behavior may be determined by both anti-surveillance and anti-exploitation preferences. First, she considers her taste for the others experiencing—or "seeing"—this data. Second, she contemplates her taste for others exploiting—or "making money" from—this data in a secondary market. As examples, the number of data recipients who can access her data increases exposure, and information about her recipient's ability to earn profits after obtaining data can be a relevant signal for her exploitation tolerance.

To organize and demonstrate how privacy behavior can change in response to the salience of data exploitation, a framework for inattention is provided in DellaVigna (2009, p. 349). First, to follow the notation from the previous section, suppose some amount of (negative) utility the individual experiences from data-sharing is $V \leq u(w, p^{+}) - u(w, p^{-})$. Consider this loss in utility to be made up of two components: $V = v(e) + o(e)$.[5] The first component v is this individual's taste for others' experiencing her personal information. The second component o is this individual's taste for others exploiting her information. Both components are a function of $e \geq 0$, representing the exposure (e.g., number of data recipients) in data-sharing.[6] However, if the individual is inattentive over

[5]One can also generalize to a decision where there is positive utility gained by data and information exposure as well as exploitation; meaning taste would be negative distaste and non-aversion would be negative aversion.

[6]Intuitively and practically, e can be thought of as the number of data recipients she is sharing her information with, normalized to some recipient that is not herself. I assume that in the degenerate case $c(0) = 0$, sharing data with no one results in no dis-utility from either experience or exploitation.

the o component, she perceives $\hat{V} = v(e) + [1 - \theta(s)]\, o(e)$. Here, θ 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 non-decreasing in salient signals).

This study measures privacy responses to exogenous manipulations of exposure e and the salience of data exploitation s. The first manipulation varies the number of data recipients the individual is considering sharing data with. Thus, if changes in this number influences the individual's taste for the recipient experiencing and exploiting her data, this is revealed in her reservation price r for sharing data. The second manipulation varies the signal strength of information about the data recipient's ability to exploit personal data for profit in a secondary market, which should influence her taste for the recipient exploiting her data and not her taste for the recipient experiencing her data. If strengthening this information influences the individual's awareness for data exploitation and she is averse to this data-sharing consequence, then her taste for data exploitation is revealed in her price for data-sharing.

3.3.3 Using Responses to Data Exposure as a Qualitative Benchmark

Finally, the influence of exposure and exploitation over privacy responses can be compared. This study designs one comparison case (i.e., one and thirty data recipients) for only one particular kind of data (i.e., psychometric data). This study will use the change in privacy choices under greater exposure (to more recipients) as a qualitative benchmark for individuals' aversion to data exploitation in secondary markets. However, the results of this comparison may vary depending the context of the decision, the type of data in question, and the amount of exposure considered.

Empirical evidence has shown disclosure choices are weak and insensitive to "notice and choice" policies that increase transparency on information security. For example, Athey et al. (2017) and Adjerid et al. (2013) find that transparency about the security of personal information exchange does not greatly reduce individuals' disclosure behavior. These empirical findings suggest that transparency on information security or, alternatively, assurances for preventing others from experiencing or "invading" into one's personal information, may not always be the most important concern in the individual's privacy calculus.

Hypothesis 3.3 *Individuals' privacy-seeking response to increased data exposure can be weaker than their response to data exploitation.*

Seminal works theorizing the existence and origins of the intrinsic value of privacy would theorize—practically, by definition—that the value of privacy is determined by minimizing one's personal information from being exposed to others (Westin, 1967). However, few studies have empirically documented a relationship between the value of privacy and increased exposure to more parties who can experience the information in one's personal data. In one study I am aware of, Schudy and Utikal (2017) experimentally examined sharing personal addresses with varying numbers of data recipients, and they found that willingness-to-share decreased with exposure to more anonymous recipients.

3.4 Online Experiment

The online experiment was designed to simulate a personal data market where users generated their psychometric data and faced real decisions to share that data with others in return for benefits. It was conducted at a U.S. institution's business school lab (hereafter the "Lab"). The Lab maintains an Institutional Review Board (IRB) approved subject pool for online and in-person studies. This study was advertised with the title "How well do you know yourself? An economic decision study" in order to avoid priming potential participants with the idea that the study was meant to examine privacy preferences.[7] In fact, the words "privacy" and "security" are not used for the entirety of the study, up until the exit survey questions related to privacy attitudes.[8] Also advertised was a minimum payment of $2 in Amazon gift cards for a 15 minute study, with the possiblity to earn more based on the survey taker's decisions within the study.

The data was collected across three different samples, with a combined total of 1,188 participants in Spring 2019, Fall 2019, and Summer 2020.[9] The primary conditions, experimental paradigm, and survey recruitment and collection methods were replicated exactly for each sample collected. A set of three secondary experimental conditions varied by sample. A summary of the survey

[7]The collection and intended use of respondents' data, the data-sharing decisions subjects would be asked to make, and the real outcomes of data-sharing were approved by the University's Institutional Review Board (IRB) in 2017 prior to any interventions.

[8]For example, Adjerid et al. (2019) showed that individuals' had a higher propensity towards privacy outcomes when prompted to make decisions about their "privacy" settings versus their "survey" or "app" settings.

[9]The the four-condition, within-subject design (including, but not limited to, the method of randomization, the primary and secondary outcome elicitation methods, prices, type of data being shared) was pre-registered with the AEA RCT Registry. Prior to samples collected subsequent to Sample 1, the pre-registration was updated with the newly anticipated total number of participants.

chronology is shown in Figure 3.3. The details of the experimental design are described in the proceeding sections.

3.4.1 Collecting Personal Identifiers

The Lab was well positioned to conduct this study, as the personal identifiers of subjects (e.g., names, emails) were available to the researcher prior to the survey launch and then attached to survey takers' psychometric data during the experiment.[10] In general, 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. The subject pool is primarily made-up of students affiliated with the university, and many register with accurate personal identifiers in order to receive course credit.[11]

Other platforms with arguably more nationally representative samples could have been used (e.g., Amazon Mechanical Turk, Qualtrics Panel) for this online study. However, none offered higher quality identifiers on their participants. Eliciting identifiers within the survey could have contaminated or be contaminated by the experimental treatments. Subjects also received personalized survey links through Qualtrics to the email they registered with at the Lab. In these emails, the content was addressed to the subject's name, so that surveyers were reasonably aware that the data generated from their personality assessment could be linked to their identity.

[10] In order to have been in the Lab's subject pool and participated in any of its advertised studies, individuals needed to first register with names, email addresses, and answer some prescreening questions. Surveys were distributed through personalized email links.

[11] While this experiment only offered monetary payment, many other studies overlapping in time with this study offered course credits.

Figure 3.1: 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.	○	●	○	○	○

Figure 3.2: Example User-Generated, Psychometric Data

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

3.4.2 User-Generated Psychometric Data

This experiment collected self-disclosed responses to a 50-item Five Factor questionnaire about the respondent's attitudes, personality, and habits. This survey was taken from the standard sample of Likert-type assessment statements in the International Personality Item Pool (IPIP), which is widely used in psychology research. A small selection of the items are shown in Figure 3.1. Responses to these self-assessments generated each person's Five Factor personality scores across the traits: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect. All self-assessment statements required a response from the participant. Skipped questions prevented the survey from continuing, and a failure to complete prevented the respondent from earning the $2 participation earnings. After completing all 50 items in the self-assessment, respondents were presented with their scores and personal identifiers (see example in Figure 3.2) as well as information on how to interpret their scores.[12]

[12]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.

Similar to previous work measuring privacy valuations over personal data gathered within the study, untruthful responses could have occurred in the information people disclosed. A number of considerations were included in this design to minimize this occurrence. First, this self-assessment occurred at the start of the study, without revealing to participants that the study intended to explicitly elicit their willingness-to-share data with other persons in the study. Subjects would have been reasonably, myopically focused on the immediate benefits of receiving a personality assessment. Second, the Lab is hosted in a business department, where MBA students regularly take Five Factor score assessments in their leadership courses. Third, endogenous motivations (e.g., "nothing to hide" or "something to hide") of privacy behavior should be orthogonal to the experimental treatments in this study; certain scores are not more economically valuable that other scores and this indifference is made clear to subjects both prior to the self-assessment and prior to data-sharing decisions.[13] Fourth, it is unclear whether untruthful answers, if revealed, are more costly as a false representation of a person's characteristics. Finally, and most compellingly, the IPIP surveys have been extremely successful in countless psychology research studies in eliciting accurate and discriminating personality measures that can predict the real behaviors and traits of individuals.

3.4.3 Data-Sharing Conditions

In the second stage of the survey, conditions were created to elicit and quantify individuals' willingness-to-accept downstream data exposure and secondary

[13]Prior to the self-assessment portion of the experiment, respondents are told, "Your responses for each statement will NOT determine your earnings in this study." Prior to the data-sharing decisions, participants are not given any information that their scores determine the available prices or the exploitation abilities of data recipient(s).

market exploitation. There are six distinct conditions that vary the number of data recipients and the salience of a data recipient's exploitation abilities. For the latter variation, there are either 1 or 30 data recipients.

Table 3.1: Experimental Conditions

Condition Label	Number of Recipients	Information Provisions on Recipient's Secondary Market Data Exploitation Ability	Sample(s)
$[1, N]$	1	No info	1, 2, & 3
$[30, N]$	30	No info	1, 2, & 3
$[1, F]$	1	Full info (able to earn $ from data by selling to a "third party")	1, 2, & 3
$[1, P]$	1	Partial info (able to earn $ from data)	1
$[30, F]$	30	Full info (able to earn $ from data by selling to a "third party")	2
$[1, F']$	1	Alternative full info (able to earn $ from data by selling to a "another participant")	3

As shown in Table 3.1, the six data-sharing conditions were $[1, N]$, $[1, P]$, $[1, F]$, $[1, F']$, $[30, N]$, and $[30, F]$. All subjects made decisions under three conditions $[1, N]$, $[30, N]$, and $[1, F]$ (i.e., these conditions were replicated across all three samples). Their fourth data-sharing condition varied by sample: Sample 1 made decisions under $[1, P]$; Sample 2 made decisions under $[30, F]$; and Sample 3 made decisions under $[1, F']$. Each of the subjects' four data-sharing decisions were randomly assigned to either their first, second, third, or fourth decision period. A limit of four data-sharing decisions per participant was a feature of the design in order to reduce inattention and survey fatigue.

Prior to any data-sharing decisions, the survey explained to the subject they would be considering "several" scenarios for sharing their data with others in the study. To make the decisions incentive-compatible, they were told that one of their data-sharing scenarios would be made real at the end of the study. One week following survey submission, the experimenter randomly selected data recipients to receive personal data from other study participants and bonus earnings from a hypothetical commercial data activity.[14]

3.4.4 Eliciting Willingness-to-Share Data and Other Outcomes

Monetary rewards are not only appropriate for eliciting reservation values for sharing data, empirical privacy work has found monetary rewards to be effective in obtaining private information (Hui et al., 2007; Xu et al., 2010). For each data condition, decision-makers in the experiment considered five potential prices—\$0.01, \$0.49, \$0.99, \$1.99, \$2.99—to either accept or reject.[15] Following the Becker-Degroot-Marschak (BDM) (Becker et al., 1964) incentive-compatible method for eliciting willingness-to-accept, the subjects were told one of the prices would be selected at random to be the final price for sharing their data. 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 (**?**). For the incentive-compatibility to work, the individual should understand that the potential, final price is randomly drawn from the price range. Therefore, prior to knowing the final price, it is in the best interest of the decision-maker to reveal their true, minimum willingness-to-accept.

Implementing a BDM price elicitation is not without its challenges. First, the instructions for how to choose a minimum price given an unknown final price

[14]Subjects had a less than one percent chance of being selected as a data recipient who could earn money from others' personal data.

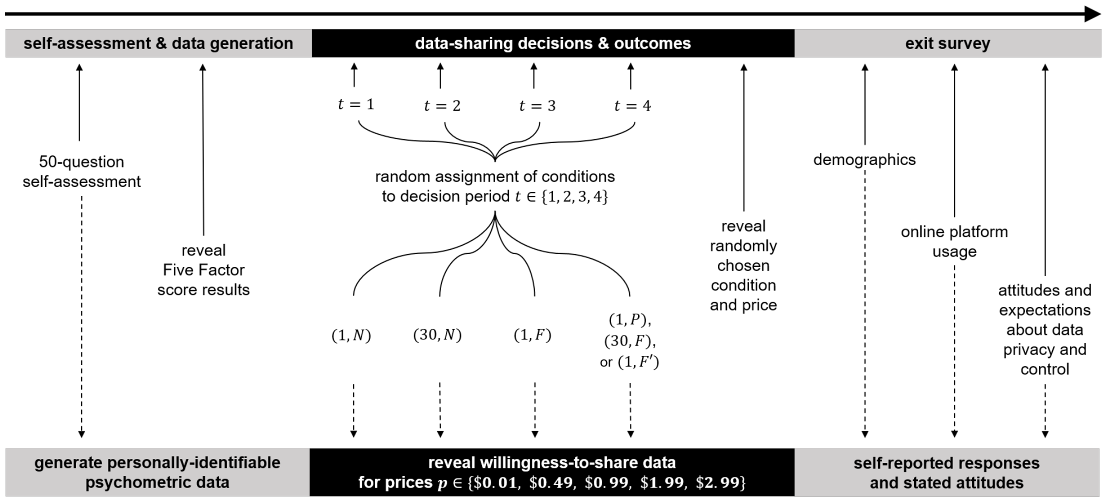
[15]The maximum price range was chosen based on pilot surveys of this study, where participants were asked to share their data under a one-price choice.

drawn from a random distribution can be unusual and confusing. It is easy to conflate choosing the *minimum* acceptable with *any* acceptable price. Moreover, it is important to recognize that data are not sold for explicit prices in the real world,[16] so it is reasonably challenging for the average decision-maker to deduce their reservation price without prior experience with market prices for data. To remedy this challenge, I use a short and coarse list of only five prices (i.e., the survey taker did not need to consider a continuum of prices), which is neither long nor granular. The survey taker could easily consider each price in sequence, and imagine whether they 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. On the other hand, price-reversing behavior in a multiple price list can happen (e.g., accepting $1.99 while rejecting $2.99). However, I did not design survey mechanisms to prevent this, in order to observe whether survey takers were attentive to the price elicitation in post-experimental analyses.

