

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

Privacy, Disclosure, and Social Exchange Theory

### Permalink

<https://escholarship.org/uc/item/5hw5w5c1>

### Author

King, Jennifer

### Publication Date

2018

Peer reviewed|Thesis/dissertation

**Privacy, Disclosure, and Social Exchange Theory**

by

Jennifer King

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Information Management and Systems

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Deirdre K. Mulligan, Chair  
Professor Coye Cheshire  
Professor David Wagner  
Professor Steven Weber

Spring 2018

**Privacy, Disclosure, and Social Exchange Theory**

Copyright 2018  
by  
Jennifer King

## Abstract

Privacy, Disclosure, and Social Exchange Theory

by

Jennifer King

Doctor of Philosophy in Information Management and Systems

University of California, Berkeley

Professor Deirdre K. Mulligan, Chair

Maintaining the privacy of one's personal information—one's choice of when to disclose it and to whom, how one maintains control over it, and the risks of disclosure—is one of the most important social issues of the internet era. For the past decade, privacy researchers have focused on several domains, including: documenting public opinion about privacy attitudes and expectations; understanding how user interfaces affect disclosure; and focusing on understanding interpersonal privacy dynamics within social media settings, to name a few. All of this work shares the goal of furthering our collective understanding of how people think about information privacy in online settings, what they expect when disclosing their personal information, and why they make the disclosure choices they make. A common element missing from the extant privacy research is an accounting of social structures. More specifically, as researchers consider the various factors that affect personal disclosure, they often do not consider the *relationship* between the discloser and the recipient, and how aspects of that relationship may directly or indirectly affect one's decision to disclose. A specific form of relationship I examine here is that between individuals and the companies to whom they disclose their personal information.

This dissertation explores how the structure of relationships between individuals and companies influences individuals' decisions to disclose personal information. I accomplish this through a mixed-methods approach; first, I conducted twenty exploratory qualitative interviews with ten users of the 23andMe genetic testing service and ten women who used mobile apps to track their pregnancies. I interviewed all twenty participants about their experiences using online search engines. I then conducted three online survey experiments, using a hypothetical wearable fitness device that collects personal information as the premise of the study. The experiments tested a set of hypotheses and further explored themes that emerged from the qualitative research.

These studies examine the ways in which the *relationship* between individuals and the companies they disclose to, and in particular the *distribution of power within the relationship*, affects the individuals' decisions to disclose. I use social exchange theory (SET) as the theoretical framework for this inquiry because the transfer of personal information in exchange for a service is an exchange between social actors. Thus, SET provides an empirically tested scaffolding for exploring

key features of these relationships and their impact on the normative aspects of exchange that affect disclosure choices, specifically: individuals' perceptions of trust, fairness, power, and privacy.

This dissertation forges new ground in the analysis of information privacy and personal disclosure. Namely, the results of my mixed-method studies demonstrate the utility of the relational analytic approach for identifying social structural factors that affect personal disclosure. Further, it demonstrates the influence of *power* on personal disclosure—the extent to which individuals can control the terms under which personal information is exchanged, the options available to them to obtain similar resources elsewhere, how fair the exchange is, and the extent to which individuals benefit from it. This approach yields a different set of insights into the dynamic of personal disclosure and information privacy. It reveals the impact of power differentials on personal disclosure, demonstrating that imbalances in power between individuals and companies can affect individual decisions to disclose.

*To my stars and moon: Deneb, Seren, and Sashi*

*Thank you for all the love in the universe, and for your support through this crazy journey.*

# Contents

<b>Contents</b>	<b>ii</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Introduction and Literature Review</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Problem Space . . . . .	1
1.3 Research Motivations . . . . .	2
1.3.1 Research Focus . . . . .	3
1.3.2 Dissertation Structure . . . . .	3
1.4 Related Literature . . . . .	3
1.4.1 Information Privacy . . . . .	4
1.4.2 Privacy as Control . . . . .	5
1.4.2.1 Research Based on Privacy As Control . . . . .	6
1.4.2.2 Privacy As A Rational Calculus . . . . .	7
1.4.3 The Privacy Paradox . . . . .	8
1.4.4 Privacy and Behavioral Economics . . . . .	10
1.4.4.1 Economics of Privacy . . . . .	11
1.4.5 Privacy as Contextual Integrity . . . . .	12
1.5 Social Exchange Theory . . . . .	14
1.5.1 Forms of Exchange . . . . .	15
1.5.1.1 Direct Negotiated Binding Exchange . . . . .	16
1.5.1.2 Direct Reciprocal Exchange . . . . .	16
1.5.1.3 Generalized Exchange . . . . .	16
1.5.2 The Object of Exchange . . . . .	16
1.5.3 Assurances and SET . . . . .	18
1.5.4 Power and SET . . . . .	18
<b>2 Motivations, Research Questions, and Study Hypotheses</b>	<b>21</b>
2.1 Interleaving SET and Privacy . . . . .	21

2.2	Motivating the Studies . . . . .	24
2.2.1	Qualitative Study . . . . .	24
2.2.1.1	Interview Contexts . . . . .	24
2.2.1.2	Qualitative Research Questions . . . . .	25
2.2.2	Experimental Survey . . . . .	25
2.2.2.1	The Paradox of Control . . . . .	26
2.2.2.2	Experimental Research Questions . . . . .	27
2.3	Study One: Hypotheses . . . . .	27
2.3.1	Trust . . . . .	28
2.3.2	Power . . . . .	28
2.3.3	Fairness . . . . .	29
2.3.4	Privacy . . . . .	29
2.4	Studies 2A and 2B . . . . .	30
2.4.1	Hypotheses . . . . .	30
<b>3</b>	<b>Qualitative Methods</b>	<b>32</b>
3.1	Study Overview . . . . .	32
3.1.1	Study Context . . . . .	32
3.1.1.1	Pregnancy Tracking Apps . . . . .	33
3.1.1.2	Direct-to-Consumer Genetic Testing (DTCGT) . . . . .	34
3.1.1.3	23andMe Company Background . . . . .	34
3.1.1.4	Genetic Testing Process . . . . .	35
3.2	Participant Recruitment . . . . .	36
3.2.1	Search Interviews . . . . .	37
3.2.2	Interview Instrument . . . . .	37
3.2.3	Contextual Questions . . . . .	38
3.2.4	SET Thematic Questions . . . . .	40
3.2.4.1	Relationships . . . . .	41
3.2.4.2	Fairness and Benefits . . . . .	41
3.2.4.3	Reciprocity and Negotiation . . . . .	42
3.2.4.4	Assurances . . . . .	42
3.2.4.5	Personal Disclosure . . . . .	44
3.2.4.6	Risks and Trade-Offs . . . . .	44
3.2.5	Interviews . . . . .	45
3.2.6	Data Analysis . . . . .	46
3.2.7	Generalizability and Limitations . . . . .	46
<b>4</b>	<b>Qualitative Interview Findings</b>	<b>48</b>
4.1	Overview . . . . .	48
4.2	Interviews: Setting the Context . . . . .	49
4.2.1	Pregnancy Tracking Apps . . . . .	50
4.2.1.1	Selecting Pregnancy Applications . . . . .	50



4.2.1.2	Company and App Impressions . . . . .	50
4.2.1.3	Personal Disclosure and Revealing the Pregnancy . . . . .	51
4.2.2	Genetic Testing . . . . .	52
4.2.2.1	Motivations for Genetic Testing . . . . .	52
4.2.2.2	Participant Impressions of 23andMe . . . . .	53
4.2.2.3	Complicating Social Norms . . . . .	54
4.2.3	Search Engines . . . . .	55
4.2.3.1	Evolving Search Norms . . . . .	56
4.2.4	Summary . . . . .	56
4.3	Social Exchange Theory and Relationships . . . . .	57
4.3.1	The Nature of the Relationship . . . . .	57
4.3.2	Pregnancy Tracking . . . . .	58
4.3.3	Genetic Testing . . . . .	60
4.3.4	Search Queries . . . . .	61
4.3.5	Summary . . . . .	62
4.3.6	Benefits and Fairness . . . . .	64
4.3.6.1	Pregnancy Tracking . . . . .	65
4.3.6.2	Genetic Testing . . . . .	66
4.3.6.3	Search Queries . . . . .	68
4.3.6.4	Summary . . . . .	71
4.3.7	Assurances and Trust . . . . .	72
4.3.7.1	Pregnancy Tracking . . . . .	72
4.3.7.2	Institutional Assurances . . . . .	74
4.3.7.3	Genetic Testing . . . . .	75
4.3.7.4	Institutional Assurances . . . . .	76
4.3.7.5	Search Queries . . . . .	76
4.3.7.6	Institutional Assurances . . . . .	77
4.3.7.7	Summary . . . . .	78
4.4	Privacy: Personal Disclosure, Risk, and Trade-Offs . . . . .	79
4.4.1	Personal Disclosure . . . . .	79
4.4.1.1	Pregnancy Tracking . . . . .	79
4.4.1.2	Pregnancy and Advertising . . . . .	80
4.4.1.3	Genetic Testing . . . . .	81
4.4.1.4	Search Queries . . . . .	83
4.4.1.5	Summary . . . . .	86
4.5	Risks and Trade-Offs . . . . .	87
4.5.1	Pregnancy Tracking . . . . .	87
4.5.1.1	Risks . . . . .	88
4.5.1.2	Specialized Sharing . . . . .	88
4.5.1.3	Trade-Offs . . . . .	89
4.5.2	Genetic Testing . . . . .	90
4.5.2.1	Risks . . . . .	90

4.5.2.2	Specialized Sharing . . . . .	91
4.5.2.3	Trade-Offs . . . . .	91
4.5.3	Search Queries . . . . .	92
4.5.3.1	Risks . . . . .	92
4.5.3.2	Nothing To Hide . . . . .	93
4.5.3.3	Trade-Offs . . . . .	95
4.5.3.4	Summary . . . . .	95
4.6	Summary . . . . .	97
<b>5</b>	<b>Experimental Methods</b>	<b>98</b>
5.1	Introduction . . . . .	98
5.2	Study One: Negotiated Relationships . . . . .	99
5.3	Study Two: Assurances . . . . .	101
5.4	Experimental Design . . . . .	102
5.5	Study Assumptions . . . . .	103
5.5.1	Data Collected . . . . .	104
5.6	Independent Variables . . . . .	105
5.6.1	Study One . . . . .	105
5.6.2	Study Two . . . . .	105
5.7	Dependent Variables . . . . .	106
5.7.1	Trust . . . . .	106
5.7.2	Power . . . . .	107
5.7.3	Fairness . . . . .	107
5.7.4	Privacy . . . . .	108
5.8	Mediating Factors . . . . .	108
5.8.1	Demographic Factors . . . . .	108
5.8.2	Trust and Caution Scale . . . . .	109
5.8.3	Information Technology Knowledge Scale . . . . .	109
5.8.4	Privacy Scale . . . . .	109
5.9	Vignette-Specific Control Factors . . . . .	110
5.9.0.1	Scales . . . . .	111
5.10	Survey Instrument Development . . . . .	111
5.10.1	Recruitment and Study Procedures . . . . .	112
5.10.2	Study Two Procedures . . . . .	115
5.11	Analytic Strategy . . . . .	116
5.11.1	Study One . . . . .	116
5.11.1.1	Stage 1: Bivariate Hypothesis Testing . . . . .	116
5.11.1.2	Regression Analysis . . . . .	117
5.11.2	Study Two . . . . .	117
<b>6</b>	<b>Experimental Survey Findings</b>	<b>118</b>
6.1	Study One Findings . . . . .	118

6.1.1	Hypothesis 1: Trust . . . . .	118
6.1.2	Hypothesis 2: Power . . . . .	118
6.1.3	Hypothesis 3: Fairness . . . . .	119
6.1.4	Hypothesis 4: Privacy . . . . .	119
6.2	Stage Two: Regression Analysis . . . . .	120
6.2.1	Trust Measure . . . . .	120
6.2.1.1	Model One . . . . .	120
6.2.1.2	Model Two . . . . .	122
6.2.1.3	Model Three . . . . .	122
6.2.1.4	Model Four . . . . .	122
6.2.1.5	Discussion . . . . .	122
6.2.2	Power Measure . . . . .	123
6.2.2.1	Model One . . . . .	124
6.2.2.2	Model Two . . . . .	124
6.2.2.3	Model Three . . . . .	124
6.2.2.4	Model Four . . . . .	124
6.2.2.5	Discussion . . . . .	124
6.2.3	Fairness Measure . . . . .	125
6.2.3.1	Model One . . . . .	127
6.2.3.2	Model Two . . . . .	127
6.2.3.3	Model Three . . . . .	127
6.2.3.4	Model Four . . . . .	127
6.2.3.5	Discussion . . . . .	127
6.2.4	Privacy Measures . . . . .	128
6.2.4.1	Model One . . . . .	128
6.2.4.2	Model Two . . . . .	128
6.2.4.3	Model Three . . . . .	128
6.2.4.4	Model Four . . . . .	131
6.2.4.5	Discussion . . . . .	131
6.2.5	Study One Discussion . . . . .	132
6.3	Study 2A Findings . . . . .	132
6.3.1	Summary . . . . .	133
6.4	Study 2B Findings . . . . .	134
6.4.1	Summary . . . . .	134
6.5	Study Two A+B Discussion . . . . .	135
6.5.1	Assurances in Detail . . . . .	135
6.5.1.1	Too Little Risk, or Risk Mismeasured . . . . .	137
6.5.1.2	Entrenched Cynicism Among Participants . . . . .	139
6.5.1.3	The Paradox of Disclosure . . . . .	139
6.5.1.4	Design Flaws . . . . .	140
6.6	Summary . . . . .	140

<b>7</b>	<b>Summary, Synthesis, and Conclusion</b>	<b>141</b>
7.1	Research Questions . . . . .	141
7.1.1	Qualitative Research Questions . . . . .	141
7.1.2	Experimental Survey Research Questions . . . . .	142
7.2	Summary of Qualitative Findings and Emergent Themes . . . . .	142
7.2.1	The Nature of the Relationship . . . . .	142
7.2.2	Benefits and Fairness . . . . .	142
7.2.3	Assurances and Trust . . . . .	142
7.2.4	Personal Disclosure . . . . .	143
7.2.5	Risks and Trade-Offs . . . . .	143
7.3	Emergent Themes . . . . .	143
7.3.1	The Importance of Power in Disclosure Relationships . . . . .	144
7.3.2	The Length of the Relationship . . . . .	145
7.3.3	Mediating Effects of Indirect Generalized Exchange . . . . .	145
7.4	Summary of Survey Experiment Findings and Emergent Themes . . . . .	146
7.4.1	Study One . . . . .	146
7.4.2	Privacy Attitudes . . . . .	147
7.4.3	Implications for Design . . . . .	147
7.4.4	Negative Perceptions of Company Relationships . . . . .	148
7.4.5	Privacy Paradox . . . . .	150
7.5	Synthesis and Discussion . . . . .	152
7.5.1	Contributions to Privacy Literature . . . . .	153
7.5.2	Contributions to SET Literature . . . . .	154
7.5.3	Reconciling Assurances Between Studies . . . . .	154
7.5.4	Suggested Interventions . . . . .	155
7.6	Conclusion . . . . .	157
7.6.1	Final Words . . . . .	157
<b>A</b>	<b>Experimental Survey Scales</b>	<b>159</b>
<b>B</b>	<b>Mean Comparisons for Studies 2A and 2B</b>	<b>162</b>
<b>C</b>	<b>Reference Tables for Chapter Seven</b>	<b>164</b>
<b>D</b>	<b>Survey Instrument</b>	<b>166</b>
	<b>appendix</b>	<b>184</b>
	<b>Bibliography</b>	<b>184</b>

## List of Figures

5.1	Vignette Text . . . . .	114
5.2	Example of the split screen in the experimental survey . . . . .	115
6.1	Boxplot of Assurance responses . . . . .	136
6.2	Boxplot of Vignette-Specific Assurance responses . . . . .	137
B.1	Mean comparisons for Study 2A . . . . .	162
B.2	Mean comparisons for Study 2B . . . . .	163
C.1	Means with Standard Deviations, Median, and Scale for dependent variables . . . . .	164
C.2	Characterizations of the Company-respondent relationship . . . . .	165

## List of Tables

3.1	Summary of Mobile Pregnancy Tracking App Participants . . . . .	36
3.2	Summary of 23andMe Participants . . . . .	37
3.3	Search Engine Participant Summary . . . . .	38
6.1	Trust Measure Combined . . . . .	121
6.2	Power Measure Combined . . . . .	123
6.3	Fairness Measure . . . . .	126
6.4	Privacy Measure – Disclosure . . . . .	129
6.5	Privacy Combined Measure . . . . .	130
6.6	Means and Standard Deviations of Responses to Suggested Standard Industry Practices. . . . .	138
7.1	Anonymization risk scenarios. . . . .	151
7.2	Identifiable data risk scenarios. . . . .	151
A.1	Responses to the real product question . . . . .	160
A.2	Real product responses summarized as a dummy variable. . . . .	161

## Acknowledgments

First, none of this would have been possible without the support of the amazing Deirdre Mulligan. Deirdre has steadfastly supported me from the moment I first walked into her office in 2005 and asked her if she could use someone like me with her research. I'm so grateful the answer was yes. I could not have asked for a better mentor during this journey.

Coye Cheshire has always been so generous with his time, starting with his first year teaching at Berkeley in 2005 when he helped me navigate my first experience with CPHS. His encouragement throughout this entire process, particularly the hours spent with me hashing out the early details of applying SET to privacy in his office, were invaluable. His patience with my occasional emotional turmoil during this process is much appreciated, as well as his administration of chocolate when needed.

Steve Weber and David Wagner graciously agreed to serve on my dissertation committee, for which I am eternally thankful.

Several people provided me with invaluable research support and assistance for this project. Brandon Shalchi helped me with background research, protocol development, and was a insightful sounding board as I got the project off the ground. He was instrumental to helping me develop the interview protocol.

Qing Huang and Matt Nagamine spent a summer with me combing through interview transcripts and classifying and organizing the excerpts, for which I am very grateful.

This dissertation would have never been finished without Rena Coen. Rena coded the interviews, classified quotes, and both copy edited and commented on the draft document. Rena, you are a rockstar.

I received institutional support from Berkeley's Center for Long Term Cybersecurity, which provided me with a seed grant for this research, and the Center for Technology, Society, and Policy's director's fund during my 2016-2017 academic year residency.

Many others helped me along the way. I want to express my gratitude to: Chris Hoofnagle, Airi Lampinen, Nancy Van House, John Chuang, Mary Madden, Heather Patterson, Liz Goodman, Ashwin Mathew, Aaron Smythe, Elaine Sedenberg, Richmond Wong, Sebastian Benthall, Alessandro Acquisti, and Nathan Good. I also must thank the participants of the I-School's Doctoral Theory and Research Workshop in 2015, the participants at my session of the 2015 Privacy Law Scholars Conference at Berkeley where I first presented this material, and the participants at the Security and Human Behavior Workshop at Harvard in 2016. I received valuable feedback from the participants in all of these venues.

None of this would have been possible without the loving support of my sons, husband, parents, and extended family. I thank you all with all of my heart.

# Chapter 1

## Introduction and Literature Review

### 1.1 Introduction

In this chapter, I introduce the topic of this dissertation and the problem space. I review the questions this research study answers, and provide an overview of the remainder of the dissertation.

### 1.2 Problem Space

Maintaining the privacy of one's personal information—one's choice of when to disclose it and to whom, how one maintains control over it, and the risks of disclosure—is one of the most important social issues of the internet era. For any person that engages with information technology in the twenty-first century, issues of personal information privacy have moved from margin to center. Concerns about the exposure of one's personal information online are no longer the province of tinfoil hat paranoia. Whether the risks are posed by hackers possessing one's health records, a credit agency exposing the details of one's financial life, or the collection of one's personal data by third parties, few (if any) internet users are free from persistent online information collection. A 2015 Pew Internet survey found that only nine percent of the U.S. public felt they had “a lot of control over how much information is collected about them and how it is used.” [68]

For the past decade, privacy researchers have focused on several domains, including: documenting public opinion about privacy attitudes and expectations; understanding how user interfaces affect disclosure; and focusing on understanding interpersonal privacy dynamics within social media settings, to name a few. All of this work shares the goal of furthering our collective understanding of how people think about information privacy in online settings, what they expect when disclosing their personal information, and why they make the disclosure choices they make. While there are many excellent studies that further these goals, a common element missing from the extant privacy research is an accounting of social context. Meaning, that as researchers consider the various factors that affect personal disclosure, they often do not consider the relationship between the discloser and the recipient, and how aspects of that relationship may directly or indirectly affect one's decision to disclose. One of the specific relational aspects that I believe is overlooked



but likely has a considerable impact on personal disclosure is the influence of power—the extent to which each actor in the relationship can control the terms under which personal information is exchanged, the options available to each actor to obtain similar resources elsewhere, how fair the exchange is to each actor, and the extent to which each actor benefits from it. The specific form of relationship that I believe needs more examination is that between individuals and the companies to whom they disclose their personal information in order to obtain access to a service.

This dissertation examines the effect of social context on personal disclosure between individuals and companies using social exchange theory (SET) as the framework for inquiry. Using the exchange of personal information for access to a service as the basis for study, I use the predictive aspects of SET to understand the role relational factors such as power play in personal disclosure. By incorporating SET's relational analytic into an empirical analysis of information privacy, I seek to move beyond a focus on individual cognition as a primary factor in understanding personal disclosure.

### 1.3 Research Motivations

I chose this topic for my dissertation after spending more than two years reviewing the research devoted to understanding information privacy decision-making—why people make the disclosure choices that they do.[32] [7] [2] [3] I was especially intrigued by the so-called *privacy paradox*: the fact that despite well-documented claims that people care about their information privacy, the public continues to use information-intensive services that require substantial personal disclosure. [88] [12]

While the studies I cite above (and others) attack the problem of understanding information privacy decision-making from multiple angles and disciplines, a common element is a focus on approaching it through understanding or decomposing individual cognition. While this approach is valid and has aided our understanding of personal disclosure, excluding the effect of social context only tells us part of the story. This dissertation is an attempt to broaden the focus of privacy research on personal disclosure by incorporating the influence of social structure.

The extant research on disclosure from individual to institution, specifically to the companies that collect personal information, is limited, especially as compared to privacy research focusing on technically mediated *interpersonal* disclosure. Much of the research that does examine institutional privacy issues, including some of my own, focuses on consumer expectations and how they align with specific outcomes, such as whether users are satisfied or comfortable with a company's data collection and sharing practices. There is also scant extant research that examines disclosure relationships between individuals and companies as a form of social exchange, or of privacy at all, from the perspective of SET. Additionally, in studies of personal disclosure much attention has been paid to one's *initial decision to disclose* (and the impact of privacy policies, user interfaces, mental models, and individual privacy attitudes, among others, on that decision) with little examination of the ongoing relationship that the user builds with the company after the initial point of disclosure. This focus isn't surprising; in many of the relationships that consumers build with companies, the bulk of personal disclosure occurs when first signing-up for a service.

Further, examining one's longer term disclosure during use of a product or service can be difficult to study, especially if the researcher is doing so without cooperation from the company in question. Regardless of these challenges, both my own scholarship as well as my review of the existing literature has persuaded me that moving the locus of analysis from individual cognition and behavior to include the relationship between the discloser (individual) and the recipient (the company) will yield a richer understanding of why people disclose.