Following the data-sharing decisions and outcomes, the exit survey included voluntary questions on the individual's demographics, social media usage, and general attitudes and expectations about their data privacy. Demographic questions included gender, age, employment, marital status, and education. Social media questions included usage and frequency of using the digital platforms Facebook, LinkedIn, and Twitter. 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.

[16] Rather, there are implicit prices in trading data for goods and services.

Figure 3.3: Survey Chronology



3.5 Measures and Estimation

Given the nature of the price elicitation, acceptable prices were chosen simultaneously with whether to accept *any* price in the available range. Selecting "I do not accept" for all prices allowed individuals to exit the experiment's data market. My method of analyzing data market outcomes is to separately measure (1) the odds or likelihood of data market participation by individuals and (2) the price demanded among those who did participate. Unlike many labor market participation questions in economics that focus on changes to the price variable, the question of selecting into data markets is more interesting and relevant for firm data strategy. Real data markets operate under an "all or nothing" and rarely under a "how much" in exchange for data. In addition, while measuring price changes are useful for inferences internal to the experiment, it is difficult to map these magnitudes to external contexts, especially where benefits for data are not monetary amounts.

While I recognize there exist behavioral mechanisms that influence data market entry or exit (i.e., the selection decision) separately from price, identifying mechanisms is neither the focus of nor measurable from 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 the results of this study. First, it estimates how the explanatory variables impacted data market participation. Whether individual i chose to participate in the data market under decision-period t is indicated by

$$\mathit{Participation}_{it} = \begin{cases} 1, & y^*_{it} \leq 2.99 \\ 0, & y^*_{it} > 2.99. \end{cases} \tag{3.1}$$

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 they were willing to accept. The observed minimum acceptable price when $Participation_{it} = 1$ is

$$\mathit{Price}_{it} = \begin{cases} 0.01, & y^*_{it} \leq 0.01, \\ 0.49, & 0.01 < y^*_{it} \leq 0.49, \\ 0.99, & 0.49 < y^*_{it} \leq 0.99, \\ 1.99, & 0.99 < y^*_{it} \leq 1.99, \text{and} \\ 2.99, & 1.99 < y^*_{it} \leq 2.99. \end{cases} \tag{3.2}$$

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

$$[\textit{Participation}_{it}, \textit{Price}_{it}] = \alpha + \beta \cdot \mathbf{Condition}_{it} + \delta \cdot \mathbf{T}_t + \gamma \cdot \mathbf{Y}_i + \theta_i + u_{it} \quad (3.3)$$

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 are denoted as $\mathbf{T}_t$. These are included to capture effects specific to each decision period that can presumably affect all individuals uniformly. The variable $\mathbf{Y}_i$ is a set of participant-specific characteristics. The first characteristic includes the sample number of the individual, which is used when pooling results from all three samples. The second characteristic controls for whether the individual was randomly assigned to no exploitation information (N) in the first two decision periods and any exploitation information (P, F, or F') in the last two decision periods. Finally, as set of optional characteristics are included as an opportunity to analyze variation in privacy behavior across different types of participants. The participant-specific random effect is denoted by θ_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 (θ_i and u_{it}).

3.6 Data & Results

The data was collected from 1,188 participants in three samples spanning 16 months; each sample was conducted in four consecutive, weekly sessions. The first sample in March and April 2019 included 413 subjects; the second sample in October and November 2019 included 420 subjects; and the third sample in May and June 2020 included 355 subjects.

3.6.1 Participant Characteristics

The experiment collected self-reported activities with three major online platforms (Facebook, Twitter, and LinkedIn), and approximately 79% of individuals self-reported using Facebook with a non-anonymous account and accessed the platform at least once per week.[17] Data collected from these questions provide an approximate understanding of the survey respondents' out-of-experiment engagement with online platforms that collect user-generated data for internal or external business uses. The experimental subjects were predominantly female, in their early-20s, and students.[18] Notably, previous research on generational trends find that older people are much less likely to reveal personal information and are more concerned about their privacy (Goldfarb and Tucker, 2012). Finally, less than a quarter of participants in the sample self-identified as employed (full or part-time).

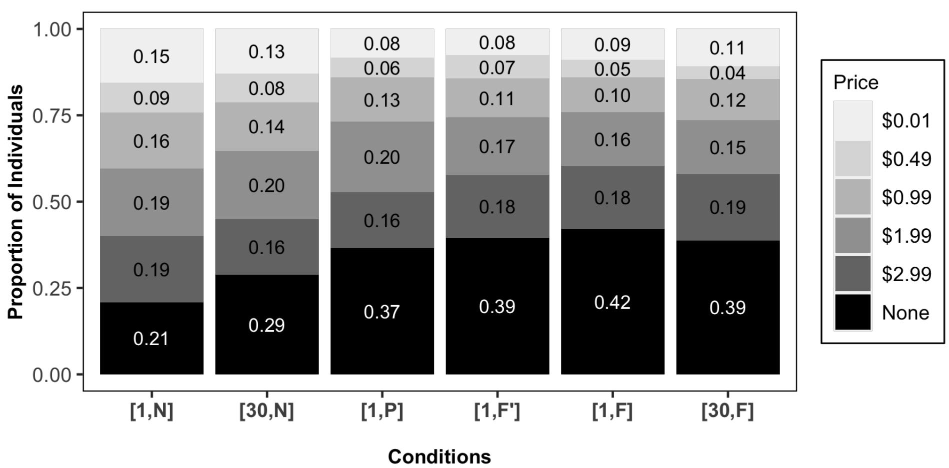
[17]Following this same definition of platform usage, 50.8% and 24.3% of participants self-reported as users of LinkedIn and Twitter, respectively. A significant minority of those surveyed reported active engagement with these platforms. For example, 19% were Facebook users that posted or shared content 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.

[18]Over 72% of participants self-identified as female. Their mean reported age was 23.8 years ($SD = 6.95$). Over 70% reported to be students as opposed to employed (full or part-time). This study's demographic makeup is not unusual compared to other studies conducted at the Lab as well as other behavioral labs at the University.

3.6.2 Descriptive Summary of Privacy Choices

After individuals disclosed personal information in the self-assessment, they displayed privacy-seeking behavior over the data created from those assessments in downstream choices (see the proportion of individuals' minimum price they were willing-to-accept by experimental condition in Figure 3.4).[19] Changes in behavior towards more private outcomes were more prevalent when there were information provisions about a data recipient's secondary market exploitation abilities. As shown in Figure 3.5, nearly a quarter of all individuals who were *willing* to share personal data with *one* recipient (in return for prices less than or equal to $2.99) were *unwilling* to share when treated with information about their data recipient's secondary market exploitation abilities.[20]

Figure 3.4: Reservation Prices for Data-Sharing by Condition



When observing each sample separately and leveraging the data-sharing conditions unique to each sample, I find further evidence that privacy choice

[19]For example, under condition $[1, N]$, approximately 60% of individuals' minimum willingness-to-accept was greater than $0.99 for sharing data with *one* other study participant and in the *absence* of any information about secondary market data exploitation.

[20]In contrast, only 10 percent of individuals were willing to participate in the $[1, N]$ market but not willing to participate in the $[30, N]$ market to share data with *thirty* recipients.

changes were more responsive to exploitation provisions than to increased exposure. Providing partial information about a recipient's "ability to make money" (i.e., excluding the secondary market information on "selling to a third party") discouraged data-sharing for 20 percent of $[1, N]$ participants who exited the data market under $[1, P]$.[21] Increasing exposure under salient data exploitation information (i.e., to *thirty* recipients in condition $[30, F]$) had surprisingly little impact on the privacy behavior of those who where willing to share data with *one* data exploiter.[22] However, provisioning secondary market activities of data recipients still increased the privacy-seeking behavior of those willing to expose their data to many recipients.[23] Finally, privacy-seeking behavior in response to provisions about a data recipient's ability to sell data was robust to using "another person" in the study (i.e., instead of using the term "third party").[24]

3.6.3 Regression on Participation and Prices

As previously mentioned, inferences are made in a two-part fashion—as opposed to assuming a natural censoring on the dependent variable—to appropriately acknowledge that the decision to select any price may follow a different decision process than the price chosen. Estimates of individuals' participation and conditional (on participation) prices include three specifications of random effects panel regressions. All regressions compare experimental conditions to a

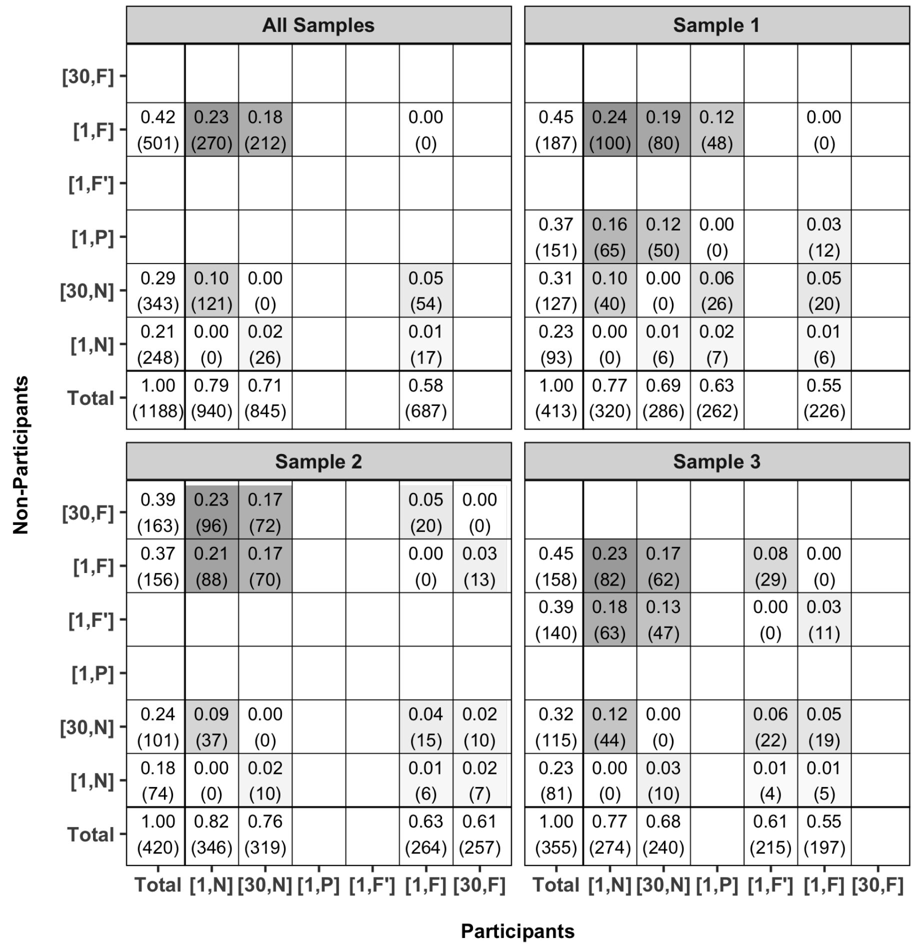
[21]Eighteen percent of those who participated under $[1, P]$ did not participate in the market under $[1, F]$.

[22]Only 8 percent of $[1, F]$ participants exited the data market under $[30, F]$.

[23]Over 22 percent of $[30, N]$ participants exited the market under $[30, F]$.

[24]Twenty-three percent of of $[1, N]$ participants exited the market under $[1, F']$. Approximately 11 percent of $[30, N]$ participants exited under $[1, F']$. And, finally, 13 percent of $[1, F']$ participants exited under $[1, F]$.

Figure 3.5: Data Market Participation and Non-Participation across Conditions



Note: Proportions out of sample and number of individuals in parentheses.

"leave-out" condition. For the sake of comparing the treatment of interest (i.e., a F, P, or F') to both $[1, N]$ and $[30, N]$, the results tables present the likelihood of participation or decreasing prices (i.e., non-privacy-seeking behavior) under the "leave-in" relative to the "leave-out" conditions.

The first, baseline model, includes dummy variables indicating the condition a privacy choice was made under, leaving out the comparison condition. The second specification extends the first by including a control for *NoInfoFirst*, to indicate whether the individual experienced no exploitation information (N)

conditions in the first two of four decisions (with F,P, or F' in the latter two). The third specification includes optional controls for individual-specific characteristics, including demographics and whether the individual scored above 30 (on a scale of 10 to 50) in each psychometric trait score.[25] All specifications include controls for decision-period, and regressions with pooled data control for sample. All regression specifications are compared using a Wald Test, to determine whether the inclusion of regressors have meaningful explanatory power.

Table 3.2: Data Market Participation and Price Results (Pooled)

Variables	**Logit**			**OLS**		
Dependent:	*Participation*			*Price* ($)		
Comparison: $[1, F]$	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	0.242*	0.473***	1.009**	3.566***	3.442***	3.095***
	(0.105)	(0.119)	(0.383)	(0.086)	(0.096)	(0.276)
$[1, N]$	1.023***	1.016***	1.037***	-0.267***	-0.284***	-0.287***
	(0.064)	(0.064)	(0.065)	(0.048)	(0.048)	(0.049)
$[30, N]$	0.590***	0.583***	0.596***	-0.247***	-0.264***	-0.267***
	(0.059)	(0.059)	(0.060)	(0.048)	(0.048)	(0.048)
NoInfoFirst		-0.370***	-0.396***		0.197*	0.213*
		(0.108)	(0.110)		(0.088)	(0.088)
Demographic controls	No	No	Yes	No	No	Yes
Psychometric controsl	No	No	Yes	No	No	Yes
Sample controls	Yes	Yes	Yes	Yes	Yes	Yes
Decision-period controls	Yes	Yes	Yes	Yes	Yes	Yes
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)
$Pr(> \chi^2)$	0.001	0.001	0.000	0.025	0.124	0.040

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

Table 3.2 summarizes the estimated impact on data market participation and

[25]Respondents who self-identified as female (versus male, non-binary, or did not wish to respond) are indicated by *Female*. Respondents who self-identified as single (never married) are indicated by *Single*. Respondents who self identified as employed part or full time (as opposed to looking for work or as student) are indicated by the variable *Employed*. Variable *Facebooker* indicates whether the respondent self-reported to be a Facebook user (who uses a non-anonymous account and accesses the Facebook platform at least once per week) and either posts or shares content on Facebook at least once a week. The dummy variables *Extravert*, *Agreeable*, *Conscientious*, *EmoStable*, and *Intellectual* indicate whether the respondent scored an score of greater than 30 in each trait, on a scale of 0 to 50.

minimum acceptable prices of $[1, N]$ and $[30, N]$ conditions relative to $[1, F]$. As shown Model (3a), the odds of participating in the data market in condition $[1, N]$ was 2.82 times greater than that of condition $[1, F]$ ($\hat{\beta}^{[1,N]}_{[1,F]} = 1.04$, $p < 0.0001$).[26] Salient information about a recipient's exploitation abilities also resulted in stronger privacy behavior than compared to considering 30 data recipients with no salient exploitation abilities.[27] Prices demanded (among respondents willing to accept a price less than or equal to \$2.99) were approximately \$0.29 and \$0.27 ($p < 0.0001$) lower under $[1, N]$ and $[30, N]$ conditions, where exploitation signals were absent.