### 1.3.1 Research Focus

This dissertation explores how the structure of relationships between individuals and companies influences individuals' decisions to disclose personal information. It seeks to fill a gap in the privacy literature, which has primarily sought explanations for the so-called "privacy paradox" in individual cognitive biases. I examine the ways in which *the relationship between the discloser and the recipient*, and in particular *the distribution of power within it*, affects individuals' decisions to disclose. I use SET as the framework for this inquiry because the transfer of personal information in exchange for a service is an exchange between social actors. Thus, SET provides an empirically tested scaffolding for exploring key features of these relationships and their impact on normative aspects of exchange (*i.e.*, individuals' perceptions of trust, fairness, and benefits) that affect disclosure choices. Incorporating these dimensions into analyses of relationship-based disclosure will help to explain individual decisions that appear paradoxical yet are predictable when examining them through a relational analytic framework. Further, it illuminates the role of structural power on individual decision-making, which I believe is both significant and under-appreciated in extant research on privacy and disclosure.

### 1.3.2 Dissertation Structure

The remainder of this dissertation is structured as follows. In this chapter, I review the literature in privacy and social exchange theory that relates to this research. Chapter 2 provides an overview of how I interleave privacy and social exchange theory, as well as the research questions and experimental hypotheses. Chapter Three reviews the qualitative study methods, and Chapter 4 details the findings from the qualitative interview study. Chapter 5 provides an overview of the goals and methods of the experimental survey portion of my research, and Chapter 6 reviews the detailed findings from the survey studies. Finally, I offer my synthesis of findings and conclusion in Chapter 7.

## 1.4 Related Literature

In this section I review the theories and related literature motivating this work: theories of information privacy and Social Exchange Theory (SET).

### 1.4.1 Information Privacy

Privacy is an essentially contested concept. While few disagree that privacy is a core human need, what privacy actually means across multiple contexts, how much is optimal, and how to practice it are topics of debate. Even Professor Daniel Solove, a leading legal privacy theorist, calls privacy “a concept in disarray.”[106]

Information privacy online encompasses many contexts, from selective disclosures to friends on social networks, to attempting to control profiling by search engines and advertisers, to name just two. Defining “information privacy” is in itself contested, particularly when many people feel as if they’ve already lost control of their personal information (and hence their sense of privacy). Interestingly, information science scholars have had only modest contributions to theoretical developments in information privacy. Information ethicist Luciano Floridi has written on the topic, offering a definition of information privacy based on “ontological friction:” the forces that oppose the flow of information within the infosphere, and the amount of work required for an agent to obtain information.[43] In other words, structural constraints (or a lack thereof) define whether information privacy can exist. However, when it comes to defining theories of privacy, Floridi draws on privacy as informational self-determination: “the right of individuals . . . to control the life-cycle . . . of their information and determine for themselves when, how, and to what extent their information is processed by others.”[43] H. Jeff Smith, Tamara Dinev, and Heng Xu, all professors in the information management and systems field, co-authored an interdisciplinary literature review of information privacy (defined as: “access to individually identifiable personal information”) that focuses on theories that are based on human interpretations of privacy by legal scholars and social scientists rather than inherent qualities of information. They divide the prevailing theories into two primary groups: value-based and cognate-based.[105] The values-based approach defines privacy as “a human right integral to society’s moral value system.” The cognate-based approach is “related to the individual’s mind, perceptions, and cognition rather than to an absolute moral value or norm.”[105] Within each of these groups there are further subdivisions; within the values-based approach, the authors identified two strains: privacy as a right to be protected by the state, and privacy as a commodity subject to economic principles. Within the cognate-based approach the two sub-strains are: privacy as a state of being (*e.g.*, of limited access or anonymity), and privacy as control, specifically the ability to control access to the self.

In this dissertation, I draw upon two theories of information privacy that incorporate disclosures between individuals and organizations: political scientist Alan Westin’s conception of privacy as control over one’s personal information, and philosopher Helen Nissenbaum’s theory of contextual integrity, focusing on context and information flows as the key elements defining information privacy. In terms of defining privacy for this dissertation, I will rely on Nissenbaum’s theory of contextual integrity to define whether a disclosure of personal information violates privacy or not. Operationalizing privacy is the practice of selective disclosure—the ability of an individual to choose to whom she wishes to disclose, what, when, and why—which follows Westin’s definition.

### 1.4.2 Privacy as Control

The late political scientist Alan Westin famously defined privacy as “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others,” and it is this definition that is most widely referred to when articulating privacy as an issue of personal control.[117] Privacy was, in Westin’s view, necessary for individuals to achieve self-realization, and individualistic in nature rather than of value to society as a whole. Professor Priscilla Regan notes that Westin viewed privacy “as fundamentally at odds with social interests.”[96] Westin developed his theory during an era (the 1960s) when both commercial and government data banks were on the rise, as records moved from paper to the first forms of digital media. There was growing public concern with the idea of one’s personal information being collected and stored, particularly without the ability for redress. First-party collection and use was the primary concern; third-party collection and resale of information wasn’t yet an issue as it was impractical, requiring both more computing power than was commercially available to most organizations, and necessitating the exchange of physical media (cloud storage did not exist—data was typically stored on magnetic tape). The very notion that one’s personal information could be stored and used by computers was perceived as sinister to some, particularly in a world where the abuses by governments of citizenship lists and the like during World War II were not far in the past.

Privacy articulated as control was a reasonable response to address the growing concerns over information privacy in the 1960s and early 1970s, given that in practical terms, the notion that one could control data collection was conceivable given how few entities collected personal information in digital form. Westin’s definition is focused primarily on identifiable data of the kind that creditors and governments collected, and he developed his theory at a time when personal disclosures were straightforward and made upon direct request: filling out a paper form, or talking to another person by phone. Data mining, inferential profiling, and online tracking were all non-existent (and the stuff of science fiction). The assumption that one could actually exert control over personal disclosure was not (at least, initially) practically unrealistic, though one’s ability for redress—to correct errors about information once gathered and digitized—was of concern. The operationalization of this theory on the ground was the establishment of the FIPs: the Fair Information Practice Principles, guidelines first recommended in the Report to the Secretary of Housing, Education, and Welfare by the U.S. Congressional Advisory Committee on Automated Personal Data Systems in 1973[48], and the U.S. applied them to federal agencies with the passage of the Privacy Act of 1974. The FIPs provide a framework of data protection based on self-determination.[83] Despite their inclusion in the Privacy Act, they have never been incorporated into an omnibus U.S. federal law regulating commercial uses of personal information. U.S. privacy law today remains sectoral in nature, applied narrowly to industries such as banking or health care. The collection of personal information by companies outside of specific sectors remains unregulated, with the Federal Trade Commission’s authority in those areas limited to policing unfair and deceptive trade practices, such as when companies violate their own privacy policies.

This control-based approach has evolved into the practice of notice and choice, where individuals are given notice of a company or organization’s information collection practices and can

freely choose to provide consent—or not—based on their reading and judgment of those practices. According to Bob Gellman, an expert on privacy and the FIPs, “Notice and choice is sometimes presented as an implementation of FIPs, but it clearly falls well short of FIPs standards.”[48] Notice and choice relies upon a model of the individual as a rational actor with full access to the information required to make an informed decision about her privacy. Functionally, when choosing to disclose personal information to a company, this requires an individual to both read a company’s privacy policy and comprehend it prior to disclosure. Armed with this information, the individual is presumed to be freely able to make an informed decision about whether to disclose or not—a choice that is generally presented on ‘take-it-or-leave-it’ terms that she may only accept or reject. One’s personal information is considered as part of the cost of the exchange for a service; as philosopher Helen Nissenbaum notes, “all is deemed well if buyers are informed of a seller’s practices collecting and using personal information and are allowed freely to decide if the price is right.”[85]

Privacy as control has shown its limitations in practice as advances in technology have made sharing and selling personal information widespread and instantaneous. The theory assumes a world where an individual can, in fact, exert practical control over information about herself. The evidence shows otherwise; a 2015 survey by the Pew Research Center found that only nine percent of respondents said they had “a lot of control” over how much information is collected about them and how it is used on a typical day.[68] Fifty percent responded “not much” or “no control at all.” By this measure, when half of the American public thinks they have little to no control over how their personal information is collected on a daily basis, then arguably the efficacy of defining privacy as control is questionable. Further, the theory relies upon a narrow view of personal information, which while still relevant, doesn’t reflect the evolution in the types of personal information now available about individuals, such as: information that qualitatively describes an individual (*e.g.*, her thoughts, emotions, preferences, opinions); information about an individual’s relationships (*e.g.*, social graphs), and information that today is on the cutting edge, such as brain wave patterns or our DNA. As Daniel Solove points out, classifications of particular kinds of information as public or private assume “that these are qualities that inhere in the information,” but that “no particular kind of information or matter, however, is inherently private.”[106] In sum, using categories of information (*e.g.*, social security numbers) or its disclosure status as the basis for determining privacy protection potentially leaves exposed information types that many would argue should be protected. This weakness highlights one of privacy-as-control’s largest omissions: the importance of context to privacy.

#### 1.4.2.1 Research Based on Privacy As Control

Researchers seeking to understand how individuals protect their privacy using the model of control, as well as why individuals appear to make paradoxical disclosure choices, have looked at information disclosure decision making through several lenses: by exploring privacy as a rational process of weighing costs and benefits and risk (the “privacy calculus”); by exploring the divergence between people’s expressed preferences for privacy and behaviors that put privacy at risk (privacy paradox); and, by examining how cognitive heuristics and biases influence privacy

decision-making specifically through the framework of behavioral economics. Across these categories, this line of research shares a fundamental reliance on protecting information privacy through personal control—that, at its core, privacy is a value that can and should be practiced through individuals exercising control over personal disclosure.

### 1.4.2.2 Privacy As A Rational Calculus

One area where privacy as control has been operationalized consistently is in the area of opinion polling research focusing on understanding and documenting people's attitudes and expectations regarding information privacy, specifically the public's perception of control over their personal information. Information privacy surveys in the U.S. were initially dominated by Alan Westin. From the 1970s onward Westin conducted over thirty surveys where he analyzed public sentiment based on the three privacy categories he created: privacy fundamentalists, privacy pragmatists, and the privacy unconcerned.[94] While Westin's surveys did address aspects of online information privacy, historically they were focused on pre-internet business and institutional relationships, and did not address the impact of the internet and other technologies on information sharing and disclosure. Westin ceased conducting surveys in the mid 2000s, but his influence remained strong for about a decade until several of us in the field raised concerns with the validity of his categories both generally and specifically in the context of internet-enabled data sharing. I published a paper in 2014 at a workshop devoted to critically assessing the applicability of Westin's categories, in which I used data from a 2009 survey I co-authored in a series of regression analyses demonstrating that Westin's categories were not useful for predicting privacy attitudes.[60]

There is also a separate line of survey work focused on building causal models for understanding privacy attitudes and choices, identifying antecedents in decision-making, and privacy scale development.[70]; [103]; [95] A core assumption underlying this work is the privacy calculus, as described by both Culnan and Armstrong and Dinev and Hart: that people follow a rational process of weighing risks and benefits when making a privacy disclosure decision.[30][33] For example, Dinev and Hart constructed a causal model of privacy (IUIPC) based on the relationship between one's personal dispositional factors, demographic covariates, the type of information requested, and an individual's context-specific trust and risk beliefs, which they find influence her behavioral intention (*e.g.*, decision to disclose). This work assumes fixed privacy beliefs and attitudes that are borne out by deliberate actions. While this and other research in this vein have uncovered interesting associations related to privacy decisions, what it has not explained is the privacy paradox—the divergence between the generally high and consistent support for privacy found in both the opinion polling and survey modelling research and behavior. The calculus itself relies upon an assumption that may feed the paradox: namely, a reliance upon risk evaluation as a key decisional component.[65] While certainly many overt privacy choices have a risk component, not all privacy decisions people make involve conscious or direct calculations of risk. Thus, to the extent that privacy choices are evaluated as 'paradoxical' because they appear to be poor calculations of risk, from the decision-maker's perspective, the choice may not have been one of risk evaluation whatsoever. This may be especially likely in cases of personal disclosure on social networks or in social media, where disclosures are often made in an interpersonal context—conversation with

others (with a lack of understanding or visibility about the true reach of one’s audience)—and not as explicit choices of weighing risks and benefits.[37] [75][115]