The results shown in Table 3.3 suggest that there was both a "money" and "third party" effect.[28] Relative to $[1, P]$, the odds of participating in the data market were 2.01 times higher ($\hat{\beta}^{[1,N]}_{[1,P]} = 0.70$, $p < 0.0001$) under $[1, N]$ and 1.3 times higher ($\hat{\beta}^{[30,N]}_{[1,P]} = 0.26$, $p = 0.0007$) under $[30, N]$.[29] The results of individuals' reservation prices follow the same pattern. As shown in column (3b), data market participants chose approximately \$0.22 lower prices under conditions with no exploitation signals relative to $[1, P]$.[30]

Surprisingly, individuals exhibited little change in privacy behavior in $[1, F]$ versus $[30, F]$, suggesting that increasing exposure to more data recipients who *each* had exploitation abilities were not so different from just one data recipient

[26]For example: If C is the condition of interest and $\bar{C}$ is the baseline condition, 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.

[27]The odds of participating in the data market under $[30, N]$ was 1.81 times greater than under $[1, F]$ ($\hat{\beta}^{[30,N]}_{[1,F]} = 0.60$, $p < 0.0001$).

[28]Table 3.3 compares the Sample 1 outcomes of $[1, N]$, $[30, N]$, and $[1, F]$ conditions relative to $[1, P]$.

[29]The odds of *not* participating were 1.45 times higher ($\hat{\beta}^{[1,F]}_{[1,P]} = -0.38$, $p < 0.0001$) under $[1, F]$ relative to $(1, P)$ (i.e., the marginal "third party" effect).

[30]Price differences between $[1, P]$ and $[1, F]$ are statistically inconclusive.

having these abilities.[31] This result suggests that privacy-seeking behavior is not linearly related to the number of data recipients who can exploit one's data for profit.

Finally, replacing the words "third party" with phrase "another participant" was robust to the finding that individuals were more motivated towards privacy in response to data exploitation than to data exposure.[32] This result suggests that even when the third party is contained within the experiment—effectively sharing data with *two* recipients, this was less preferred than sharing data with *thirty* recipients without salient exploitation abilities.

Notably, the *NoInfoFirst* randomization indicates a relationship between information provision and experience in privacy decision-making. Across all samples, participants who were treated with exploitation signals in the latter half of their decisions exhibited a stronger privacy response.

3.6.4 Discussion

Throughout three replications and across various versions of the main treatment effect, results show that individuals consistently respond to information signals about a data recipient's data exploitation abilities. This result begs the question: what is the behavioral mechanism that drives these privacy responses? While further research is needed to disentangle this question, results suggest a behavioral motivation for privacy-seeking behavior beyond extrinsic factors associated with data exposure (e.g., data breaches, surveillance). One obvious

[31] Table 3.3 compares the Sample 2 outcomes of $[1, N]$, $[30, N]$, and $[1, F]$ conditions relative to $[30, F]$.

[32] Table 3.3 compares the Sample 3 outcomes of $[1, N]$, $[30, N]$, and $[1, F]$ conditions relative to $[1, F']$.

Table 3.3: Data Market Participation and Price Results (By Sample)

Variables	**Logit**			**OLS**		
Dependent:	*Participation*			*Price* ($)		
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
Comparison: $[1, P]$			**Sample 1 Only**			
(Intercept)	0.636***	0.760***	0.193	3.441***	3.336***	3.435***
	(0.117)	(0.151)	(0.583)	(0.088)	(0.113)	(0.435)
$[1, N]$	0.686***	0.677***	0.699***	-0.210**	-0.216**	-0.223**
	(0.097)	(0.096)	(0.099)	(0.072)	(0.072)	(0.071)
$[30, N]$	0.262**	0.253**	0.262**	-0.208**	-0.217**	-0.223**
	(0.094)	(0.093)	(0.097)	(0.073)	(0.073)	(0.072)
$[1, F]$	-0.363***	-0.366***	-0.379***	0.088	0.094	0.082
	(0.076)	(0.077)	(0.080)	(0.064)	(0.064)	(0.065)
NoInfoFirst		-0.261	-0.320†		0.204	0.209
		(0.180)	(0.183)		(0.141)	(0.137)
Comparison: $[30, F]$			**Sample 2 Only**			
(Intercept)	0.496***	0.790***	2.624***	3.411***	3.306***	2.368***
	(0.114)	(0.151)	(0.604)	(0.095)	(0.118)	(0.387)
$[1, N]$	1.092***	1.086***	1.152***	-0.294***	-0.310***	-0.323***
	(0.119)	(0.118)	(0.124)	(0.078)	(0.078)	(0.077)
$[30, N]$	0.698***	0.690***	0.737***	-0.277***	-0.289***	-0.299***
	(0.099)	(0.098)	(0.103)	(0.073)	(0.073)	(0.073)
$[1, F]$	0.070	0.073	0.079	0.027	0.028	0.030
	(0.058)	(0.060)	(0.065)	(0.049)	(0.049)	(0.048)
NoInfoFirst		-0.658***	-0.671***		0.219	0.219
		(0.190)	(0.195)		(0.146)	(0.145)
Comparison: $[1, F']$			**Sample 3 Only**			
(Intercept)	0.499***	0.573***	1.124	3.593***	3.467***	3.377***
	(0.122)	(0.156)	(0.766)	(0.100)	(0.126)	(0.489)
$[1, N]$	0.792***	0.786***	0.801***	-0.268**	-0.283***	-0.283***
	(0.103)	(0.102)	(0.104)	(0.083)	(0.082)	(0.082)
$[30, N]$	0.308**	0.302**	0.308**	-0.221*	-0.234**	-0.233**
	(0.100)	(0.100)	(0.102)	(0.087)	(0.087)	(0.085)
$[1, F]$	-0.209**	-0.210**	-0.214**	-0.081	-0.083	-0.085
	(0.073)	(0.073)	(0.075)	(0.066)	(0.066)	(0.066)
NoInfoFirst		-0.158	-0.163		0.252	0.248
		(0.192)	(0.196)		(0.154)	(0.153)
Decision-period controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	No	No	Yes	No	No	Yes

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

one is fairness-related—that is, individuals believe they should split the surplus gained by a recipient from the exchange. Other mechanisms can be specific to personal data characteristics, such as data inalienability and psychological ownership of personal data. Despite a transfer of information, individuals may still feel ownership over what is "their" data and demand data-royalties in exchange for exploitation.

Furthermore, why have we not seen this evidence in prior experimental works on secondary use such as Buckman et al. (2019)? One possibility is that the consequences of secondary use have previously been focused on risks related to data exposure, such as unwanted surveillance and data breaches. I do not believe my results contradict prior studies; rather, my findings extend our understanding of privacy preferences related to secondary use—specifically, on the ability of others to monetize one's data. As the landscape of privacy research has suggested: privacy-related concerns are indeed wide-ranging and can include non-normative factors. Moreover, my study accommodates for and confirms that "all-or-nothing" privacy responses are prevalent in data-sharing decisions. Valuation elicitations that do not include an opt-out choice can miss this critical set of behavioral responses to notice-and-choice regimes.

There are several conclusions to be made about the negative impact of *NoInfoFirst*—i.e., exploitation provisions coming later in decisions—on privacy responses to increased salience about the second party's monetization abilities. First, and most obviously, is that information has spillovers across conditions—i.e., individual remember exploitation signals from past decisions. Perfect information spillover is of course the expected underlying assumption for rational, privacy-decision makers. However, when the exploitation signals "leave" in the latter half of decisions, on average individuals do not fully-reveal the same set of privacy preferences. Second, it could be that there is a behavioral, non-normative privacy response to being suprised by the news of the data recipient's exploitation abilities after having made a few data-sharing choices.

There are some interesting patterns to conclude from the set of controls in the analysis (see Table B.3 in the appendix). First, there is a order effect (i.e., in the decision period number), and privacy-seeking behavior co-moves with the progression of decisions. This implies an interesting learning and experience effect that data-sharing decision-making has on revealed preferences. Second, Sample 2 (i.e., Fall 2019) has a less-privacy seeking set of behaviors in comparison to the other two samples collected in Spring and Summer. Third, those who were treated with exploitation signals in the first half of their decisions reacted with stronger changes to privacy-behavior than those in the reverse. Finally, and most interestingly, there is a negative effect on those with a high Conscientiousness score. This can include two indistinguishable conclusions: this character predicts (1) more inattentive behavior under no salient signals or (2) stronger aversion to salient signals about data exploitation. Either way, this provides evidence of at least one characteristic that may be vulnerable to a lack of salience about the monetization abilities of recipients.

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

Second, there are known to be strong anchoring effects of price elicitation survey questions. While the relative valuations in this study provides 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 is to directly ask participants to declare a minimum, acceptable price. However, as mentioned in 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 their 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. That is, 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 rejected $2.99. Wider ranges of price lists cannot easily correct for this due to the aforementioned anchoring effects. Furthermore, extremely high upper-bound prices can be less believable to decision-makers, which effectively render them opt-out choices and difficult to interpret for researchers.

Finally, and most importantly, I cannot rule out the potential significance and influence of experimenter demand effects. There are several mitigation strategies, including the inclusion of real outcomes and incentive-compatibility.

Moreover, an analysis of between-subjects behavior in only the first decision period confirms the same conclusions about aversion to data exploitation. However, as the subject proceeds through decisions two through four, I cannot disentangle and measure the degree to which these choices are more experienced privacy-decisions—and, therefore, less prone to miscalculation of privacy tradeoffs—or decisions influenced by experimenter demand.

There are some interesting patterns to conclude from the set of controls in the analysis (see Table B.3 in the appendix). First, there is a order effect (i.e., in the decision period number), and privacy-seeking behavior co-moves with the progression of decisions. This implies an interesting learning and experience effect that data-sharing decision-making has on revealed preferences. Second, Sample 2 (i.e., Fall 2019) has a less-privacy seeking set of behaviors in comparison to the other two samples collected in Spring and Summer. Third, those who were treated with exploitation signals in the first half of their decisions reacted with stronger changes to privacy-behavior than those in the reverse. Finally, and most interestingly, there is a negative effect on those with a high Conscientiousness score. This can include two indistinguishable conclusions: this character predicts (1) more inattentive behavior under no salient signals or (2) stronger aversion to salient signals about data exploitation. Either way, this provides evidence of at least one characteristic that may be vulnerable to a lack of salience about the monetization abilities of recipients.

3.7 Future Research & Implications

This work demonstrates the importance of studying privacy decision-making as a set of choices that depend on secondary data market activities. Privacy behavior responds to salient information about how parties—who acquire personal data assets—can benefit from trading and exploiting those assets in secondary markets. Even after individuals disclose information to a first party, they have privacy preferences over personal data that remain unrevealed for later, downstream markets. This work has a number of important implications for research into firms' privacy strategies, individuals' privacy preferences, and regulators' privacy policies. A number of extensions of this work should be done to continue the discourse about privacy issues in digital markets.

First, firms in the U.S. and Europe are beginning to implement new consumer privacy policies that adhere to strict guidelines for obtaining opt-in consent for data usage by third parties.[33] Research has already demonstrated GDPR policy impacts on firm strategy, effectively reducing the number of third party web connections, and market power consequences that favor larger firms (e.g., Google) (Peukert et al., 2020). Research on privacy preferences should continue to be done to understand the balance of costs to competition in digital markets and benefits to consumer privacy welfare.

Second, research on how secondary markets influence privacy preferences should be extended beyond the lab. This study was conducted in a research university setting. Individuals' propensity to share data likely depends on whether

[33]Many of these new firm strategies are responding to the European GDPR and U.S. CCPA. Moreover, firms are also investing in privacy protecting strategies in general. For example, Apple's "User Privacy and Data Use" is requiring that users of iOS 14.5 give apps consent for their data to be tracked across apps and websites owned by other companies by the end of 2021. See https://developer.apple.com/app-store/user-privacy-and-data-use/.

the collecting agency is non-profit (e.g., a research university) or for-profit (e.g., a marketing agency). The fact that self-assessment responses were so readily disclosed by those in my study likely relates to the study being marketed as a research project. Future research should examine the impact of this study's information provisions in a for-profit setting.

Third, this work is only one perspective on how individuals' privacy preferences depend on secondary market activities of data recipients. This study shows how the exploitation ability of a recipient can matter more than exposure to more recipients. The latter finding suggests that future research can study other data types—beyond the psychometric data in this study—to see whether aversion to information exploitation (by others) dominates aversion to information experience (by others).

Fourth, there are various mechanisms that potentially explain the finding that data exploitation matters for privacy behavior, which I encourage future researchers to examine both theoretically and empirically. The most compelling and privacy-specific mechanism is the theory that individuals may never fully accept a loss of ownership over their personal data—despite any economic transaction that transfers that ownership, which stems from the *inalienability* feature of personal data (Koutroumpis et al., 2019). In addition, fairness concerns (i.e., "Why should they make so much money off of my data?") and hold-out behavior (i.e., "I want the best deal they can offer for my data.") can also explain the aversion to personal data exploitation.

Fifth, beyond the environmental treatments that are exogenously assigned and examined in this study, there remains many unobserved factors that influence privacy preferences that are unaccounted for in empirical research. Future work in behavioral and psychology research can examine more challenging empirical questions about endogenous determinants to privacy choices. This study provides (non-randomly assigned) data on personality traits and privacy attitudes and, thus, some suggestive evidence that certain characteristics predict more or less privacy behavior.

Sixth, existing data property rights are inadequate and inflexible towards the usage-dependent privacy preferences of individuals. In particular, more flexible exclusionary abilities on behalf of individuals can account for differentiating disclosure and access rights between different kinds of data recipients and for different usages. Regulators should consider policies that adjust for changing downstream privacy choices. The recent data exploiting firms *Cambridge Analytical* (in the case of psychometric data) or *Clearview AI* (in the case of facial imaging data) demonstrate there are strong incentives for firms to create products and services commercialized with personal data in the absence of data licensing agreements with the individuals who supply personal. Recent theoretical work by Jones and Tonetti (2020) found that giving data property rights to firms can result in data hoarding behavior and socially inefficient outcomes, whereas consumer data property rights are more optimal and can balance the economic gains of selling data with privacy concerns. However, the long term consequences to consumer welfare in the absence of sophisticated data licensing structures has not yet been measured.

3.8 Conclusion

Understanding how individuals make privacy trade-offs is critical for firm strategies involving external exploitation of data assets. This project evolves our understanding of the motivations for individual privacy-seeking behavior in the personal data market. As consumers become better-informed agents in data markets, the way that firms collect, curate, manage, and share data will be priced-into the consumer's privacy behavior—not only in their propensity to disclose, but also in their demand for better goods and services. If consumers suffer dis-utility from being exploited in the personal data market, then this motivates privacy regulation and protections for individuals to control their personal data in digital markets. If consumers are inattentive to how their data are used in secondary markets, information provisions about how firms use and benefit from exploiting data can influence consumers' (1) decisions related to exiting the data market (e.g., data deletion requests or not consenting to third party data usage) or (2) demand greater benefits in return for data-sharing.

3.9 Acknowledgments

Thank you so much to my advisors, mentors, and peers who have helped and supported me and this project along the way. Thank you to my committee members Aija Leiponen, David Just, and Vicki Bogan for their continuing encouragement. Special thanks to Ted O'Donoghue and Chris Forman for their encouraging and helpful guidance. I also thank the members of the Behavioral Economics Research Group at Cornell and graduate students in Applied Economics and Management at Cornell for their support. The continued improvement of this work was also made possibly be the comments from discussants, anonymous reviewers, and participants of the TIME Colloquium in Munich, ETH Zurich, LMU Munich, 2021 Academy of Management Annual Conference, 2019 Consortium on Competitiveness and Cooperation Doctoral Conference.

The funding for this this project was provided by my Professor Aija Leiponen's Cornell Unversity's Institute for Social Science Small Research Grant. The ability to publish is not tied to the results of this paper. This experiment was approved by the Cornell University Institutional Review Board (#1806008035) and pre-registered at the American Economic Association's Randomized Control Trials Registry (AEARCTR-0004005). All errors are my own.

CHAPTER 4

A COMPARISON OF STATED AND REVEALED PRIVACY PREFERENCES

4.1 Introduction

A common conclusion among researchers is that there is an empirical "privacy paradox" between individuals' privacy attitudes and privacy behavior (Norberg et al., 2007). Individuals' stated concerns about personal data privacy are unreliable in predicting actual willingness-to-share data. Observational studies point out that individuals often "give away" their data for small benefits (Athey et al., 2017). However, many cultural, political, and corporate agendas are in favor of stronger privacy protection regimes.[1] For example, Apple's recent iOS 14.5 update only allows Apps to track their users' data across third parties after users opt-in and provide their explicit consent.[2]

There are underlying assumptions to the paradox conclusion that may not portray an accurate understanding of individuals' preferences about privacy. First, it assumes that stated preferences are stable and predict behavior. However, talk is often cheap to do, and social norms about what individuals ought to say can easily sway their statements away from true beliefs. Second, it assumes that general attitudes and abstract concerns about rights to personal privacy can directly apply to specific information-sharing choices online. In fact, the fram-

[1]See, as examples, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016, https://ec.europa.eu/info/law/law-topic/data-protection; CALIFORNIA CONSUMER PRIVACY ACT, 2018 *Cal. Legis. Serv.* Ch. 55 (A.B. 375) (WEST), https://oag.ca.gov/privacy/ccpa.

[2]"User Privacy and Data Use," *Apple Inc.*, 2021, https://developer.apple.com/app-store/user-privacy-and-data-use/ (accessed: June 15, 2021).

ing of privacy decisions in the real world is far removed from considerations of privacy-related rights, norms, and policies.

In Chapter 3, I examine an incentive-compatible personal data-sharing market for respondents to trade real, identifiable psychometric data with others in return for monetary benefits. Consumers make similar trade-offs everyday in their digital activities. Researchers, including myself, may classify these decisions as an observation of users' "privacy calculus" (Laufer and Wolfe, 1977), but whether general beliefs and normative stances about privacy regimes are related to these real data-sharing activities remains an open question. Moreover, as I show in Chapter 3, disclosure behavior is a poor proxy for privacy preferences, as the decision-maker can be vulnerable to a lack of awareness about secondary data market activities. These same salience vulnerabilities should be similarly concerning for researchers when eliciting beliefs about privacy, especially in the absence of explicit risks, consequences, and outcomes for data-sharing.

Privacy belief questions are varied and context dependent. Westin's surveys spanning from the 1970s to 1990s occurred prior to the era of rising digitization and user-generated data. Beliefs about government surveillance popular at the time are distinct from today's concerns over firms' predictive and persuasive power over online behavior. Researchers cannot easily empirically analyze responses to "are you concerned about privacy?" without a context for the potential privacy violation, a specific aspect of privacy, or potential consequences of privacy losses. More normative questions about privacy policies are also needed to assess preferences for regimes separately from stances on privacy worries. Finally, the strength of an individual's awareness about the free movement of data in secondary markets should also be captured to understand its interaction with privacy preferences.

The data used in my study include a collection of likert-type attitudes about (1) data security and exposure concerns, (2) control rights over one's data, (3) attitudes towards data-sharing, and (4) expectations about how data is being used. While binary yes or no responses are easy to standardize, they lack information about the strength or uncertainty of respondents' agreement or disagreement. Likert-type scales can capture some information about users' strength (or lack of) certainty in agreement.

To understand how stated attitudes can align with actual privacy behavior, evaluating them in close proximity is one method. Evaluating attitudes and actions that are far apart in time and context becomes challenged with many competing explanations about why observed behavior and beliefs are misaligned. The data presented in this article include stated attitudes elicited immediately proceeding individuals' data-sharing behavior.

There are also related latent variables that are often difficult to account for empirically. Evaluating attitudes in a rigorous way introduces the challenge of accounting for other endogenous factors. Personality characteristics is one source mentioned by Adjerid et al. (2019). Among demographic characteristics, age can have explanatory power over privacy preferences (Goldfarb and Tucker, 2012). Beyond surveys on privacy attitudes, demographics like gender strongly interact with self-reported beliefs (Exley and Kessler, 2019). In my study, I collect data on these individual characteristics, including demographics, social media usage, and Five Factor personality scores.

The results show that information security, data control rights, and sharing attitudes can be aligned with individuals' data-sharing activities. Users' who exhibit weak behavior displayed relatively weaker attitudes and vice versa. Moreover, individuals who seek privacy due to second parties' data exploitation activities are less likely to state an expectation that these activities occur when businesses collect their personal data. These results suggest that (1) concern about data privacy, (2) demand for control rights and restrictions on the free movement of data, and (3) awareness about data exploitation are more aligned with a person's privacy behavior than previously believed.

4.2 Data

This data is used in the experimental study of the prior chapter, which analyzes the privacy behavior of individuals in response to various information provisions about the ability of data recipients' to trade data in secondary markets. Despite its experimental focus on valuation and awareness of data markets, this data is also useful for examining stated attitudes and concerns. The first portion of the survey (prior to any data-sharing and privacy items), individuals participated in a survey to receive their Five Factor personality scores.[3] After the experimental portion, the respondent was asked to provide demographic information—including age, gender, marital status, employment status—and social media usage. Stated attitudes elicited at the end of this study are primed with salient information about the consequences of data-sharing. This data set, therefore, provides a rich set of individual characteristics, personality traits, and behaviors that can be be linked to their attitudes about data privacy and control.

[3] All responses to this self-assessment were required for survey completion.

Participants reported stated attitudes and expectations about data privacy, ownership, and usage intentions by businesses (see Table 4.1). Stated attitudes about data privacy, control, usage, and expectations are elicited in four general categories of items. For each item, a seven-point likert scale (ranging from "Strongly Agree" to "Strongly Disagree") is used to capture individuals' impressions. The first category elicits respondents' concerns about their data privacy related to identity, access, usage, and sensitivity of information. A second category captures attitudes about control over one's data, including exclusion and retraction rights. A third category elicits attitudes about the usage rights of third parties over their personal data. Finally, a fourth category elicits general awareness and expectations about a second party's ability to internally and externally exploit data. Respondents were given the option to not respond to any of these questions.

While Cho and Larose (1999) and Day (1975) argue that online surveys may have severe sample selection and low response rates due to its privacy-invasive questions, my survey combats these challenges by marketing it as a research study on economic decisions, delaying demographic and privacy questions until the end of the study, and incentivizing survey completion and participation. In fact, approximately 84 percent of registered participants completed the required portions of the survey; among those, over 99 percent were willing to continue and complete the exit survey on privacy attitudes in return for a chance to win monetary rewards.

4.3 Data Analysis

4.3.1 Stated Attitudes

Responses to the Likert-type items on data privacy, ownership, and secondary data markets is presented in Table 4.1. Overall attitudes were skewed towards extreme agreement about data privacy concerns, including when their data is identifiable and especially when their data is "highly sensitive". The results of this survey is suggestive of normative attitudes about the desire for privacy or the need to be concerned about privacy.

Over half of those surveyed stated some disagreement with the internal or external exploitation third parties' data exploitation abilities. However, when "others can use my information for their own purposes" is re-packaged in a statement about a third party's right to "use my data for their own purposes" results are less skewed and even tend towards agreement. This suggests that reported attitudes are sensitive to the priming or context in which they are elicited (e.g., in the context of when one is concerned about privacy, or in a context when one considers the rights of data recipients).

In terms of the general expectations of a firm's (i.e., a second party's) data usage, self-reported awareness varied according to internal versus external data exploitation. There was general agreement that when data is shared with a business, individuals expected the business to use the data *internally* for commercial purposes. However, *external* exploitation (i.e., data sale to other businesses) was skewed towards more disagreeable attitudes.

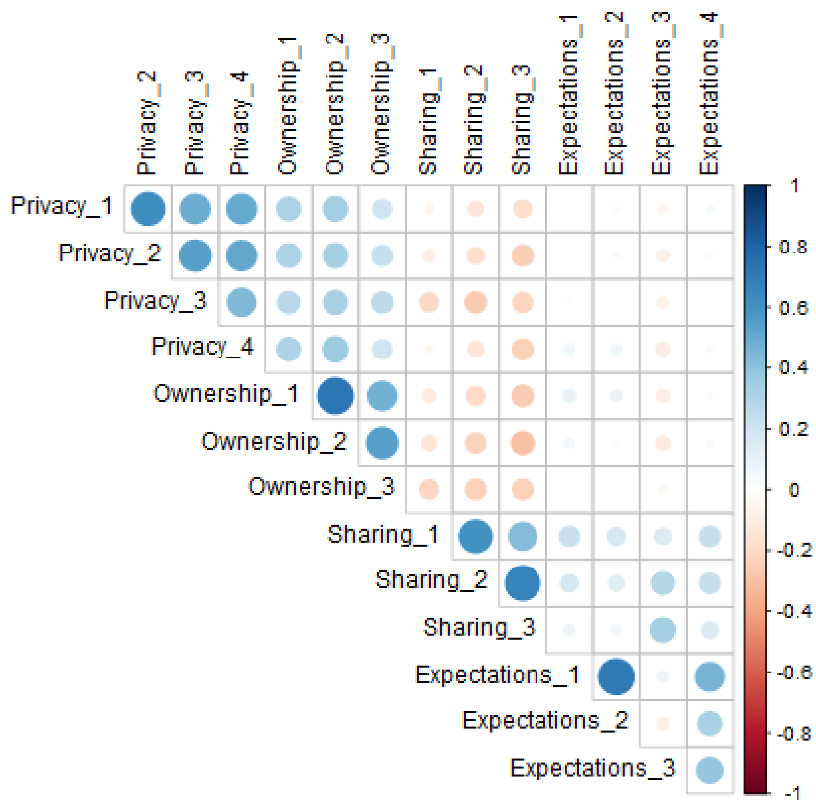
This survey also contributes observational data to Taylor (2004), which de-

Table 4.1: 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 …									
_1	… I can be personally identified."	55.1	25.7	11.4	2.9	2.4	1.3	1.2	0.0
_2	… 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
_3	… 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
_4	… 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 …									
_1	… choose who I share my data with."	65.5	24.9	5.5	2.2	0.8	0.5	0.5	0.2
_2	… exclude others from accessing my data."	60.3	26.3	7.6	2.8	1.8	0.7	0.5	0.1
_3	… 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 …									
_1	… use my data for their own purposes."	9.3	24.4	21.9	9.8	9.8	10.8	13.7	0.3
_2	… use my data to make money."	4.9	10.6	15.2	9.9	14.2	16.9	28.1	0.1
_3	… 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 …									
_1	… 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
_2	… 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
_3	… sell my data to other businesses."	11.4	14.3	15.1	8.1	11.5	17.5	22.1	0.0
_4	… 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, 2=Agree, 3=Somewhat agree, 4=Neither agree nor disagree, 5=Somewhat disagree, 6=Disagree, 7=Strongly disagree. Statement blocks and statement lines presented to participants in a random order.

Figure 4.1: Correlation Plot of Stated Attitude Items



Note: Items labeled in order according to items in Table 4.1.

velops a model that rational consumers keep information private to prevent the future price discrimination (a consequence of firms trading data with each other in secondary markets). However, whether consumers are attentive towards this possibility is an open question. The results of my data show that respondents are generally aware of these activities, when price discrimination activities by a second party is explicitly described (i.e., "When I share my data with a business, I expect that they will use my data to understand whether I would be willing to pay more for their products or services." in Table 4.1). This suggests that consumers are neither fully blind to nor fully against firms' ability to use data for price discrimination.