### 1.4.3 The Privacy Paradox

In 2006, Professor Susan Barnes published a piece in *First Monday* that introduced a new turn of phrase. Reflecting on a then new phenomenon spreading among teenagers, she observed:

“Social networking sites create a central repository of personal information. These archives are persistent and cumulative. Instead of replacing old information with new materials, online journals are archive-oriented compilations of entries that can be searched. While American adults are concerned about how the government and corporations are centrally collecting data about citizens and consumers, teenagers are freely giving up personal and private information in online journals. Marketers, school officials, government agencies, and online predators can collect data about young people through online teenage diaries. Herein lies the privacy paradox. Adults are concerned about invasion of privacy, while teens freely give up personal information. This occurs because often teens are not aware of the public nature of the Internet.”[12]

With this article Barnes introduced the “privacy paradox,” a phrase that inspired a new wave of privacy research. “Paradox” appeared to be the perfect word to explain a growing global phenomenon that continues to exist today. In the U.S., despite survey research consistently reassuring researchers that privacy is an important value that the majority of the public cares about (including teens and young adults), growing evidence demonstrated that people acted to the contrary. Over the past decade, social networking services and social media tools, especially Facebook and Twitter, experienced exponential growth based upon their users’ willingness to share often very sensitive and personal information about themselves. Despite claiming to care about privacy, many of us willingly feed the beast, posting photos, tweets, status updates, profiles, and personal information of every kind and flavor, across the Internet.

The privacy paradox isn’t limited only to personal disclosure on social networking and social media platforms. As researcher Mary Madden reported in a 2014 privacy survey she conducted for the Pew Internet, Science and Technology Project, the U.S. public expresses great anxiety about the privacy of their personal information in the hands of corporations and government and across different communication channels.[67] Yet, as journalist Claire Cain Miller points out in a 2014 *New York Times* article about the survey, “Americans say they are deeply concerned about privacy on the web and their cellphones. They say they do not trust Internet companies or the government to protect it. Yet they keep using the services and handing over their personal information.”[76] Madden reported: “Some 55 percent “agree” or strongly agree that they are willing to share some information about themselves with companies in order to use online services for free.”[67]

Despite broader evidence of the paradox (as my own research into mobile app usage suggests), the research community has largely focused on the paradox as Barnes originally specified it, primarily focusing on the disconnect between attitudes or beliefs and actual disclosure behaviors in

social networks.[59] [62] Work by Roberts draws upon the theory of planned behavior to help explain the paradox, arguing that the theory provides a framework for empirically testing it.[97] A key assumption is that personal beliefs form intention, which in turn facilitate behavior. Roberts identifies informed awareness (privacy issues and context), subjective norms (one's social circle, media attention focusing on privacy), and control (ease of use of the interface, complexity) as the determinant factors influencing the paradox and the disconnect between intended behavior and actual behavior. A series of studies by Sabine Trepte and her colleagues provides a clearly specified approach to the paradox supported by empirical work. Dienlin and Trepte have dissected it by both differentiating between types of privacy (informational, social, and psychological) and using the theory of planned behavior to link privacy behavior with privacy attitudes, concerns, and intentions.[32] In one study, the authors argue that "disparities between attitudes and behaviors might not indicate paradoxical and inconsistent behavior" and that instead they may "each address different dimensions of privacy." [110] They note that one of the assumptions behind the privacy paradox is that privacy behaviors can be predicted by privacy attitudes, and that "behavior can be successfully predicted by attitudes only under certain preconditions." One of their key findings from a longitudinal study of SNS users was to suggest that future research assess and operationalize privacy behavior on different dimensions, as their results "suggest that in previous research, attitudes and behaviors were unrelated because the ways in which they were operationalized did not adhere to the boundary conditions of attitude-behavior consistency." The authors also question the normative assumptions of the paradox, noting that it relies upon a view of rational users engaged in risk/benefit analysis who "deliberately choos[e] privacy risks to take full advantage of the benefits offered by the services of the social web," which is in turn presented as both risky as well as manageable though tools such as privacy settings.

That assumption of choice is one that this dissertation questions. My colleague Chris Hoofnagle has questioned the knowledge gap between how people think their privacy is protected and the actual ways in which it is or is not in the law.[54] In his telling, the paradox is explained by the fact that people act as they do because they assume their privacy is offered greater protection by the law than it actually is. Hoofnagle also suggests that consumers often articulate their privacy concerns using 'voice' (based upon the work of Albert Hirschman), and that one of the assumptions of the paradox—that consumers would simply exit, or cease using suspect products and services if they were truly dissatisfied—is flawed. One reason, he suggests, is that many of the products and services to which we disclose personal information have high switching costs. Journalist Julia Angwin recently documented this problem in her book, *Dragnet Nation*, illustrating the extremes to which she had to go to in order to use modern information communication technologies that did not compromise her privacy.][9] These examples highlight that one of the underlying premises of the paradox—that consumers will vote with their feet if they feel their privacy is being violated—may, in fact, be nearly impossible to achieve if they wish to participate at all in a digitally networked society.



#### 1.4.4 Privacy and Behavioral Economics

Of all the empirical privacy research that has influenced this dissertation, the most central to it is the research conducted within the framework of behavioral economics (BE). The BE line of privacy research is primarily conducted by Alessandro Acquisti in collaboration with colleagues (Leslie John, George Loewenstein) and former students (Laura Brandimarte, Idris Adjerid). Acquisti and his colleagues have questioned the normative model on which the privacy calculus is based in a decade-plus body of work. Acquisti and Grossklags' 2005 article, *Privacy and Rationality*, predates the coining of the privacy paradox in its critique of the individual as a 'rational economic agent,' noting inconsistencies in privacy decision-making and behavior based on "incomplete information, bounded rationality, and systematic psychological deviations from rationality".[2] While BE generally relies upon a definition of privacy that I find limiting (discussed in depth below in Section 1.4.4.1), at the same time this research has highlighted the importance of context in understanding privacy decision-making and the extent to which individual privacy preferences are malleable and exploitable, particularly by the design of user interfaces.

A 2015 article in *Science* by Acquisti, Brandimarte, and Loewenstein summarized the extant empirical information privacy research in BE and the social sciences generally (though it also included work authored by information scientists, computer scientists, and others outside social science disciplines).[3] The authors focused on three themes: individual preferences and uncertainty about the nature of privacy trade-offs; the context-dependence of privacy preferences; and the malleability of privacy preferences by those with greater insight into their determinants. They argue that these themes are interdependent, with context-dependence amplified by uncertainty, and privacy preferences and behaviors "malleable and subject to influence in large part because they are context-dependent and because those with an interest in information divulgence are able to manipulate context to their advantage." [3]

They define uncertainty as "incomplete and asymmetric information"; individuals generally do not know what companies or other actors know about them, which makes it difficult for them to make fully informed decisions about personal disclosure. Further, the privacy harms caused by disclosure are often intangible, and sometimes even untraceable. The authors further note that privacy is rarely an "unalloyed good"; it typically involves trade-offs (most often between convenience and disclosure), and that "elements that mitigate one or both of these exacerbating factors, by either increasing the tangibility of privacy harms or making trade-offs explicit and simple to understand, will generally affect privacy-related decisions." [3] They reference a number of studies explicating the privacy paradox, which they suggest is evidence of individuals' uncertainty about privacy preferences, as well as an inability for surveys (where people generally express high support for privacy) to predict behavior.

They cite the importance of context for privacy, following the theory of contextual integrity: "[t]he rules people follow for managing privacy vary by situation, are learned over time, and are based on cultural, motivational, and purely situational criteria." Finally, they focus on one of the central contributions of BE research to privacy: the malleability of privacy preferences, "a term we use to refer to the observation that various, sometimes subtle, factors can be used to activate or suppress privacy concerns, which in turn affect behavior." [3]) They argue specifically

that “some entities have an interest in, and have developed expertise in, exploiting behavioral and psychological processes to promote disclosure.”[3] At a time when many are beginning to question the influence of technology such as smartphones and social media on our lives this may not seem controversial, but this evolution in popular thinking marks a sea change in the popular conception of our relationship with technology. There appears to be a fundamental questioning of an individual’s ability to have an equitable and fair relationship with technology companies (and that in turn individuals can make beneficial and informed disclosure choices) given the emerging evidence that companies use our vulnerabilities against us to their advantage. If we are being manipulated into making choices that aren’t in our self-interest, how can a regime such as notice and choice, based entirely on one’s ability to assert her self-interest, ensure we are not being unfairly exploited?

Acquisti *et al* summarize the relationship between the themes of uncertainty, context, and malleability: “[u]ncertainty and context-dependence imply that people cannot always be counted on to navigate the complex trade-offs involving privacy in a self-interested fashion. Malleability, in turn, implies that people are easily influenced in what and how much they disclose.”[3] This interpretation strongly questions an approach to privacy based on the rational actor approach, both from a design and a policy standpoint. To anyone following developments in the behavioral sciences, this viewpoint is evidence of the influence of Kahneman and Tversky’s important contributions on heuristic thinking (System One) versus methodical thinking (System Two), and the operationalization of their findings into choice architectures.[58] Their work has impacted both policy (from arguments around nutrition to financial services) and product design.[109] However, it has yet to have a direct impact on privacy policy; while in the U.S. the Federal Trade Commission has recognized the importance of context in its consumer guidance[25], there hasn’t been as yet any reckoning with how the consent model specifically might transform, as well as consumer protection in general, if the rational actor approach is modified or abandoned for what Thaler and Sunstein call “libertarian paternalism”: the application of behavioral science and economics to “steer people’s choices in welfare-promoting directions without eliminating freedom of choice.”[108] Acquisti, Brandimarte, and Loewenstein conclude their work with the suggestion that resolving these issues will require a policy intervention, rather than industry self-compliance, noting: “[a] goal of public policy should be to achieve a more even equity of power between individuals, consumers, and citizens on the one hand and, on the other, the data holders such as governments and corporations that currently have the upper hand.”[3] I have taken this statement to heart; this dissertation attempts to explore that dimension explicitly as part of the problem—essentially, that the structural power differential between individuals and companies and institutions allows the companies to deploy this bag of tricks in their relationships with individuals, and the existing notice and consent structure is both ineffective in countering it, as well as contributes to the structural inequity that makes it possible.

#### 1.4.4.1 Economics of Privacy

While behavioral economics integrates social psychology into its set of tools for understanding decision-making, its basis for defining privacy is rooted firmly in economics, which I believe

limits its interpretation. In *The Economics of Privacy*, Acquisti and coauthors Taylor and Wagman provide a summary of how the field of economics conceptualizes privacy: “[a]t its core, the economics of privacy concerns the trade-offs associated with the balancing of public and private spheres between individuals, organizations, and governments. Economists’ interest in privacy has primarily focused on its *informational* dimension: the trade-offs arising from protecting or sharing of personal data.”[6] They define privacy as control over the act of disclosure, with the gains and losses from disclosure conceived of in economic terms. Both privacy and personal information are viewed generally as economic goods with a value that is context-dependent.

My concern with this view is that while I do not dispute that disclosure can have economic effects, conceiving of privacy as an economic good (as opposed to, for example, a social good) forecloses the aspects of privacy without an economic basis, and focuses discussions of privacy harms on those with economic consequences. For example, some theorists argue that privacy is important for the dignitary value it offers to humanity (and in fact, European Union law defines it as such), an aspect that is difficult, if not impossible, to quantify. My more specific concern with an economic conceptualization of privacy is that based on my own research experience people generally do not articulate either their interest or concern with privacy, nor the harms they experience, in economic terms and that conducting empirical research into privacy using an economic framework as the basis may too narrowly frame people’s motivations for their privacy-related decisions. This is particularly the case in an information economy where the monetary value of personal information is difficult to assess, and most people exchange information with companies for access to ‘free’ services.

To be fair, not all research in BE approaches privacy in purely economic terms, and of course there are circumstances in which examining privacy as an economic good is both necessary and useful. However, limiting interpretations of privacy to it as an economic good, I argue, limits the inclusion of power differentials as a contributing factor to understanding disclosure. Acquisti and his co-authors note that while privacy economics “focuses on measurable, or at least assessable, privacy components . . . using economics to study privacy does not imply the belief that such other, noneconomic dimensions do not exist or are unimportant.” Hence, the marriage of cognitive and social-psychological theoretical frameworks with economics. While some of Acquisti’s studies do examine privacy explicitly with an economic focus [4] [5], others instead evaluate privacy based on cognitive and interface aspects, such as framing effects [19] [8] These studies define privacy as control and focus primarily on trade-offs, but do not attach an explicit economic value to privacy.

### 1.4.5 Privacy as Contextual Integrity

The other theory of privacy motivating this dissertation is the theory of Contextual Integrity (CI), which the philosopher Helen Nissenbaum developed to specifically incorporate the influence of context in privacy. CI introduces the concept of relativity into privacy and disclosure with the theory’s emphasis on context-dependent information flows. CI posits that an individual doesn’t lose her expectation in the privacy of her personal information simply because she has disclosed it; the context in which it has been disclosed and the norms governing the information flows within the context are what determines whether a disclosure of one’s personal information violates privacy.

The flow of information is analyzed with conformance with the norms appropriate for the given context. If the flow of information disrupts existing norms, it violates the discloser's privacy. The relevant privacy norms themselves "can be assessed in terms of how they affect the interests of relevant parties ("stakeholders") and how they impinge on societal values, such as equality, justice, fairness and political liberties." [16]

Examining disclosure through the lens of CI resolves—at least, theoretically—many of the shortcomings inherent to privacy as control, although Nissenbaum argues that CI can accommodate privacy as control and thus one need not choose between the two theories: "[t]he idea that privacy implies a limitation of access by others overlaps, generally, with the idea of an informational norm." [86] Control, she explains, is one of several transmission principles central to the theory. CI is more holistic than privacy as control, relying upon both context and prevailing social norms to define violations of privacy. As Nissenbaum writes, "[t]he decision heuristic derived from the theory of contextual integrity suggests that we locate contexts, explicate entrenched informational norms, identify disruptive flows, and evaluate these flows against norms based on general ethical and political principles as well as context-specific purposes and values." [85] A core aspect is the ability to locate and identify contexts. Nissenbaum argues that even in online contexts, which may seem divorced from physical reality, "online realms are inextricably linked with existing structures of social life" and that the "equilibriums achieved in familiar contexts may provide analogical guidance" for online life. [85]

But while CI results in a more robust and nuanced understanding of privacy than privacy as control, how to operationalize CI into practice isn't obvious. Nissenbaum has collaborated on several projects with computer scientists to model applications of CI ([13]; [42]; [14]). A 2017 paper Nissenbaum co-authored with Seda Gürses and Sebastian Benthall surveyed twenty papers authored by computer scientists to better understand how they applied CI to their research, identifying gaps in the theory: a need for CI to be more "modular" ("giving guidelines for design and research at specific levels of the technical stack"); more clarity on the concept of context ("a theoretical account of how social spheres relate to socio-technical situations") as well as addressing situations encompassing multiple contexts; and connecting norms in the abstract to norms on the ground ("provide a way of translating from the information norms of social spheres into a characterization of enumerated and discrete privacy threats"; "articulate the special place for user expectations, preferences, and control within the general framing of appropriate flow"; "clarify the relationship between social spheres and the law"). [16] Finally, they also identified the need for CI to recognize the temporal aspect of privacy—when the risks posed by information flows evolve due to the passage of time. [16]

Perhaps the biggest challenge to CI is the not insignificant question of how to define context. Nissenbaum elaborated on this challenge in a paper published in 2015 in response to the 2012 White House report on consumer privacy which included a *Principle of Respect for Context*. [87] She identified four interpretations of context in response to the report: "context as technology platform or system, context as sector or industry, context as business model or practice, and context as social domain." [87] Of these four, she argues that representing context as a social domain "is far more likely to yield positive momentum and meaningful progress in privacy law and policy" than the other three, and that "respect for context amounts to respect for contextual integrity." [87]

I would further argue that defining context in terms of the social domain is by definition people-centric and inclusive of human values, and is necessary if the object of policy protection is people (as opposed to businesses). However, this is only one example of the struggle implicit in defining context—even within the field of human computer interaction there is not a simple agreement around defining and implementing context.[34]

Further, CI's default assumption that norms provide an inherently positive compass for how information should flow is of concern. For example, as the internet has evolved over the past twenty years, the norms governing personal information disclosure in the U.S. have shifted, with people disclosing ever increasing quantities and types of personal information to companies. I argue that this shift has been driven by private companies, who have normalized personal disclosure such that it is nearly impossible to interact with any digital system without disclosing some amount of personal data. While we may still be in the midst of assessing how thoroughly these changes “impinge on societal values,” one of the arguments I will explore later in this dissertation is the connection between structural power and norms, and whether power imbalances may aid powerful actors in shifting norms in their favor. Perhaps, as described above, this is a debate over what constitutes context, with our legal and policy structures embracing context as business model or technology platform and thus deeming information flows as appropriate when they adhere to those norms. However, it is important to ask: how are norms established within a given context? Can more powerful actors influence the development of norms in their favor? As an example, as I write this dissertation several groups have joined together to try to force technology platforms to shift their design choices towards building products that are less ‘addictive’ as well pose less harm towards children. [47] [18] [71]; The products and the product designers have established norms of use that a growing number of people are concerned cause harm, and a countermovement is growing to shift these norms in a different direction. I am uncertain whether this is a battle over context or a battle over norms (or both), but in any case, the product designers were likely designing for the good of their employers and not for the good of society.

## 1.5 Social Exchange Theory

In this section, I introduce the concept of Social Exchange Theory and discuss the related research I draw on for this dissertation.

Social Exchange Theory (SET) is in actuality a set of theories focusing “on the benefits people obtain from, and contribute to, social interaction.”[80] The main assumptions (or scope conditions) of social exchange theory are that:

1. Behavior is motivated by the desire to increase gain and to avoid loss;
2. Exchange relations develop in structures of mutual dependence (that there some reason to engage in exchange to obtain resources of value);
3. Actors engage in recurrent, mutually contingent exchanges with specific partners over time;
4. Valued outcomes obey the psychological principle of satiation [77][80].

Working from these initial assumptions, researchers can use SET to make predictions about the behavior of actors within exchange relations, as well as the effects of different factors on exchange outcomes, that meet these four conditions.

Social exchange theory emerged principally from the early writings of sociologists George C. Homans [53], Phillip Blau [17] and Richard Emerson [40] [39] [38]. These theorists were interested primarily in the micro-level social processes that occur between individuals or small groups, and applied microeconomic theory to understand them. As Emerson describes it, “the exchange approach in sociology might be described, for simplicity, as the economic analysis of non-economic social situations.”[41] As such, social exchange theory shares many of the same core assumptions as microeconomics, but as Emerson elaborates, “neoclassical economic theory is organized so heavily around rational individual decision making in a perfectly competitive market that its applicability to tradition-bound or normatively regulated behavior outside of competitive markets is placed in doubt, yet goods are produced and distributed through exchange.”[41] Molm and Cook elaborate on this point further: “whereas classical microeconomic theory typically assumed the absence of long-term relations between exchange partners and the independence of sequential exchange transactions, social exchange theory took as its subject matter and its smallest unit of analysis the more or less enduring relations that form between specific partners.”[80]

### 1.5.1 Forms of Exchange

Working from these four core assumptions there are several different forms that exchanges can take. The first distinction Molm and Cook make is between direct exchange and indirect exchange. In direct exchanges, two actors exchange with one another. In an indirect exchange, three or more actors are linked in an exchange relation, but the benefits from the exchange are not directly reciprocated; for example, Actor A may make an exchange with Actor B that indirectly benefits Actor C. Within the classifications of direct and indirect exchange several sub-forms have been identified, based on characteristics such as reciprocity (reciprocal exchange), negotiation (negotiated exchange), and contributions to groups (generalized exchange, between three or more actors), with different outcomes based on the type of exchange. Cheshire *et al* argue that the dominant factor contributing to this differentiation is the “underlying difference in types of risk and levels of uncertainty involved in each mode of exchange.”[24] Risk and uncertainty are core features of exchange theory, as they required for the development of trust in social exchange. In a binding exchange, assurance structures (discussed below) exist to reduce uncertainty and to provide a form of enforcement that the terms of the exchange will be upheld by a third party or external structure.

Trust plays a central role in exchange relationships. Molm *et al* tested the classical proposition in SET that trust was more likely to develop in reciprocal rather than negotiated relationships and found strong support: negotiated exchanges presented lower amounts of risk and uncertainty as compared to reciprocal exchanges. This difference was attributable to the reliance in negotiated exchanges on jointly negotiated agreements that bind the actors to a specific outcome (as well as assurance structures which I discuss below, although this aspect was not tested in this study), whereas negotiated relationships are based on trust that develops between actors.[82] While trust

may be a stronger component of reciprocal relationships, it is still a necessary ingredient in negotiated relationships, though its salience may be mediated both by risk and assurances.[79]

In this dissertation, I will be focusing on the following forms of exchange: *direct negotiated binding exchange*, *direct reciprocal exchange*, and *generalized exchange*.

### 1.5.1.1 Direct Negotiated Binding Exchange

In a direct negotiated binding exchange (DNBE), the exchange occurs directly between two actors with the expectation that the exchange itself has been negotiated, meaning the product of the exchange is agreed upon based on a joint decision process and known to each actor. According to Cheshire *et al.*, “[i]n negotiated direct exchange the only risk involved is the risk of not concluding a successful exchange by failing to reach an agreement, since jointly reached agreements are binding on the actors.”[24] Given the specification of the terms and the presence of assurance structures, DNBEs are considered a low risk/low uncertainty exchange condition. A common example of a DNBE is of an exchange between a buyer and seller on eBay, where eBay’s guarantee policies act as a binding assurance between the two actors to guarantee that both actors will uphold their side of the purchase agreement.

### 1.5.1.2 Direct Reciprocal Exchange

A direct reciprocal exchange is an exchange between two actors “in which the terms are not negotiated.”[80] According to Molm, reciprocity is “the giving of benefits to another in return for benefits received.”[78] A reciprocal exchange includes both high uncertainty and high risk, as exchange partners do not know in advance what might be exchanged, or whether a partner will reciprocate. Reciprocal exchanges abound in social life; to give a single example, allowing a neighbor to borrow one’s lawn mower with the undefined expectation that you may borrow something from her in turn one day is an exchange based on reciprocity.

### 1.5.1.3 Generalized Exchange

Generalized exchange includes “indirect reciprocity between three or more individuals.” [24] The resources each individual contributes may go toward other individuals, or a collective group, without the expectation of a direct exchange in return. Donating blood is an example of a generalized exchange, where the blood donor gives blood to a pool without an expectation that she will benefit from her specific donation.