Figure 4.1 demonstrates the pair-wise correlations for within-individual attitudinal responses for each of the items in Table 4.1. Negative correlations indicate indirect relationships between items. Positive correlations indicate co-movement in the same direction of the Likert scale. For example, the agreement for items related to privacy concerns are associated with agreeing attitudes for ownership and control over data. However, privacy-concerning attitudes are negatively associated with agreement towards exchange and usage of personal data by third parties. Little association between awareness about second party data exploitation activities if found with the other three categories. This suggests that awareness of how businesses use one's personal data seems to have little correlation with one's normative attitudes around data privacy, control, and secondary markets.

4.3.2 Attitudes and Behavior Comparison

In order to document and understand whether actual privacy behavior have relationships with these stated attitudes, I turn to regression analysis. In addition, I can examine whether Five Factor scores (which is a tool that elicits robust and stable psychometric factors) have explanatory power in resolving disparities in individuals' stated and revealed preferences. The econometric approach examines the relationship between actual (ex-ante) privacy decisions and stated attitudes.

The dependent variable is an indicator for whether the individual reported an attitude more extreme than the average tendency of those in the sample. For example, the category of items under "I am concerned about my data privacy when ...”; an above average concern would indicate whether the individual’s response was more extreme than the average tendency for *each* item in the category. While Likert-type responses are notoriously difficult to interpret and standardize for measurement, the measure used in this empirical strategy can capture some indicator for the strength (or lack of uncertainty) in beliefs, relative to others in the sample. Finally, the average is used as the central tendency—rather than the median—since many items’ median and modal impressions from respondents were also the boundary of the scale and, therefore, without responses more extreme than the median.

The main independent variables of interest are the actual privacy responses (i.e., willingness-to-share) psychometric data with a data recipient(s) in four different conditions with varying information provisions about (1) the number of data recipients and (2) the ability for recipients to exploit data in secondary markets. A binary variable *Private Behavior* indicates whether the individual chose to not disclose their data (in return for monetary benefits) in at least three out of four privacy decisions. Second, an indicator for *Exposure Averse* labels individuals who either did not share data in a condition with thirty data recipients or demanded more benefits (i.e., higher prices) compared to one recipient. Third, an indicator for *Exploitation Averse* similarly labels individuals who either did not share data in a condition with signals about the recipient’s ability to exploit data in secondary markets or had a higher reservation prices when compared to conditions without this information provision. All respondents were required to complete their decisions for the four data conditions in order to proceed to the exit survey and earn their participation fees. Therefore, no missing data are included. Throughout these analyses, one specification includes all three of these privacy behavior indicators. However, *Private Behavior* a strong, multi-collinearity with *Exposure Averse* and *Exploitation Averse*. This is due to the construction of the latter responses variables, which include those who chose private behavior in at least two conditions.

The secondary independent variables of interest are the psychometric attributes of individuals, elicited through the Five Factor self-assessment at the start of the study. Each score ranges between 10 and 50, and low scores on a Factor indicate low value in that personality trait. For example, a low score on *Extraversion* indicates high introversion. Similar to the actual experimental protion, all respondents were required to complete the self-assessment in order to earn their participation fees.

Finally, the last set of explanatory variables are included to explore variation in stated attitudes that are associated with demographics like age, gender, martial status, employment status, and social media usage. While the demographic questions of the exit survey were voluntary, these indicators only represent opt-in disclosure of one type. For example, *Female* indicates whether the individual selected female as their gender identity, instead of "male", "non-binary", "prefer not to say," or a non-response. *Age* is an integer provided by the respondent regarding age. An indicator is included for those who select their marital status as *Single (never married)*. Individuals who are employed as opposed to "unemployed," "looking for work," "retired," or "student" is indicated by *Employed (full or part-time)*. Finally, individuals who indicated that they access Facebook at least once a week under a non-anonymous account are indicated by *Regularly Uses Facebook*.

In the first set of attitudes in Table 4.2, above average privacy concern for personal data that is identifiable, easily accessed by others, can be used by others

Table 4.2: Attitudes, Behavior, and Individual Traits

Variables	**Privacy** Above Avg. Concern		**Ownership** Above Avg. Attitude		**Sharing** Above Avg. Attitude		**Expectations** Above Avg. Aware	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
(Intercept)	-4.566**	-5.041***	-3.731*	-4.029**	-1.136	-1.672	-3.469*	-3.029†
	(1.456)	(1.442)	(1.520)	(1.517)	(1.629)	(1.616)	(1.684)	(1.677)
Private Behavior	0.879***		0.680***		0.920***		-0.389*	
	(0.146)		(0.166)		(0.155)		(0.163)	
Exposure Averse		0.364**		0.293*		0.484**		0.103
		(0.135)		(0.140)		(0.151)		(0.153)
Exploitation Averse		0.422**		0.388**		0.319†		-0.405*
		(0.146)		(0.142)		(0.166)		(0.164)
Extraversion	-0.002	-0.003	-0.008	-0.008	0.007	0.006	-0.012	-0.013
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Agreeableness	0.027*	0.023†	0.037**	0.035**	0.012	0.009	0.028*	0.030*
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Conscientiousness	0.030**	0.030**	0.007	0.006	0.026*	0.026*	0.020†	0.021†
	(0.010)	(0.010)	(0.010)	(0.010)	(0.012)	(0.012)	(0.011)	(0.011)
Intellect	-0.004	-0.001	0.045***	0.047***	-0.002	0.001	0.012	0.010
	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)
Emotional Stability	-0.027**	-0.028**	-0.001	-0.002	-0.004	-0.005	0.002	0.003
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Female	-0.221	-0.192	-0.038	-0.020	0.039	0.074	0.210	0.199
	(0.151)	(0.149)	(0.154)	(0.154)	(0.171)	(0.169)	(0.166)	(0.165)
Age	0.809*	0.927*	0.315	0.387	-0.502	-0.358	0.792†	0.683
	(0.386)	(0.381)	(0.409)	(0.408)	(0.436)	(0.431)	(0.457)	(0.455)
Single (never married)	0.331	0.306	0.066	0.034	-0.263	-0.289	0.197	0.197
	(0.255)	(0.253)	(0.262)	(0.262)	(0.275)	(0.274)	(0.281)	(0.281)
Employed (full or part-time)	-0.068	-0.071	0.321†	0.313†	-0.052	-0.054	-0.144	-0.149
	(0.164)	(0.163)	(0.174)	(0.173)	(0.185)	(0.184)	(0.184)	(0.184)
Regularly Uses Facebook	-0.285†	-0.319†	-0.291†	-0.314†	-0.083	-0.113	-0.189	-0.148
	(0.173)	(0.172)	(0.171)	(0.171)	(0.192)	(0.191)	(0.187)	(0.187)
Sample Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1139	1139	1137	1137	1136	1136	1139	1139
Adj. Pseudo-R^2	0.030	0.020	0.029	0.028	0.017	0.005	-0.003	-0.004

Note: Each dependent variable indicates respondents who self-reported beliefs more extreme (towards privacy-protecting behavior) than the central tendency of those in the sample for *each* item in the category. Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

for their own intentions, and contain highly sensitive information is correlated with unwillingness to share data in nearly all real privacy decisions related to their identifiable, psychometric data. The odds of stating above average concern is 2.42 ($p < 0.001$) times greater for those who rejected data-sharing in the majority of decisions. Those who behaved privately in the face of greater exposure—to more data recipients—and with salient information about recipient's data exploitation abilities also exhibit greater odds of above average stated concerns. Personality traits that have a positive association with above average concern include Agreeableness and Conscientiousness, and negatively associated with Emotional Stability. A one point increase in each trait's score (on a scale of 10 to 50) is associated with an approximately 3 percent increase in the odds of above average concern, and a 3 percent decrease in the odds of above average concern. Moreover, those who do not regularly use Facebook have a 1.32 times greater odds of stated concern.

I find that age positively and significantly associates with stated privacy concerns information (see Table 4.2), consistent with Goldfarb and Tucker (2012) which finds that age gaps in privacy concerns seem to be widening (with older adults more concerned). A positive association—although not significant—is also found between age and respondents' attitudes about data control retraction rights. However, a negative association (although not significant) is found for attitudes regarding third-party usage.[4] These results are also aligned with early survey work by Hoofnagle et al. (2010), which finds that younger and older American adults are largely in harmony regarding the maintenance of privacy, especially related to norms and policy suggestions. Results of this survey

[4]Consistent with research by Hoofnagle et al. (2010), which finds that younger and older American adults are largely in harmony regarding the maintenance of privacy, especially related to norms and regulations.

emphasize the multi-dimensional aspects of privacy concerns. Age may be correlated with certain presentations of privacy concerning attitudes, but have less or insufficient association with other attitudes on privacy-protecting regimes.

Interestingly, as shown in Models (4a) and (4b) in Table 4.2, older adults are associated with stronger expectations and awareness about a second party's ability to internally and externally exploit data. While prior work has found age gaps in privacy knowledge—where younger adults answer online privacy questions correctly (Hoofnagle et al., 2010)—this survey demonstrates a reverse age gap in expectations about a data recipient's economic activities with data.

4.3.3 Paradoxical Attitudes and Behavior

This section includes analyses on a search for paradoxical behavior among outliers of the sample. In the tables that follow, demographic variables are tested for association with attitudes in which stated preferences and slanted from revealed behavior. It is important to note that these are endogenous predictors of paradoxical behavior. For example, paradoxical behavior predicted by a personality trait can be due to that personality trait's relationship with Likert-type responses in general, and not necessarily related to paradoxical privacy preferences. Nonetheless, this section provides a useful documentation for potential confounding factors when interpreting the distance between stated attitudes and actual behavior.

Across all regressions, there are two categories of paradoxical behavior, either (1) strong stated attitudes and weak privacy behavior or (2) weak stated attitudes and strong privacy behavior. Within each category, there are three estimated forms of revealed privacy preferences. First, there a those behave privately strongly (weakly) and did not share personally, identifiable psychometric data in a majority (minority) of their data-sharing decisions. Second, strong exposure responses indicate those who did not share data or responded more privately (i.e., high reservation prices) when exposed to thirty data recipients—alternatively, under weak behavior, those that did share data and did not respond more privately. Third, strong exploitation responses indicate those who did not share data or responded more privately when treated with information about their data recipient's exploitation abilities—similarly, under weak behavior, those that did share data and did not respond more privately.

In Table 4.3, *Emotional Stability* demonstrates a positive association with individuals who have strong stated concerns about data privacy and weak responses towards exposure.[5] Strong stated concerns includes attitudes about data privacy concerns more extreme than the average in the sample. These weak responses indicate data-sharing psychometric data with thirty data recipients in return for monetary benefits.

In Table 4.4, *Conscientiousness* negatively associates with weak stated data ownership and control attitudes and strong exploitation response.[6] *Intellect* negatively associates with strong stated attitudes and weak exploitation responses.[7] In terms of demographic characteristics, *Female* positively associates with strong stated attitudes about data ownership and control, with weak revealed privacy behavior (i.e., sharing data in a majority of data-decisions).[8]

---

[5]For a one point increase in Five Factor score, $\hat{\beta} = 0.021$, *S.E.* $= 0.009$.
[6]For a one point increase in Five Factor score, $\hat{\beta} = -0.026$, *S.E.* $= 0.012$.
[7]For a one point increase in Five Factor score, $\hat{\beta} = -0.042$, *S.E.* $= 0.015$.
[8]For *Female*, $\hat{\beta} = 0.806$, *S.E.* $= 0.380$.

Table 4.3: Paradoxical Privacy Concerns and Privacy Behavior

Variables	**Strong Stated Concern and Weak . . .**			**Weak Stated Concern and Strong . . .**		
	Private Behavior	Expos. Response	Exploit. Response	Private Behavior	Expos. Response	Exploit. Response
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	-3.420	-1.963	0.771	-1.326	-0.823	-2.521
	(2.083)	(1.576)	(1.414)	(1.652)	(1.872)	(2.361)
Extraversion	0.010	0.008	0.003	0.003	0.001	0.002
	(0.013)	(0.009)	(0.008)	(0.009)	(0.010)	(0.013)
Agreeableness	-0.010	0.002	0.007	0.018	0.004	0.020
	(0.018)	(0.013)	(0.012)	(0.013)	(0.015)	(0.019)
Conscientiousness	-0.009	-0.018	-0.014	0.017	0.007	-0.014
	(0.015)	(0.011)	(0.010)	(0.011)	(0.013)	(0.016)
Intellect	0.010	0.008	0.002	-0.019	-0.010	-0.019
	(0.019)	(0.014)	(0.012)	(0.014)	(0.015)	(0.020)
Emotional Stability	0.016	0.021^{*}	$0.016^{\dagger}$	$-0.016^{\dagger}$	-0.017	-0.020
	(0.013)	(0.009)	(0.008)	(0.010)	(0.011)	(0.014)
Female	0.313	0.122	0.185	$-0.293^{\dagger}$	-0.111	-0.298
	(0.233)	(0.165)	(0.147)	(0.164)	(0.188)	(0.234)
Age	0.255	-0.019	-0.515	-0.100	-0.196	0.371
	(0.551)	(0.419)	(0.379)	(0.443)	(0.504)	(0.633)
Single (never married)	-0.055	0.167	-0.175	0.435	0.326	0.318
	(0.360)	(0.277)	(0.244)	(0.304)	(0.345)	(0.451)
Employed (full or part-time)	0.151	0.177	-0.004	-0.142	-0.176	$-0.577^{\dagger}$
	(0.236)	(0.175)	(0.160)	(0.186)	(0.213)	(0.298)
Regularly Uses Facebook	-0.099	0.019	0.218	-0.045	0.140	-0.233
	(0.256)	(0.185)	(0.163)	(0.190)	(0.207)	(0.289)
Sample Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1139	1139	1139	1139	1139	1139
Adj. Pseudo-R^2	-0.023	-0.011	-0.010	-0.008	-0.016	-0.017

Note: Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

In Table 4.5, *Intellect* negatively associates with weak stated attitudes about the free movement of data in secondary markets and strong exploitation responses.[9] *Age* strongly and positively associates with paradoxical privacy preferences: older individuals have stronger attitudes (against) the sharing and use of personal data in the secondary markets, but demonstrate weak privacy be-

[9] For a one point increase in Five Factor score, $\hat{\beta} = -0.049$, *S.E.* $= 0.024$.