## 1.5.2 The Object of Exchange

In all social exchange interactions, the object of exchange is a key factor. One early effort to create a typology of exchange goods is Resource Exchange Theory.[45] This theory attempts to identify the structure of the exchange interactions between individuals by classifying the nature of resources exchanged. These resources include love, services, money, goods, status, and information. Cheshire specifically argues that information is a good that can be an object of exchange as

it is a good “much like any other good, since it can be transferred and it has value,” albeit with different properties than physical goods.[22]

In this dissertation, the object of exchange I examine is personal information, specifically *disclosures of personal information made in exchange for access to specific information-based products or services*. I define an act of disclosing personal information for services as a type of information exchange—the act of revealing personal information to another person, group, or organization.

According to Molm and Cook, an exchange must meet the four scope conditions of the theory described earlier. For the qualitative portion of this dissertation I selected three information-based exchange contexts to examine in depth that met those conditions, which are:

- Pregnancy tracking applications: users are required to exchange personal information for recurring access to the app;
- Direct-to-consumer genetic testing: users are required to exchange personal information with the genetic testing company in order to receive test results, health information, and information about genetic relatives;
- Internet search engines: users provide queries to search engine providers in exchange for relevant information. A single isolated query may not contain personal information, but repeated queries to a single search engine over time may contain personal information or reveal information about the user in aggregate.

Additionally, for the experimental portion of this project, I created a model information exchange context requiring the participant to imagine exchanging of a range of personal information with a company on an ongoing basis in order to obtain access to personalized health and fitness information through a wearable tracking device. The four exchanges complied with the four scope conditions:

- There was a mutual dependence on the other actor for a benefit. Across all cases, the companies needed the individual to disclose personal information in order to provide the service, while the individual could only access the benefit of the service through disclosure (though the value of the benefit to each individual varied);
- The act of disclosing led to a positive outcome or benefit for the individual: *e.g.*, needed information, personal insight;
- The relationship was recurrent, though the frequency and depth of the exchanges varied; for example, both pregnancy tracking and genetic testing typically required the primary disclosure at the start of the relationship with far less required disclosure on an ongoing basis;
- The outcomes obeyed the principles of satiation or diminishing marginal utility; for example, an individual’s knowledge seeking behavior is satiable in that once the specific information need is fulfilled, one need not continue to repeat the behavior. An online search to find



a nearby pizza restaurant may require several queries, but ultimately the information need will be adequately satisfied (or abandoned). At some point additional information will not continue to address the need.

### 1.5.3 Assurances and SET

As mentioned above, trust is a key component of any exchange relationship, but negotiated exchanges rely less on trust and more specifically on the jointly negotiated binding agreement between actors. Agreements may be bound by assurances: “[m]echanisms that provide assurance include legal or normative authorities that impose sanctions for violations of agreements or failure to fulfill one’s obligations, guarantees such as collateral that protect against loss, warranties that assure certain standards of quality, and so forth.”[82] Yamagishi and Yamagishi define the difference between trust and assurance in exchange relationships in a 1994 paper investigating the disparity between Japanese and Americans’ self-assessments of trust.[120] Despite the limitations of the study itself (a non-representative survey of 1,136 Japanese and 501 Americans), the authors’ theoretical findings have been highly influential for subsequent theory development and research into the forms of social exchange. They define trust as expectations of benign behavior based on inferences about a partner’s personal traits and intentions, and assurance as knowledge of the incentive structure surrounding the relationship.[120] Both trust and assurances work to reduce social uncertainty between partners, with social uncertainty defined as “a mixed motive incentive structure in which the actor does not have the capability of correctly detecting the partner’s intentions.”

Molm, Takahashi, and Peterson build on the Yamagishi paper in a 2000 study where they explicitly tested the proposition that negotiated exchanges rely on assurance, while reciprocal exchanges rely on trust.’[82] Their findings, based on a lab experiment, support their propositions, including that a reliance on assurance structures in negotiated exchanges precluded the need for trust between partners. Further, they observed that behavioral commitment between partners was no more likely to develop in one form of exchange, but that “important determinant of commitment is the structure of power and its effects on the opportunity costs of commitment.”[82] Finally, one qualitative study exploring the role of assurances in an exchange relationship following a similar design to the qualitative portion of this dissertation is Lampinen and Cheshire’s work on negotiated exchange relationships on the Airbnb platform.[63] In a purposive sample of twelve Airbnb hosts, the authors discovered that the Airbnb hosting platform functioned as an assurance structure, mitigating the risk and uncertainty with opening one’s home to strangers.

### 1.5.4 Power and SET

According to Molm and Cook, “*power* is a potential that derives from the structural relations among actors—their relative dependence on one another.”[**emphasis in original**, 80] Much of the research examining power differentials in SET has focused on network structures, such as how an actor’s relative position in a network can lead to predictable differences in the use of power.[29] According to Molm, “Emerson’s theory of power-dependence relations provided the impetus for the focus on structural power by proposing that structure determines power use, regardless of ac-

tors' intentions to use power or their awareness of the power structure.”[77] Cook and Emerson argued that “power is an attribute of position in a network structure observable in the occupant’s behavior, even though the occupant does not know what position or what amount of power s/he possesses.”[27] Emerson’s prediction that “an imbalance or inequality in structural power produces corresponding inequality in exchange benefits that favors the less dependent actor in an exchange relation is one of the more robust findings in the social science literature.” [77] Other work, primarily by Molm, has focused more specifically on coercion and punishment power within the context of reciprocal exchanges.[29]

Power dependency is another key aspect: “an actor is *dependent* on another to the extent that outcomes valued by the actor are contingent on exchange with the other.” [80] Exercises of power occur when one actor uses this potential to gain advantage in exchange over another; “actors with few or no alternatives are vulnerable to exploitation.”[80] Key to power dependency is the availability of alternatives. One’s dependence *increases* as one’s access to exchange alternatives *decreases*, and the actor with greater control over the valued resource in the exchange has more capacity to exercise power.

The explicit recognition of power in exchange relationships differentiates the theory from the basic model of economic exchange, where actors are generally portrayed as on equal footing. It incorporates both the reality of inequality between social actors as well as the influence of social network structure on exchange relationships. As Cook and Cheshire describe it, “[p]ower inequality is an inevitable outcome of differentiation in resources and structural position. Over time some actors gain positions of advantage in their exchange relations (or networks of exchange relations) and thus have the capacity to exploit this advantage.”[119] The authors also noted that without differentiation in resources and preferences between actors there would be little reason for exchange; universal equality would not foster exchange. Power differentials between actors are not inherently bad, nor do they exclusively result in poor outcomes for the power-disadvantaged actor.

However, Emerson argued that power-unbalanced relations are inherently unstable. According to Cook *et al.*, “[t]he important feature of power inequality is that it creates strains in exchange relations and provides an impetus for structural change.”[26] In unbalanced relations power use by the more powerful actor tends to increase over time, and while power use can be purposive, it need not be—the structure of the relation itself may simply create power differentials.[80] Emerson identified four balancing operations that can occur to correct unstable relations: withdrawal from the relation by the weaker actor; network extension, where disadvantaged actors seek alternative partners; coalition building among weaker actors; and status giving from the powerful actor to the weak actor.[26] Cook and Emerson also suggest that “normative concerns operate as constraints upon the use of power in exchange networks.”[27]

While assessing the effect of structural power on the negotiated exchange is beyond the scope of this study, it is worth noting that many information-intensive companies benefit from network effects, which provide them with greater structural power both over other companies as well as individual consumers. The rise of the “Frightful Five,” as *The New York Times* technology reporter Farhad Manjoo calls Apple, Amazon, Google, Facebook, and Microsoft, exemplifies the effects of structural power.[72] Madden and Rainie’s 2015 Pew Internet study found that over 90 percent of the public believed that they had lost control over their personal information.[68] The subtext

of these findings points both to a loss of individual power as well other factors I will assess in this study: respondents' perceptions of fairness and distribution of benefits. In short, some people may disclose their personal information in part because they feel powerless to affect the terms by which they disclose. Further, the effects of structural power are also felt when users attempt to cut ties with a powerful company, as the 2018 movement to delete Facebook demonstrates.[56] Many Facebook users, fed up with the combination of privacy and 2016 election hacking scandals, confronted the challenges of doing so, given that the platform remained the single most useful way many stayed in contact with family and friends. Assessing how fair respondents find these relationships, and who stands to benefit more, will provide additional insight into these issues.

## Chapter 2

# Motivations, Research Questions, and Study Hypotheses

In this chapter, I discuss how this dissertation interleaves social exchange theory with privacy research. I discuss how the theories interrelate, and review the empirical research from both SET and information privacy that informed my research design. I conclude with the questions and hypotheses that motivate the two research studies I conduct.

### 2.1 Interleaving SET and Privacy

The explicit application of SET to an analysis of information privacy is new territory; at the time of publication, I found only two papers that explicitly incorporated SET and privacy. Stanton and Sham used SET in a study examining how increased capacity for technological surveillance affected employer-employee relationships, including the IT staff who could enable electronic monitoring.[107] The authors used indirect generalized exchange as the “explanatory framework for examining the dynamics of power and information control” within a set of non-profit organizations. Using qualitative interviewing as their method, they found support for their use of SET in examining the power relationships between managers, employees, and IT staff. Luo produced a paper arguing that consumers’ growing privacy concerns with e-commerce (circa 2002) could be resolved by building trust relationships between companies and customers.[66] The author used SET and relationship marketing theory as the basis of his analysis, arguing that the theories should inform the development business to customer relationships, with a focus on repeated transactions to build trust and decrease privacy concerns. Neither paper contributes directly to this dissertation, though Stanton and Sham’s study is useful as an exemplar of the use of similar methods.

In terms of integrating SET and privacy in this dissertation, I start first with the assumption that privacy is necessary for societies to function as it provides the psychological space for individuals to live authentically in society without the threat of constant surveillance. It is an essential feature of personal relationships and is practiced through selective disclosure. However, *one’s capacity to choose what, when, and how to disclose in a relationship is affected by the structural features of*

*the relation itself*. In this sense, privacy is not simply individualistic in nature and a question of individual control. As Professor Priscilla Regan argues in her brilliant book *Legislating Privacy*, privacy also has a *social value*, a value to society at large, which she defines in three specific ways: it is a *common value* in that all people value some degree of it; it is a *public value* in that it holds value to our common democratic system; and, it is a *collective value* in that “technology and market forces are making it hard for any one person to have privacy without all persons having a similar minimum level of privacy.”[96] From this perspective, the structural effects on personal disclosure do not simply affect individuals—they also affect society.

According to Molm and Cook, social exchange theory “attempts to explain how relations between social actors (both individuals and groups) develop and change, how the structure of networks in which relations are embedded affects processes of interaction, and how processes such as power use and coalition formation lead, in turn, to changes in social structure.”[80] One of the key features of this articulation is the focus on the relation between social actors as the smallest unit of analysis, rather than on the actors themselves. Emerson’s 1972 formulation of SET relied upon operant psychology explicitly in order to facilitate “the development of a theory that emphasized structure rather than individuals’ thoughts or needs,” particularly after Homans and Blau’s earlier efforts were critiqued respectively for their reliance on rational actors and existing norms to explain individual motivations and social structure.[80]

It is this focus on the relation that differentiates the analytic framework of SET from existing theories used in extant privacy research and introduces another way to approach privacy’ analysis. In moving the unit of analysis away from individuals and individual cognition to the relation between actors, SET allows us to consider how the structure of the relation affects disclosure. When one actor practices *selective disclosure* (privacy) as part of the exchange relation, this disclosure is in part dependent on the actor’s *power dependency*: her ability to obtain valued resources from other actors. And one’s capacity to choose what, when, and how to disclose in an exchange relationship is affected by the structural features of the relation itself, including power differentials. Further, the form of the relationship can enable us to make predictions about how exchanges will unfold. Armed with these insights, I argue that examining many of the disclosure relationships that critics call paradoxical aren’t so when reconsidering them as direct negotiated exchanges and taking into account the influence of structural power. An individual decision to disclose personal information to Google’s search engine may appear paradoxical given its inherent risk to one’s privacy, but less so when considering that the company exerts near-monopolistic control over its resource, there are few viable alternatives to it, and the company controls the terms of the exchange.