Table 4.4: Paradoxical Ownership Attitudes and Privacy Behavior

Variables	**Strong Stated Attitude and Weak ...**			**Weak Stated Attitude and Strong ...**		
	Private Behavior	Expos. Response	Exploit. Response	Private Behavior	Expos. Response	Exploit. Response
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	-3.532	-0.799	2.100	0.322	1.251	0.303
	(2.815)	(2.054)	(1.784)	(1.377)	(1.480)	(1.765)
Extraversion	0.019	0.012	0.011	-0.003	-0.005	0.000
	(0.018)	(0.012)	(0.010)	(0.008)	(0.008)	(0.010)
Agreeableness	-0.031	-0.008	-0.014	0.020†	0.008	0.002
	(0.024)	(0.017)	(0.014)	(0.011)	(0.012)	(0.014)
Conscientiousness	0.018	0.005	-0.006	-0.002	-0.002	-0.026*
	(0.022)	(0.015)	(0.012)	(0.010)	(0.010)	(0.012)
Intellect	-0.024	-0.030†	-0.042**	0.019	0.019	0.010
	(0.026)	(0.017)	(0.015)	(0.012)	(0.012)	(0.015)
Emotional Stability	-0.005	-0.002	-0.003	0.004	-0.003	-0.002
	(0.018)	(0.012)	(0.010)	(0.008)	(0.009)	(0.010)
Female	0.806*	0.070	0.047	-0.058	0.025	-0.165
	(0.380)	(0.218)	(0.182)	(0.142)	(0.151)	(0.176)
Age	0.511	0.014	-0.366	-0.627†	-0.751†	-0.292
	(0.745)	(0.552)	(0.483)	(0.369)	(0.399)	(0.477)
Single (never married)	-0.326	-0.191	-0.148	-0.023	-0.335	-0.129
	(0.476)	(0.352)	(0.305)	(0.239)	(0.251)	(0.301)
Employed (full or part-time)	0.083	-0.274	-0.281	0.186	-0.026	-0.030
	(0.334)	(0.246)	(0.207)	(0.155)	(0.166)	(0.198)
Regularly Uses Facebook	-0.506	0.049	0.221	-0.069	0.116	-0.213
	(0.391)	(0.239)	(0.197)	(0.161)	(0.168)	(0.211)
Sample Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1139	1139	1139	1139	1139	1139
Adj. Pseudo-R^2	-0.019	-0.022	-0.003	-0.008	-0.011	-0.015

Note: Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

havior and responses to data exposure.[10]

In Table 4.6, *Conscientiousness* negatively associates with weak stated expectations about data exploitation by second parties and strong exploitation responses.[11] *Employment* negatively associates with weak stated expectations and

[10]For a one year increase in age, $\hat{\beta} = 1.618$, *S.E.* $= 0.462$ in weak private behavior and $\hat{\beta} = 1.365$, *S.E.* $= 0.390$ in weak exposure response.

[11]For a one point increase in Five Factor score, $\hat{\beta} = -0.022$, *S.E.* $= 0.011$.

Table 4.5: Paradoxical Data-Sharing Attitudes and Privacy Behavior

Variables	**Strong Stated Attitude and Weak . . .**			**Weak Stated Attitude and Strong . . .**		
	Private Behavior	Expos. Response	Exploit. Response	Private Behavior	Expos. Response	Exploit. Response
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	-7.160***	-5.892***	-2.051	0.266	-1.362	-0.020
	(1.783)	(1.479)	(1.367)	(1.993)	(2.210)	(3.110)
Extraversion	-0.009	0.002	-0.005	0.002	0.008	0.000
	(0.011)	(0.009)	(0.008)	(0.011)	(0.013)	(0.016)
Agreeableness	0.014	0.017	0.021†	0.018	0.002	0.028
	(0.016)	(0.012)	(0.011)	(0.016)	(0.018)	(0.024)
Conscientiousness	0.007	-0.004	-0.003	0.022	0.016	-0.010
	(0.014)	(0.010)	(0.010)	(0.014)	(0.015)	(0.020)
Intellect	0.000	-0.006	-0.005	-0.030†	-0.034†	-0.049*
	(0.016)	(0.013)	(0.012)	(0.016)	(0.018)	(0.024)
Emotional Stability	-0.009	0.001	0.000	-0.002	-0.010	-0.004
	(0.012)	(0.009)	(0.008)	(0.011)	(0.013)	(0.017)
Female	0.084	-0.070	0.049	-0.120	0.050	-0.077
	(0.201)	(0.153)	(0.142)	(0.198)	(0.233)	(0.296)
Age	1.618***	1.365***	0.409	-0.684	-0.045	-0.557
	(0.462)	(0.390)	(0.365)	(0.540)	(0.595)	(0.847)
Single (never married)	0.130	0.485†	0.286	-0.241	-0.082	0.279
	(0.310)	(0.261)	(0.239)	(0.333)	(0.384)	(0.559)
Employed (full or part-time)	0.187	0.188	0.215	-0.095	-0.141	0.265
	(0.204)	(0.164)	(0.155)	(0.223)	(0.257)	(0.314)
Regularly Uses Facebook	-0.363	-0.349†	0.047	0.049	-0.061	-0.196
	(0.235)	(0.180)	(0.160)	(0.223)	(0.260)	(0.353)
Sample Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1139	1139	1139	1139	1139	1139
Adj. Pseudo-R^2	0.009	0.008	-0.006	-0.015	-0.022	-0.033

Note: Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

strong exposure responses.[12] Finally, *Facebook* usage positively associates with weak stated expectations about a firm's exploitation intentions, and strong exposure responses.[13]

Based on the analyses from this section, readers should conclude that the event of "paradoxical" behavior is not limited to one particular category of in-

[12]For *Employed (full or part-time)*, $\hat{\beta} = -0.326$, *S.E.* $= 0.161$.
[13]For Facebook users, $\hat{\beta} = 0.346$, *S.E.* $= 0.162$.

Table 4.6: Paradoxical Data Exploitation Awareness and Privacy Behavior

Variables	**Strong Stated Expect. and Weak ...**			**Weak Stated Expect. and Strong ...**		
	Private Behavior	Expos. Response	Exploit. Response	Private Behavior	Expos. Response	Exploit. Response
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	-2.375	0.404	1.521	2.309†	3.276*	1.435
	(2.518)	(2.230)	(1.902)	(1.373)	(1.417)	(1.577)
Extraversion	0.012	0.003	0.011	-0.006	-0.011	-0.004
	(0.016)	(0.013)	(0.011)	(0.008)	(0.008)	(0.009)
Agreeableness	-0.034	-0.018	-0.010	0.004	-0.011	-0.013
	(0.021)	(0.018)	(0.015)	(0.011)	(0.011)	(0.012)
Conscientiousness	0.003	-0.006	-0.018	0.001	0.000	-0.022*
	(0.019)	(0.016)	(0.013)	(0.010)	(0.010)	(0.011)
Intellect	0.007	-0.009	0.001	-0.010	-0.009	-0.003
	(0.023)	(0.019)	(0.016)	(0.012)	(0.012)	(0.013)
Emotional Stability	-0.023	-0.002	-0.007	0.004	0.001	0.001
	(0.017)	(0.013)	(0.011)	(0.008)	(0.008)	(0.009)
Female	0.093	-0.261	-0.109	-0.016	0.041	-0.062
	(0.287)	(0.225)	(0.189)	(0.144)	(0.145)	(0.159)
Age	0.475	-0.240	-0.581	-0.564	-0.729†	-0.277
	(0.674)	(0.601)	(0.516)	(0.366)	(0.380)	(0.426)
Single (never married)	-0.430	-0.297	-0.184	0.018	-0.251	0.006
	(0.425)	(0.370)	(0.319)	(0.238)	(0.244)	(0.276)
Employed (full or part-time)	-0.099	-0.025	0.042	-0.257†	-0.326*	-0.291
	(0.312)	(0.256)	(0.211)	(0.155)	(0.161)	(0.182)
Regularly Uses Facebook	-0.339	0.171	0.077	0.133	0.346*	-0.052
	(0.336)	(0.254)	(0.211)	(0.163)	(0.162)	(0.183)
Sample Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1139	1139	1139	1139	1139	1139
Adj.Pseudo-R^2	-0.032	-0.025	-0.015	-0.008	-0.002	-0.009

Note: Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

dividual. Depending on the type of normative belief being elicited—whether related to privacy concern, data control rights, or the free movement of data in secondary markets—various individual characteristics correlate with a divergence in stated versus actual privacy decision-making over data. Of course, the context of this study—as with all real-world privacy choices— revealed behavior is limited to one type of data and context.

On the other hand, a search fo paradoxical outliers can have little predictive power in understanding the universe of users and their privacy choices. Privacy decisions are varied in context, benefits, and consequences. It is likely futile to create a privacy-loving or privacy-paradox index for each individual, given the ever growing categories of individual data, data-sharing terms, data recipients, and policy norms that all influence individuals' stated and revealed preferences for data-sharing in online markets. Perhaps, instead of a uniform theory of privacy, viewing privacy through a price theory lens centered around information goods—i.e., where privacy is synonymous with the trading choices people make over their self-generated information commodities—is a far more fruitful model for privacy preferences.

4.4 Conclusion

This paper examines how privacy concerns compare to actual privacy behavior. Contrary to finding overwhelming evidence of a privacy paradox among individuals, I observe that stated and revealed privacy preferences can align. Privacy attitudes more or less rationalize decisions individuals make related to their data, which—in the setting of this data—immediately precede their stated, normative beliefs about privacy regulation. In addition, other individual level characteristics—including demographics and personality traits—are observed and suggest a pattern of association with various concerns about privacy.

Some of these characteristics, including age, intellect, emotional stability, and conscientiousness, correlate with the minority that exhibit paradoxical stated and actual behavior. Notably, outliers of paradoxical behavior are not limited to strong, stated preferences and weak, revealed behavior; the opposite phenomenon is also possible. However, much of the reasons underlying the trends observed in my data lie unexplained. The data is also limited by the survey setup and priming of participants to actual data-sharing behavior and information provisions about secondary data market activities. Despite these limitations, I provide unique insights into behavior-based evidence of the existence and possible correlates to beliefs about data privacy regulation. Future research should explore the mechanisms and conditions under which stated preferences vary and deviate from behavior.

4.5 Acknowledgments

Thank you to my committee members, Aija Leiponen, David Just, and Vicki Bogan, for supporting this effort and all the valuable feedback. Special thanks to LMU Munich's Institute for Strategy, Technology and Organization (ISTO) and Institute for Behavioral Economics and Consumer Choice (IBECC) round table for providing helpful comments.

APPENDIX A

APPENDIX OF CHAPTER 2

Table A.1: Average Valuations by Group

Variable	**Licensor**		**Licensee**
	Creator	Owner	
Price (\$0 to \$15)	9.74	9.00	3.63
	(0.72)	(0.67)	(0.31)
	95% CI [8.29, 8.29]	95% CI [7.64, 7.64]	95% CI [3.01, 3.01]
Confidence (1 to 10 scale)	8.55	7.29	6.59
	(0.34)	(0.53)	(0.35)
	95% CI [7.86, 7.86]	95% CI [6.22, 6.22]	95% CI [5.89, 5.89]
Quality (1 to 10 scale)	8.48	7.19	6.36
	(0.33)	(0.52)	(0.33)
	95% CI [7.81, 7.81]	95% CI [6.15, 6.15]	95% CI [5.71, 5.71]
Number of individuals	42	42	83

Note: Standard errors of means in parentheses.

Table A.2 shows the interaction of solutions' probabilistic success (graded by the researcher and not by participants) of mean differences in evaluations among creators, owners, and buyers. For main results for all solutions, see Table 2.1 in Section 2.5.2. "High" grades were assigned to ideas that accurately solved the logic problem or included one minor error in its prescription. "Low" grades were assigned to ideas that had more than one minor error in its prescription.

Table A.2: Algorithm Valuation Differences By Grade

Variable	**Mean of Differences**					
	Low Grade Solutions			**High Grade Solutions**		
	C−L1 (1a)	C−O (2a)	O−L2 (3a)	C−L1 (1b)	C−O (2b)	O−L2 (3b)
Price ($0 to $15)	6.340***	0.640	5.480***	6.367***	0.267	4.867*
	(1.142)	(1.360)	(0.882)	(1.036)	(1.626)	(1.747)
Confidence (1 to 10 scale)	2.880***	1.680*	0.080	1.733†	0.733	0.133
	(0.738)	(0.791)	(0.850)	(0.842)	(0.746)	(0.533)
Quality (1 to 10 scale)	2.720***	1.480*	0.320	2.533*	1.133†	-0.067
	(0.629)	(0.656)	(0.818)	(0.925)	(0.584)	(0.714)
Number of ideas	25	25	25	15	15	15

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Standard errors of means in parentheses. "C" denotes creator-licensor, "O" denotes owner-licensor, "L1" denotes licensee paired with C, and "L2" denotes licensee paired with O.

Table A.3: Survey Comprehension Score Differences

Variable	**Mean of Differences**		
	Creator−Licensee 1 (1)	Creator−Owner (2)	Owner−Licensee 2 (3)
Comprehension Score	0.000	0.025	0.025
	(0.051)	(0.056)	(0.084)
Number of ideas	40	40	40

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Standard errors of means in parentheses.

Table A.4: Licensee Valuation and Comprehension Score Differences

Variable	**Mean of Differences**
	Licensee 1−Licensee 2
Price	-0.600
	(0.613)
Confidence	-1.025†
	(0.532)
Quality	-1.125†
	(0.558)
Comprehension	0.050
	(0.080)
Number of ideas	40

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Standard errors of means in parentheses.

Table A.5: Average Licensees' and Licensors' Valuations by Type

Algorithm Quality Type	**Avg. Royalty Price** ($) **with Meta-Info on Algorithm's Success**	
	Licensees' Max. WTP	Licensors' Min. WTA
High (>80% Chance of Success)	3.32	3.13
	(0.12)	(0.10)
	95% CI [3.09, 3.54]	95% CI [2.92, 3.34]
Medium (50% Chance of Success)	2.62	2.49
	(0.12)	(0.10)
	95% CI [2.38, 2.86]	95% CI = [2.30, 2.68]
Low (<20% Chance of Success)	1.28	1.69
	(0.13)	(0.11)
	95% CI [1.02, 1.55]	95% CI [1.47, 1.91]
Number of observations	135	137

Note: Standard errors in parentheses.