A key goal in this dissertation is to explore the role of structural power in an information exchange relationship with a corporate actor. While the influence of power on privacy in surveillance contexts is well-documented, its role in online disclosure relationships, particularly in experimental studies, is underexplored. Many researchers assume individuals can or have exercised their power through *consumer choice*: the assumption that people as consumers of products or services can freely decide whether or not to use a specific product or service, and barring any contractual obligations can ‘vote with their feet’ by ceasing to use a service if they are unhappy with it. Another assumption is that both actors hold equal power in the relationship to negotiate its terms, and that each has access to comparable alternatives. My concern is that power differentials have

enough of an effect on disclosure relationships to affect their outcome; that individuals' choices to engage in these relationships are not based on free choice but are affected by power-dependence, and that the structural power advantage that companies and organizations enjoy lend themselves to exchange terms that favor the companies at the expense of individuals.

The SET approach has significant consequences for examining information exchange relationships dominated by winner-take-all actors, such as Google or Facebook. The outsized position of power these actors have in exchange networks allow them to control the terms of disclosure when exchanging with any number of individual actors. And their ability to exert control over the terms of exchange allows the companies to continue to maintain and strengthen their positions of power. Subsequently, privacy as individual control becomes nearly useless in these relationships because individuals alone cannot change the power differential. Instead, as SET indicates, individuals can seek other alternatives or build coalitions to check the power of an exchange partner. In many of the information exchange relationships dominant in our society today, actors such as Google and Facebook hold near-monopoly power over their marketplace sectors. Thus, changing the social structure through coalition building appears to be an obvious option. Regan outlines in depth the civil society coalition building that emerged to successfully help pass the Electronic Communications Privacy Act of 1986 and the Employee Polygraph Protection Act of 1988, noting that “[p]rivacy advocates were most successful in achieving privacy legislation when they reached beyond the privacy policy communities to form advocacy coalitions with other groups.”[96]

While none of the SET literature I reviewed explores coalitions in depth, I will proffer that they should include the many ways in a democratic society by which the public seeks to influence democracy: advocacy groups, civil society organizations, elected representatives, and appeals to government (such as through public comment or through elected officials), to name a few. And what ‘social structure’ might be changed? One can imagine a breadth of formal and informal mechanisms, but regards to this specific problem: the regulatory structure of notice and choice; the legal options available to individuals; the sanctions companies might face for privacy harms; the norms that govern exchange relationships (*e.g.*, from business or technological to social).

Another question must be answered when interleaving privacy and SET: is SET compatible with contextual integrity? I would argue yes, based on the following rationale. First, social exchange theorists repeatedly refer to the influence of norms on exchange relationships. Cook and Emerson specifically suggest that “normative concerns operate as constraints upon the use of power in exchange networks.”[27] Thus, respect for norms as a regulation mechanism on exchange appears uncontroversial. Next, CI’s reliance on context appears to overlap conceptually with SET’s use of social structure and more specifically, the form of the exchange relation. While Nissenbaum doesn’t delve deeply into the power differentials, networks, and hierarchies that may construe a context, there is nothing in her discussions about social context that implicitly or explicitly appears to exclude them. Following her logic, it is possible that the form of exchange (negotiated, reciprocal) could also be considered a context. Finally, applying SET to a CI-based analysis of privacy may in fact help reveal instances where information transmission adheres to existing norms, but those norms reflect an imbalance of power that impinges on social values. Perhaps instead of debating the form of context that might apply, examining the form of information exchange occurring between actors might aid in identifying a power-imbalanced exchange that is normatively accepted

but in fact in conflict with social values.

## 2.2 Motivating the Studies

In this section, I review the specific research that motivates this dissertation and link it to the specific research questions and hypotheses I constructed.

### 2.2.1 Qualitative Study

In order to explore how I could interleave privacy and SET I decided to start by conducting a set of purposive semi-structured interviews. Before moving on to test specific hypotheses using experimental methods, I wanted to conduct an inductive inquiry in which I asked participants questions exploring facets of SET and personal disclosure to get a sense of whether the concepts of social exchange capture what occurs in these relationships. A goal of this approach was to use the language of SET in my questions: did concepts such as *relationships* and *exchange* resonate with participants, or would it be necessary to translate them into more accessible language? I would use these findings to inform the construction of my experimental survey and to help ensure that the survey was ecologically valid—specifically, that the questions in the survey were interpreted by participants in the way I intended.

Another goal was to investigate whether the relationships themselves took one of two forms: *direct reciprocal relationships*, or *direct negotiated relationships*. My hypothesis was that these relationships resembled direct negotiated relationships, but I wanted to capture how participants described them. Independent of the form of the relationship, my other primary hypothesis was that participants were relying upon *assurance structures* to engage in these disclosure relationships (which would make them direct *binding* negotiated relationships, and that they might suggest assurances that I did not anticipate. Thus, I asked participants open-ended questions intended to draw out what, if anything, they relied upon when making disclosure decisions to companies.

#### 2.2.1.1 Interview Contexts

In order to investigate the form of the exchange relationships, I worked to identify two primary contexts where an explicit trade of personal information was made for access to a service. After much contemplation and advice-seeking, I settled on mobile pregnancy tracking applications and direct-to-consumer genetic testing, with a goal of interviewing ten users in each context. My goal in identifying contexts was not to perform case studies of how people use each of the three services, but instead to use services as a site to investigate the applicability of SET to explicit forms of exchange. Both of these contexts involved the explicit disclosure of personal information to the company providing the service. In addition to the two primary contexts, I chose to ask all respondents about their relationship with their primary internet search engine in order to have a single context that I could compare across the entire group of twenty respondents. While not every single disclosure to a search engine necessarily reveals personal information, taken in the

aggregate, one's search queries can both uniquely identify you as well as reveal details about multiple aspects of your life. I also anticipated that examining search engine relationships would contribute an aspect to the research that the other contexts could not: the impact of a long-term (*e.g.*, five years or longer) relationship with a single company.

### 2.2.1.2 Qualitative Research Questions

I developed the research questions below to guide my qualitative work. The qualitative research phase was exploratory work conducted with the goal of understanding what people thought about conceptualizing their interaction with a company as a social exchange.

1. SET and Disclosure: Does SET accurately and adequately capture the disclosure dynamic between individuals and companies? In other words, are individual disclosure relationships with companies a form of social exchange?
2. Assuming SET does accurately describe this disclosure relationship, is the form of these relationships direct negotiated exchanges or direct reciprocal exchanges?
3. Assuming the relationships are direct negotiated exchanges, do participants view them as binding? Do they rely on any form of assurance structures to guarantee their participation?
4. What features of direct negotiated exchange do these relationships include (or not)?
5. To what extent are power differentials a feature of these relationships and how does it affect them?
6. What is the impact of the length of the relationship on disclosure? More specifically, does the duration of a relationship (past the initial sign-up/disclosure phase) have any effect on power balances between individuals and companies?

I discuss how I integrated these research questions into the qualitative study in Chapter 3.

### 2.2.2 Experimental Survey

As I describe above, my primary goal in conducting the interview study was to inductively examine facets of SET and personal disclosure in order to obtain a sense of whether the concepts of social exchange captured the dynamics of these disclosure relationships. Given that there was no extant empirical examinations of this area, I wanted to first understand if social exchange theory provided a framework that fit the respondents' perceptions of these relationships. Given that it did, I then used the findings from the interview study to inform my experimental design, specifically: to refine the experimental manipulation statements, to incorporate assurance structures that emerged from the interviews, and to write and refine the survey questions.



### 2.2.2.1 The Paradox of Control

Of the dozen or so empirical studies Acquisti and others have conducted within the boundaries of BE and information studies, the key study this dissertation builds on is Brandimarte, Acquisti and Loewenstein’s study of what they call the “paradox of control”: that as individual control over disclosure increases, willingness to disclose also increases, and that lower individual control results in less disclosure of personal information.[19] While providing control over personal information “allows one to choose how much to reveal about himself and to whom,” the authors argue that most people respond by underestimating risk and thus disclosing more, jeopardizing their privacy.[19]

The authors make a distinction between the *release* of personal information, *access* to it, and *usage* by others. In two of the three studies, release was operationalized as disclosure of personal information through a survey intended to populate a fictional university-specific social networking site; access was manipulated as publication of one’s personal profile information to the site; and usage was manipulated as one’s profile visibility to either other students at the university or to students and faculty.<sup>1</sup> Given that control is viewed as a central to managing one’s information privacy, the authors argue that the aspect that should have the most logical salience to individual decision-making is control over the usage of the information once it is disclosed. However, given individuals’ bounded rationality and limited attention (due to focusing on, in part, the decision to disclose), the authors hypothesized that many participants would instead focus on the disclosure itself rather than both access and usage of the disclosed information. The study was conducted in three distinct phases using surveys to measure individual willingness to disclose based on participant answers to survey questions of varying intrusiveness. The authors also varied levels of personal control over disclosure. Their findings, they argue, contradict the “conventional wisdom that control over personal information either implies or at most does not negatively affect privacy protection. Our results show that ‘more’ control can sometimes lead to ‘less’ privacy in the sense of higher objective risks associated with the disclosure of personal information. In other words, our results provide evidence that control over personal information may be a necessary (in ethical or normative terms) but not sufficient condition for privacy protection.”[19] Thus, despite the public’s desire for more control, and a policy response focusing on the same, paradoxically Brandimarte *et al*’s findings suggest that offloading privacy choices from regulators or companies to individuals may actually expose the public to higher potential harm.

These findings play a crucial role in this dissertation because they provide evidence that privacy decision-making may not follow a deliberative privacy calculus, and instead is subject to both instability and malleability, as the same authors argued in their literature review in *Science*. It is when we adhere tightly to the rational actor model, which the privacy calculus relies upon, that privacy choices and behaviors may appear paradoxical and in conflict. While SET is similarly based on assumptions of rationality, as discussed earlier in this chapter, the theory makes no assumptions about individual cognition. Thus, even though the focus of this dissertation is not on individual cognition, Brandimarte *et al*’s findings provide an opening for one of my hypotheses:

---

<sup>1</sup>The third study did not use a social networking site as the disclosure location. Instead, the participants were given a ten question survey on ethical behavior and told that their anonymous responses would be published in a research bulletin with no specified audience.

that the structure of the relationship, and not just one's privacy attitudes or experiences, may affect personal disclosure.

### 2.2.2.2 Experimental Research Questions

I used my results from the qualitative interviews to help refine the experimental survey, which was developed based on theoretical concepts in SET and privacy theory. The hypotheses specifically test the impact of framing personal disclosure as either mandatory or optional on a set of theoretically derived variables.

1. When framing the disclosure relationship as a direct negotiated exchange, which form of negotiation will report higher levels of trust, fairness, and benefit: an optional disclosure negotiation, or a mandatory disclosure negotiation?
2. Which form of disclosure negotiation will respondents report meets their privacy expectations? Will respondents disclose more personal information when they report the relationship as meeting their expectations?
3. Will adding assurance statements to these negotiation framings increase respondents' reports of trust, fairness, benefit? Will assurances increase personal disclosure or privacy expectations?

I developed the hypotheses below based upon these research questions.