Table A.6: Difference in Average Licensees' and Licensors' Valuations by Type

Algorithm Quality Type	**Max. WTP and Min. WTA** ($) **with Meta-Info on Algorithm's Success**	
	Difference	Ratio
High (>80% Chance of Success)	0.185	0.944
	(0.156)	
	95% CI [-0.122, 0.491]	
Medium (50% Chance of Success)	0.129	0.951
	(0.155)	
	95% CI [-0.176, 0.435]	
Low (<20% Chance of Success)	-0.41*	1.320
	(0.17)	
	95% CI [-0.754, -0.066]	

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Standard errors in parentheses.

Table A.7: Average Licensor Valuations by Type

Algorithm Quality Type	**Avg. Royalty Price** ($) **Min. WTA**	
With Meta-Info on Success:	Yes	No
High (>80% Chance of Success)	2.89	2.49
	(0.12)	(0.12)
	95% CI [2.66, 3.12]	95% CI [2.25, 2.72]
Medium (50% Chance of Success)	2.32	2.52
	(0.10)	(0.11)
	95% CI [2.13, 2.52]	95% CI [2.29, 2.74]
Low (<20% Chance of Success)	1.66	2.03
	(0.11)	(0.14)
	95% CI [1.44, 1.88]	95% CI [1.76, 2.31]
Number of individuals	141	142

Note: Standard errors in parentheses.

Table A.8: Difference in Average Licensors' Royalties for Algorithms With and Without Success Information

Algorithm Quality Type	**Valuation With vs. Without** ($) **Meta-Info on Algorithm's Success**	
	Difference	Ratio
High (>80% Chance of Success)	0.403* (0.166) 95% CI [0.076, 0.729]	0.861
Medium (50% Chance of Success)	-0.194 (0.151) 95% CI [-0.491, 0.103]	1.083
Low (<20% Chance of Success)	-0.37* (0.18) 95% CI [-0.723, -0.024]	1.225

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Standard errors in parentheses.

APPENDIX B

APPENDIX OF CHAPTER 3

Figure B.1: 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.

Table B.1: 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 B.2: Randomization Group and Condition Orders

Sample	**Group**	**Conditions (in order)**			
		$\mathbf{t = 1}$	$\mathbf{t = 2}$	$\mathbf{t = 3}$	$\mathbf{t = 4}$
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 B.2: 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.

Figure B.3: Reservation Prices for Data-Sharing by Condition and Sample

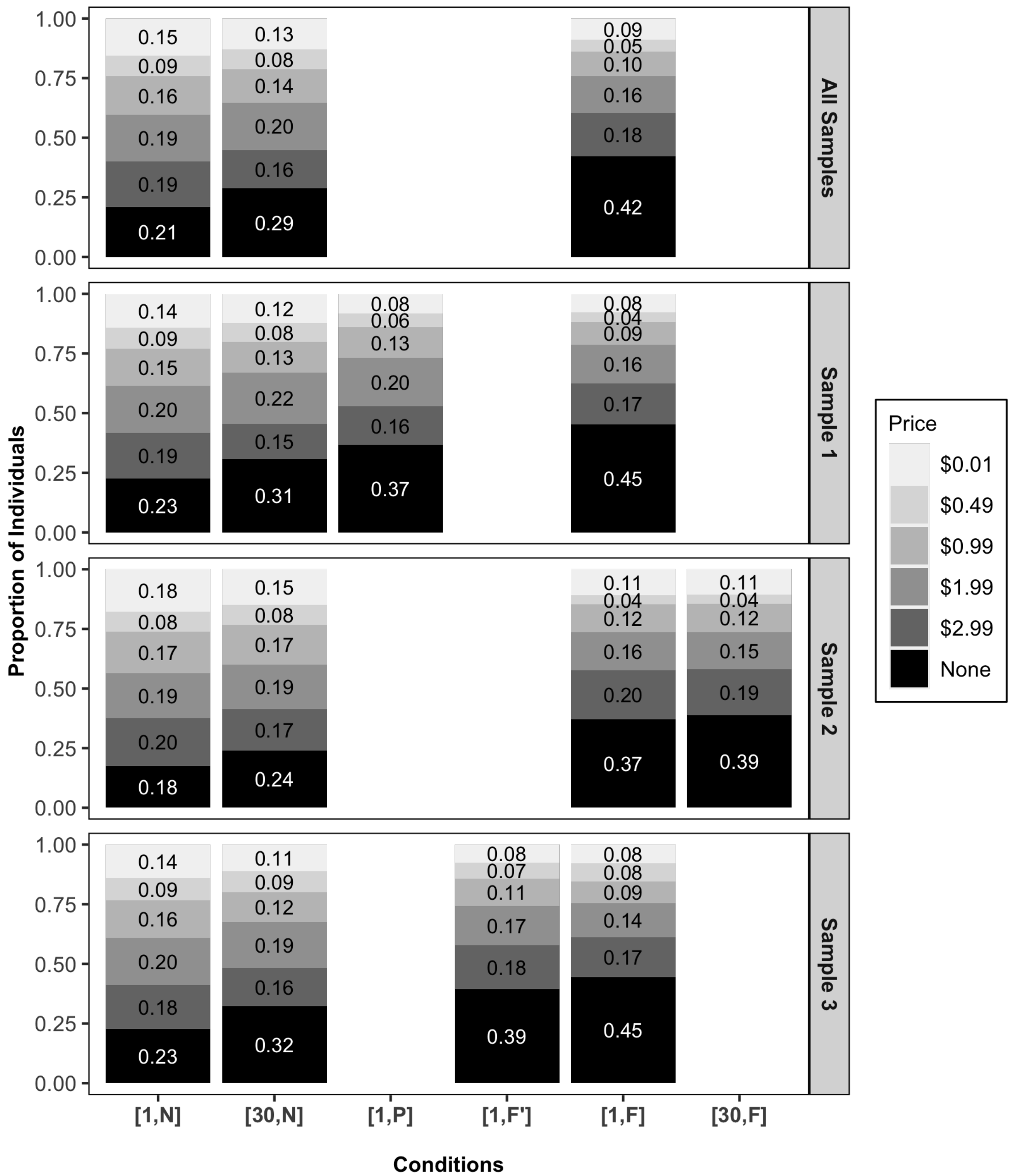


Table B.3: Data Market Participation and Price Results (Pooled, Full Table)

Variables	**Logit**			**OLS**		
Dependent:	*Participation*			*Price* (\$) $\mid$ *Participation*$= 1$		
Comparison: $[1, F]$	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	0.242*	0.473***	1.009**	3.566***	3.442***	3.095***
	(0.105)	(0.119)	(0.383)	(0.086)	(0.096)	(0.276)
$[1, N]$	1.023***	1.016***	1.037***	-0.267***	-0.284***	-0.287***
	(0.064)	(0.064)	(0.065)	(0.048)	(0.048)	(0.049)
$[30, N]$	0.590***	0.583***	0.596***	-0.247***	-0.264***	-0.267***
	(0.059)	(0.059)	(0.060)	(0.048)	(0.048)	(0.048)
NoInfoFirst		-0.370***	-0.396***		0.197*	0.213*
		(0.108)	(0.110)		(0.088)	(0.088)
Female			-0.102			-0.072
			(0.130)			(0.098)
Single			0.463**			0.002
			(0.170)			(0.146)
Employed			-0.214			-0.164
			(0.134)			(0.107)
Facebooker			0.148			0.019
			(0.145)			(0.107)
Extravert			0.102			0.122
			(0.112)			(0.086)
Agreeable			-0.316			0.334*
			(0.237)			(0.169)
Conscientious			-0.594***			0.141
			(0.161)			(0.106)
EmoStable			0.155			-0.123
			(0.114)			(0.087)
Intellectual			-0.176			-0.004
			(0.167)			(0.121)
Sample 2	0.329*	0.331*	0.311*	-0.077	-0.073	-0.060
	(0.132)	(0.132)	(0.133)	(0.102)	(0.101)	(0.101)
Sample 3	-0.018	-0.018	-0.009	-0.038	-0.040	-0.051
	(0.133)	(0.134)	(0.137)	(0.106)	(0.105)	(0.105)
Decision $t = 2$	0.087	0.085	0.085	0.051	0.048	0.042
	(0.081)	(0.079)	(0.081)	(0.054)	(0.053)	(0.053)
Decision $t = 3$	-0.103	-0.181*	-0.184*	-0.071	0.016	0.022
	(0.085)	(0.081)	(0.082)	(0.063)	(0.058)	(0.058)
Decision $t = 4$	-0.121	-0.200*	-0.207*	-0.101	-0.013	-0.013
	(0.084)	(0.079)	(0.081)	(0.066)	(0.061)	(0.061)
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)
$Pr(> \chi^2)$	0.001	0.001	0.000	0.025	0.124	0.040

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

Table B.4: Participation and Price Results (Sample 1, Full Table)

Variables	**Logit**			**OLS**		
Dependent:	*Participation*			*Price* (\$)$\vert$*Participation*$= 1$		
Comparison: $[1, P]$	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	0.636***	0.760***	0.193	3.441***	3.336***	3.435***
	(0.117)	(0.151)	(0.583)	(0.088)	(0.113)	(0.435)
$[1, N]$	0.686***	0.677***	0.699***	-0.210**	-0.216**	-0.223**
	(0.097)	(0.096)	(0.099)	(0.072)	(0.072)	(0.071)
$[30, N]$	0.262**	0.253**	0.262**	-0.208**	-0.217**	-0.223**
	(0.094)	(0.093)	(0.097)	(0.073)	(0.073)	(0.072)
$[1, F]$	-0.363***	-0.366***	-0.379***	0.088	0.094	0.082
	(0.076)	(0.077)	(0.080)	(0.064)	(0.064)	(0.065)
NoInfoFirst		-0.261	-0.320†		0.204	0.209
		(0.180)	(0.183)		(0.141)	(0.137)
Female			-0.006			-0.149
			(0.223)			(0.154)
Single			0.702*			-0.082
			(0.279)			(0.238)
Employed			-0.006			-0.227
			(0.229)			(0.175)
Facebooker			0.523*			-0.330†
			(0.228)			(0.169)
Extravert			-0.200			0.320*
			(0.186)			(0.138)
Agreeable			-0.120			-0.008
			(0.371)			(0.257)
Conscientious			-0.463†			0.184
			(0.272)			(0.190)
EmoStable			0.370†			-0.150
			(0.191)			(0.142)
Intellectual			0.372			-0.011
			(0.274)			(0.230)
Decision $t = 2$	0.079	0.077	0.082	0.094†	0.093	0.083
	(0.075)	(0.074)	(0.077)	(0.057)	(0.057)	(0.057)
Decision $t = 3$	-0.158†	-0.131	-0.135	0.071	0.089	0.084
	(0.095)	(0.095)	(0.098)	(0.078)	(0.077)	(0.076)
Decision $t = 4$	-0.252*	-0.227*	-0.237*	-0.017	0.009	0.001
	(0.101)	(0.100)	(0.103)	(0.080)	(0.080)	(0.079)
Individual clusters	413	413	413	334	334	334
Observations	1652	1652	1652	1094	1094	1094
ANOVA: Wald Test	(1a),(2a)	(2a),(3a)	(1a),(3a)	(1b),(2b)	(2b),(3b)	(1b),(3b)
$Pr(> \chi^2)$	0.148	0.019	0.016	0.148	0.042	0.031

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

Table B.5: Participation and Price Results (Sample 2, Full Table)

Variables	**Logit**			**OLS**		
Dependent:	*Participation*			*Price* (\$)$\vert$*Participation*$= 1$		
Comparison: $[30, F]$	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	0.496***	0.790***	2.624***	3.411***	3.306***	2.368***
	(0.114)	(0.151)	(0.604)	(0.095)	(0.118)	(0.387)
$[1, N]$	1.092***	1.086***	1.152***	-0.294***	-0.310***	-0.323***
	(0.119)	(0.118)	(0.124)	(0.078)	(0.078)	(0.077)
$[30, N]$	0.698***	0.690***	0.737***	-0.277***	-0.289***	-0.299***
	(0.099)	(0.098)	(0.103)	(0.073)	(0.073)	(0.073)
$[1, F]$	0.070	0.073	0.079	0.027	0.028	0.030
	(0.058)	(0.060)	(0.065)	(0.049)	(0.049)	(0.048)
NoInfoFirst		-0.658***	-0.671***		0.219	0.219
		(0.190)	(0.195)		(0.146)	(0.145)
Female			0.023			-0.094
			(0.226)			(0.167)
Single			0.383			0.207
			(0.292)			(0.243)
Employed			-0.225			-0.093
			(0.232)			(0.181)
Facebooker			-0.069			0.277
			(0.255)			(0.179)
Extravert			0.338†			-0.061
			(0.193)			(0.146)
Agreeable			-0.988*			0.782**
			(0.406)			(0.249)
Conscientious			-1.125***			0.173
			(0.276)			(0.160)
EmoStable			-0.039			-0.024
			(0.194)			(0.144)
Intellectual			-0.474†			-0.005
			(0.278)			(0.182)
Decision $t = 2$	0.215*	0.209*	0.220*	0.080.	0.082.	0.083†
	(0.085)	(0.082)	(0.087)	(0.048)	(0.048)	(0.048)
Decision $t = 3$	-0.164	-0.054	-0.066	0.119	0.148†	0.158*
	(0.107)	(0.106)	(0.113)	(0.075)	(0.076)	(0.076)
Decision $t = 4$	-0.200†	-0.092	-0.105	0.045	0.075	0.086
	(0.108)	(0.107)	(0.114)	(0.081)	(0.082)	(0.082)
Individual clusters	420	420	420	359	359	359
Observations	1680	1680	1680	1186	1186	1186
ANOVA: Wald Test	(1a),(2a)	(2a),(3a)	(1a),(3a)	(1b),(2b)	(2b),(3b)	(1b),(3b)
$Pr(>\chi^2)$	0.001	0.000	0.000	0.135	0.094	0.059

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

Table B.6: Participation and Price Results (Sample 3, Full Table)

Variables	**Logit**			**OLS**		
Dependent:	*Participation*			*Price*(\$) $\vert$ *Participation*$= 1$		
Comparison: $[1, F']$	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
(Intercept)	0.499***	0.573***	1.124	3.593***	3.467***	3.377***
	(0.122)	(0.156)	(0.766)	(0.100)	(0.126)	(0.489)
$[1, N]$	0.792***	0.786***	0.801***	-0.268**	-0.283***	-0.283***
	(0.103)	(0.102)	(0.104)	(0.083)	(0.082)	(0.082)
$[30, N]$	0.308**	0.302**	0.308**	-0.221*	-0.234**	-0.233**
	(0.100)	(0.100)	(0.102)	(0.087)	(0.087)	(0.085)
$[1, F]$	-0.209**	-0.210**	-0.214**	-0.081	-0.083	-0.085
	(0.073)	(0.073)	(0.075)	(0.066)	(0.066)	(0.066)
NoInfoFirst		-0.158	-0.163		0.252	0.248
		(0.192)	(0.196)		(0.154)	(0.153)
Female			-0.413†			-0.095
			(0.229)			(0.164)
Single			0.168			0.067
			(0.326)			(0.251)
Employed			-0.363			-0.179
			(0.224)			(0.182)
Facebooker			-0.098			0.328
			(0.272)			(0.210)
Extravert			0.151			0.173
			(0.198)			(0.151)
Agreeable			0.003			0.113
			(0.498)			(0.330)
Conscientious			-0.242			0.043
			(0.284)			(0.196)
EmoStable			0.128			-0.288†
			(0.207)			(0.158)
Intellectual			-0.249			-0.001
			(0.282)			(0.210)
Decision $t = 2$	0.133	0.132	0.138	-0.054	-0.057	-0.064
	(0.096)	(0.095)	(0.097)	(0.064)	(0.064)	(0.064)
Decision $t = 3$	-0.165	-0.149	-0.151	-0.115	-0.088	-0.089
	(0.105)	(0.104)	(0.106)	(0.086)	(0.087)	(0.085)
Decision $t = 4$	-0.238*	-0.223*	-0.225*	-0.155†	-0.128	-0.132
	(0.107)	(0.106)	(0.108)	(0.089)	(0.089)	(0.087)
Individual clusters	355	355	355	286	286	286
Observations	1420	1420	1420	926	926	926
ANOVA: Wald Test	(1a),(2a)	(2a),(3a)	(1a),(3a)	(1b),(2b)	(2b),(3b)	(1b),(3b)
$Pr(> \chi^2)$	0.412	0.437	0.483	0.100	0.521	0.285

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

APPENDIX C

APPENDIX OF CHAPTER 4

Table C.1: Survey Dates, Registration, and Completion

Sample 1	**Wave 1**	**Wave 2**	**Wave 3**	**Wave 4**	**Total**
Dates (in 2019)	3/21-24	3/28-30	4/11-14	4/18-20	
Start day (time)	Thu (8:30 AM)	Thu (10:09 AM)	Thu (11:31 AM)	Thu (8:55 AM)	
End day (time)	Sat (11:59 PM)	Sat (11:59 PM)	Sat (11:59 PM)	Sat (11:59 PM)	
# Available	200	100	200	100	600
# Registered	200	26	156	91	473
# Completed	187	18	128	80	413
Sample 2	**Wave 5**	**Wave 6**	**Wave 7**	**Wave 8**	**Total**
Dates (in 2019)	10/31-11/02	11/07-09	11/14-16	11/21-23	
Start day (time)	Thu (8:00 AM)	Thu (8:00 AM)	Thu (8:00 AM)	Thu (8:00 AM)	
End day (time)	Sat (11:59 PM)	Sat (11:59 PM)	Sat (11:59 PM)	Sat (11:59 PM)	
# Available	200	100	200	100	600
# Registered	200	100	147	69	516
# Completed	170	87	113	50	420
Sample 3	**Wave 9**	**Wave 10**	**Wave 11**	**Wave 12**	**Total**
Dates (in 2020)	5/21-23	5/28-30	6/4-6	6/11-13	
Start day (time)	Thu (8:00 AM)	Thu (8:00 AM)	Thu (8:00 AM)	Thu (8:00 AM)	
End day (time)	Sat (11:59 PM)	Sat (11:59 PM)	Sat (11:59 PM)	Sat (11:59 PM)	
# Available	200	100	200	100	600
# Registered	173	100	70	90	433
# Completed	151	90	55	58	355

Note: Registration opened between Sundays and Wednesdays prior to survey start time. Email advertisements were sent out by the Lab on Mondays. Incomplete surveys did *not* guarantee participation fee or partial earnings. Participants needed to provide an email address upon survey *completion* in order to receive electronic compensation.

Table C.2: Summary of Self-Assessment Scores

Factor	**Mean**	**SD**
Extraversion	30.6	8.22
Agreeableness	39.9	5.83
Conscientiousness	36.5	6.53
Emotional Stability	29.7	7.74
Intellect	36.8	5.65

Note: Scores ranged from 10 to 50.

Table C.3: Summary of Self-Reported Age

Mean	Min.	Max.	SD
23.75	18.00	91.00	6.95

Note: Four percent of total sample did not respond.

Table C.4: Summary of Self-Reported Demographics

Description	Proportion	99% CI	NAs
Female	0.727	[0.693, 0.759]	0.010
Male	0.238	[0.208, 0.271]	0.010
Bachelor's degree	0.321	[0.287, 0.357]	0.019
Enrolled in college	0.509	[0.472, 0.546]	0.026
Not enrolled in a program	0.223	[0.194, 0.256]	0.026
Single (never married)	0.874	[0.847, 0.897]	0.024
Student (not employed or searching for work)	0.711	[0.676, 0.744]	0.023
Employed (full or part-time)	0.236	[0.205, 0.269]	0.023

Note: Proportions out of total sample. Confidence intervals are computed using the Agresti-Coull method.

Table C.5: Summary of Self-Reported Social Media Usage

Description	Proportion	99% CI	NAs
Facebook user	0.790	[0.757, 0.818]	0.108
Posts on Facebook weekly	0.190	[0.163, 0.221]	0.215
LinkedIn user	0.508	[0.470, 0.545]	0.108
Connects on LinkedIn weekly	0.391	[0.356, 0.428]	0.492
Twitter user	0.243	[0.213, 0.277]	0.108
Tweets weekly	0.123	[0.100, 0.150]	0.758

Note: "Users" are described to respondents as individuals who access the online platform at least once a week and using a non-anonymous account where the individual can be reasonably identified. Questions on usage frequency and type only appeared for users of the platform. Proportions out of total sample. Confidence intervals are computed using the Agresti-Coull method.

BIBLIOGRAPHY

D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Too much data: Prices and inefficiencies in data markets. Working Paper No. 26296, National Bureau of Economic Research, September 2019.

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

A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security and Privacy*, 3:26–33, 2005a.

A. Acquisti and J. Grossklags. Uncertainty, ambiguity and privacy. In *Fourth Workshop on the Economics of Information Security (WEIS05)*, pages 2–3, 2005b.

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

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

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

I. Adjerid, A. Acquisti, L. Brandimarte, and G. Loewenstein. Sleights of privacy: Framing, disclosures, and the limits of transparency. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, volume 9, pages 1–11. ACM, 2013.

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

C. M. Angst and R. Agarwal. The Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *Management Information Systems Quarterly*, 33(2):339–370, 2009.

K. Arrow. Economic welfare and the allocation of resources for inventions. In R. R. Nelson, editor, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pages 609–626. Princeton: Princeton University Press, 1962.

S. Athey, C. Catalini, and C. Tucker. The digital privacy paradox: Small money, small costs, small talk. Working Paper No. 23488, National Bureau of Economic Research, June 2017.

C. K. Bagga, N. Bendle, and J. Cotte. Object valuation and non-ownership possession: how renting and borrowing impact willingness-to-pay. *Journal of the Academy of Marketing Science*, 47(1):97–117, 2019.

F. Bardhi and G. M. Eckhardt. Liquid consumption. *Journal of Consumer Research*, 44(3): 582–597, 2017.

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

R. W. Belk. Possessions and the extended self. *Journal of Consumer Research*, 15(2):139–168, 1988.

J. Bessen and R. M. Hunt. An empirical look at software patents. *Journal of Economics and Management Strategy*, 16(1), 2007.

E. Blankenship. How do i make money as an algorithm developer? *Algorithmia.com*, 2019. accessed: June 15, 2021.

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

J. R. Buckman, J. C. Bockstedt, and M. J. Hashim. Relative privacy valuations under varying disclosure characteristics. *Information Systems Research*, 30(2):375–388, 2019.

H. Cho and R. Larose. Privacy issues in internet surveys. *Social Science Computer Review*, 17(4):421–434, 1999.

A. Chowdhry. How algorithmia built the largest marketplace for algorithms in the world. *Forbes.com*, 2018. accessed: June 15, 2021.

R. H. Coase. The nature of the firm. *Economica*, 4(16):386–405, 1937.

R. H. Coase. *The Problem of Social Cost*, pages 87–137. Palgrave Macmillan UK, 1960.

M. J. Culnan. "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, 1993.

M. J. Culnan and P. K. Armstrong. Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science*, 10(1):104–115, 1999.

G. S. Day. The threats to marketing research. *Journal of Marketing Research*, 12(4):462–467, 1975.

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

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

C. L. Exley and J. B. Kessler. The gender gap in self-promotion. Working Paper 26345, National Bureau of Economic Research, October 2019.

L. Furby. Sharing: Decisions and moral judgments about letting others use one's possessions. *Psychological Reports*, 43(2):595–609, 1978.

L. Furby. Understanding the psychology of possession and ownership: A personal memoir and an appraisal of our progress. *Journal of Social Behavior and Personality*, 6:457–63, 1991.

L. R. Goldberg. The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4(1):26–42, 1992.

L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96, 2006.

A. Goldfarb and C. Tucker. Shifts in privacy concerns. *American Economic Review: Papers & Proceedings*, 3:349–353, 2012.

A. Goldfarb and C. Tucker. Digital economics. *Journal of Economic Literature*, 57:3–43, 2019.

E. Graham-Harrison and C. Cadwalladr. Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach. *The Guardian*, Mar 2018.

M. Haase and M. Kleinaltenkamp. Property rights design and market process: Implications for market theory, marketing theory, and s-d logic. *Journal of Macromarketing*, 31(2):148–159, 2011.

K. Hill. The secretive company that might end privacy as we know it. *The New York Times*, Jan 2020.

C. J. Hoofnagle, J. King, S. Li, and J. Turow. How different are young adults from older adults when it comes to information privacy attitudes and policies? 2010.

S. Hooshangi and G. Loewenstein. The impact of idea generation and potential appropriation on entrepreneurship: An experimental study. *Management Science*, 64(1): 64–82, 2018.

K.-L. Hui, H. H. Teo, and S.-Y. T. Lee. The Value of Privacy Assurance: An Exploratory Field Experiment. *Management Information Systems Quarterly*, 31(1):19–33, 2007.

S. Isaacs. *Social Development in Young Children*. Routledge & Kegan Paul LImited, 1933.

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

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

C. Korting and S. G. Otto. Choice Uncertainty and the Endowment Effect. 2019 Annual Meeting, July 21-23, Atlanta, Georgia 290841, Agricultural and Applied Economics Association, June 2019.

P. Koutroumpis, A. Leiponen, and L. D. W. Thomas. The nature of data. Innovation and Entrepreneurship Working Papers, 2019.

P. Koutroumpis, A. Leiponen, and L. D. W. Thomas. Markets for data. *Industrial and Corporate Change*, pages 1–16, 2020.

R. Laufer and M. Wolfe. Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, 33:22–42, 1977.

K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

J. A. List and J. F. Shogren. Calibration of Willingness-to-Accept. *Journal of Environmental Economics and Management*, 43(2):219–233, 2002.

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

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

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

C. K. Morewedge, L. L. Shu, D. T. Gilbert, and T. D. Wilson. Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology*, 45(4):947–951, 2009.

C. K. Morewedge, A. Monga, R. W. Palmatier, S. B. Shu, and D. A. Small. Evolution of Consumption : A Psychological Ownership Framework. *Journal of Marketing*, 85 (1):196–218, 2021.

P. Norberg, D. Horne, and D. Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1):100–126, 2007.

M. I. Norton, D. Mochon, and D. Ariely. The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, 22(3):453–460, 2012.

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

C. Peukert, S. Bechtold, M. Batikas, and T. Kretschmer. European Privacy Law and Global Markets for Data. *SSRN Electronic Journal*, 2020.

J. L. Pierce and I. Jussila. *Psychological Ownership and the Organizational Context: Theory,* *Research Evidence, and Application*. Edward Elgar Publishing, Northampton, MA, US, 2011.

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

S. Preibusch. How to Explore Consumers' Privacy Choices with Behavioral Economics. In S. Zeadally and M. Badra, editors, *Privacy in a Digital, Networked World*, chapter 14, pages 313–341. Springer International Publishing, Springer, Cham, Switzerland, 2015.

J. Reb and T. Connolly. Possession, feelings of ownership and the endowment effect. *Judgment and Decision Making*, 2(2):107–114, 2007.

S. Schudy and V. Utikal. You must not know about me—on the willingness to share personal data. *Journal of Economic Behavior & Organization*, 141:1–13, 2017.

S. B. Shu and J. Peck. Psychological ownership and affective reaction: Emotional attachment process variables and the endowment effect. *Journal of Consumer Psychology*, 21(4):439–452, 2011.

P. Sondergaard. Big data fades to the algorithm economy. *Forbes.com*, 2015. accessed: June 15, 2021.

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

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

C. R. Taylor. Consumer privacy and the market for customer information. *The RAND Journal of Economics*, 35(4):631, 2004.

L. D. W. Thomas and A. Leiponen. Big data commercialization. *IEEE Engineering Management Review*, 44(2):74–90, 2016.

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

Y.-F. Tuan. The significance of the artifact. *Geographical Review*, pages 462–472, 1980.

Y.-F. Tuan. *Dominance and affection*. Yale University Press, 1984.

H. Varian. Economic aspects of personal privacy. 1996.

A. F. Westin. *Privacy and Freedom*. Atheneum, 1st edition, 1967.

A. F. Westin. *1996 Equifax/Harris Consumer Privacy Survey.* Equifax, Atlanta, GA, 1996.

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