## 2.3 Study One: Hypotheses

Decades of laboratory experiments conducted by social exchange theorists have provided both the empirical evidence to support SET's core propositions as articulated by Emerson [39] [38], as well as expanded elements of the theory beyond its core. They have fleshed out the core forms of social exchange—reciprocal and negotiated exchanges—and identified their core features, including: the role of trust, perceptions of fairness, and power. This experiment builds on this body of research by replicating elements of it in a new context, by using information as the valued item of exchange between an individual and a company (rather than currency, which Molm and Cook note is the common resource used in lab studies).[80]

In this study, I explicitly manipulate the influence of power as optional versus mandatory personal disclosure as a form of negotiation, and observe the effect of this manipulation on participants' perceptions of trust, fairness, power, privacy, and their willingness to disclose their personal information. Researchers have demonstrated that providing individuals with more control over personal disclosure may paradoxically lead individuals to increase their levels of disclosure.[19] This study takes a different tack by examining the difference in individuals' personal disclosure and perceptions of privacy when explicitly manipulating the level of autonomy, or control, one has over disclosure within the context of an exchange relationship with a company. By doing so,

I explicitly introduce the influence of power, operationalized as negotiation, into the decision to disclose. The optional disclosure condition gives the participants autonomy over the information they disclose, presenting an exchange relationship where they possess more negotiative power over the terms of the exchange. The mandatory negotiation condition removes personal autonomy from the participants by requiring them to disclose all the information the company dictates, creating an exchange relationship where they lack negotiative power over the terms of the exchange.

In addition to examining the influence of the experimental manipulations on personal disclosure and perceptions of privacy, I also examine participant perceptions of trust, fairness, and power in the exchange relationship.

### 2.3.1 Trust

I deliberately exclude explicit assurances in Study One. Instead, I draw on Yamagishi and Yamagishi's definition of trust as expectations of benign behavior based on inferences about a partner's personal traits and intentions, I predict the participants in the optional disclosure condition will report higher levels of trustworthiness and encapsulated trust than respondents in the mandatory negotiation condition.[120] To the extent that trust is based on an expectation of benign behavior, a condition where one is constrained from exerting autonomy over personal disclosure is likely to be perceived as less benign.

**H1a: respondents in the optional disclosure condition will rate the company as significantly more trustworthy as compared to respondents in the mandatory disclosure condition.**

**H1b: respondents in the optional disclosure condition will rate the company as significantly more trustworthy with their personal data as compared to respondents in the mandatory disclosure condition.**

### 2.3.2 Power

As discussed in Chapter 1, power is a core feature of exchange relations. Actors' mutual dependence forms the basis of power in SET: "power is a potential that derives from the structural relations among actors—their relative dependence on one another." [80] In this study, I measure the participants' perceptions of individual power and control over the terms of the relationship. The power dependency between these actors is formulated in terms of the exchange of information for benefits: the company requires personal information from customers in order to develop its product. In exchange, they offer benefits derived in part from that information to their customers: greater personal insight into one's health and fitness. However, because companies in the U.S. have greater structural power than individuals, they are generally able to dictate the terms of many exchange relationships. This structural power is derived not only from a company's size relative to an individual, but from their economic resources as well as from structural benefits embedded in the law, specifically the ability for companies to create 'take-or-leave-it' contracts, such as Terms of Service agreements and 'clickwrap' agreements. Thus, negotiation in these negotiated exchanges often consists of the individual's mere acceptance or rejection of non-negotiable terms.

Here, I predict that participants in the optional disclosure condition will report having greater power and more control over the terms than those in the mandatory disclosure condition. Despite how common ‘take-or-leave-it’ terms are in technology usage, my own past research suggests that users are often dissatisfied with them.[59]

**H2a: respondents in the optional disclosure condition will rate themselves as having significantly higher measures of individual power in the relationship as compared to respondents in the mandatory disclosure condition.**

**H2b: respondents in the optional disclosure condition will rate themselves as having greater control over the terms of the relationship as compared to respondents in the mandatory disclosure condition.**

### 2.3.3 Fairness

According to Molm *et al*, “the form of exchange has strong and consistent effects on actors’ perceptions of fairness.”[80] While one might expect that the process of joint negotiation would lead to greater perceptions of fairness than a non-negotiated exchange, based on a lab experiment contrasting perceptions of fairness between negotiated and reciprocal exchanges, Molm and her colleagues found that negotiated exchanges were actually perceived as more unfair than reciprocal exchanges. They suggested this was due to the fact that the negotiation process itself highlights conflict between actors, even under conditions where the outcomes of the exchange were equivalent.[80] In later work, Molm and colleagues identified three features of negotiated exchange that contribute to perceptions of unfairness: “the establishment of inequality within bilateral transactions rather than across sequential transactions, the partner’s more active and unambiguous role in producing inequality, and a more direct and transparent relation between one actor’s cost and another’s benefit.”[80]

This study contrasts two forms of negotiation, one of which features an offer that directly benefits the company at the expense of the individual, explicitly establishing inequality between them. I predict that the participants’ perceptions of the fairness and benefit of the exchange will be higher in the optional disclosure condition, where the individual retains autonomy in the relationship and the benefits appear to be shared (at least in this bilateral transaction).

**H3a: respondents in the optional disclosure condition will rate the fairness of the exchange significantly higher as compared to respondents in the mandatory disclosure condition.**

**H3b: respondents in the optional disclosure condition will rate their benefit significantly higher as compared to respondents in the mandatory disclosure condition.**

### 2.3.4 Privacy

As reported by Brandimarte *et al*, when individuals are given increased control over disclosure, their willingness to disclose personal information increases.[19] However, there is no experimental research examining the influence of greater or lesser control over disclosure on one’s expectations of privacy. Given the negative light in which loss of control is discussed, I predict that providing more control over disclosure will not only lead to an increased likelihood to disclose, it will also

lead to higher ratings of the company's privacy practices as well as their ability to meet one's privacy expectations as compared to respondents in the mandatory disclosure condition.

**H4a: respondents in the optional disclosure condition will rate the company significantly higher in meeting their privacy expectations as compared to respondents in the mandatory disclosure condition.**

**H4b: respondents in the optional disclosure condition will be significantly more likely to disclose their personal information to the company as compared to respondents in the mandatory disclosure condition.**

**H4c: respondents in the optional disclosure condition will rate the company's privacy practices significantly higher as compared to respondents in the mandatory disclosure condition.**

## 2.4 Studies 2A and 2B

Study Two (collectively two studies, 2A and 2B) investigates the effect of assurances on the disclosure relationship between the individual and the company introduced in Study One. With the exception of the effect of privacy policies, assurances are an underexplored concept by privacy researchers. There is related work exploring the impact of interface cues on decision-making related to privacy, but this focuses on specific interface implementations and not conceptual assurances.[64][89][111][36]

As discussed earlier, assurances mitigate risk and uncertainty by providing a third party mechanism to assure a negotiated exchange. While there is extant research within the SET field validating the role assurances play in negotiated exchange, to the best of my knowledge there is none that explores the effect of different types of assurances on negotiated exchange. This study utilizes two types of assurances (formal and informal) and seeks to measure whether they have any effect on a subset of dependent variables from Study One most likely to be affected by concerns with risk and uncertainty: trust and disclosure.

### 2.4.1 Hypotheses

While assurances are typically understood to supplant trust in negotiated exchanges, some level of trust must still be present in order to facilitate exchange. I predict that respondents in any of the three assurance conditions will give higher ratings of general trustworthiness and trust the company's handling of their personal data as compared to the control. Comparatively, I predict that the the legal assurance condition will be rated significantly higher than either of the two assurances in H1a, and the anonymity assurance condition will be rated significantly higher than either of the other two assurances in H1b. While I would anticipate that the legal assurance, due to its sanctioning power may provide the highest general assurance, in this instance I would expect the anonymity assurance to perform more strongly as this assurance specifically provides protection to personal data.

**H1a: respondents in any assurance condition will rate the company as significantly more trustworthy as compared to respondents in the control condition.**

**H1b: respondents in any assurance condition will rate the company as significantly more trustworthy with their personal data as compared to the control condition.**

**H1c: respondents in the anonymity assurance condition will have significantly higher ratings than the legal and social assurance conditions.**

Decisions to disclose one's personal information are affected by risk. Thus, in general situations that pose higher risk to individuals' privacy should in turn discourage personal disclosure. I predict that respondents in any of the three assurance conditions will be more likely to perceive more control over their disclosure of their personal information than the control condition. Comparatively, I predict that the anonymity assurance condition will be rated significantly higher than either of the other two assurances in H1a as this assurance specifically provides protection to personal data.

**H2a: respondents in any assurance condition will be significantly more likely to rate themselves as having greater control over the disclosure of their personal information to the company as compared to respondents in the control condition.**

**H2b: respondents in the anonymity assurance condition will have significantly higher ratings than legal and social assurance conditions.**

I review the methods I used for these experiments in Chapter 5, and the results in Chapter 6.

# Chapter 3

## Qualitative Methods

In this chapter I describe the qualitative methods used in this dissertation to explore the disclosure of personal information to companies across three contexts: mobile pregnancy tracking apps, direct to consumer genetic testing, and search engines.

### 3.1 Study Overview

The qualitative study was an inductive exploration into privacy and personal disclosure using social exchange theory as the primary theoretical framework. The goal was to examine a specific form of exchange relationship: individuals choosing to disclose personal information to companies in exchange for a product or service. I argue that this type of exchange relationship is a *direct negotiated exchange*.

I conducted a small (20 participants) interview study where I focused on two primary contexts within the field of personal health (direct to consumer genetic testing and pregnancy tracking apps) and one general informational context (online search). I recruited participants based on their membership in one of the two primary contexts, and asked all the participants about the secondary context. I used these two specific informational contexts, both of which rely on the disclosure of highly sensitive personal information, as an entry point to delve into: individual perceptions of the form of the information exchange; participants' perceptions of the risks and benefits of the relationship; and whether and which assurance structures they viewed as mediating risk and uncertainty with regards to disclosure, among other topics. The study was approved by the U.C. Berkeley Office for the Protection of Human Subjects (#2015-12-8223).

#### 3.1.1 Study Context

In this section I review the three contexts that I used as the basis for this study.

### 3.1.1.1 Pregnancy Tracking Apps

Pregnancy tracking apps allow women to typically track multiple features of their pregnancies. The universal minimal disclosure required is one's due date, as the content the apps display are pegged to the user's current week of her pregnancy (out of a typical 40 weeks), such as the stage of the fetus's development, typical body changes, and so on. The types of features available vary, from apps focused primarily on delivering content, to those that facilitate participation through community (such as message boards and social media sharing functions). Some apps only collect a minimal amount of information (such as due date and one's email address to register for an account), while others allow for the user to optionally enter in a range of details, such as weight gain, health data (blood pressure, blood type, etc.), doctor's appointments, contraction or kick monitoring, and others.

While one's pregnancy (typically) by the end becomes impossible to hide from those who interact with you in person, exerting personal control over the process despite its visibly public outcome is important to women for a multitude of reasons: the potential for miscarriage, questions of paternity, fertility struggles, risks for discrimination, and most fundamentally, the desire to control one's own experience and body without outside interference. Most women plan for publicly acknowledging their pregnancies at some point in the process, but would like to control the terms by which they do so, particularly the window between when they first learn they are pregnant and when they choose to share their status with others. Thus, when a woman downloads a pregnancy app seeking help and advice with her pregnancy, particularly in the early stages, she might expect that an app would not share her status with anyone (individual people or 'the internet' at large) without her explicit permission. Unfortunately, this is not always the case—the same apps women turn to for advice about their pregnancies may also be the source of their disclosure to a network of advertisers and information brokers, all in pursuit of one of the most valuable marketing leads available.

Prior to recruiting participants, I conducted an analysis of the most popular mobile pregnancy apps in the Google and Apple mobile app stores, and based on their features and business practices (*e.g.*, whether they required creating an account, whether one could share personal information on their platforms with other users) I compiled a list of qualifying apps. While pregnancy tracking apps abound in both the Google Play and Apple app stores, there are approximately ten apps that dominate market share in the category. I installed each of the top apps to a tablet device and reviewed each to ensure that disclosure of personal information was required to use the app. After reviewing them, I settled on the following list of seven qualifying apps:

- Glow Nurture
- BabyBump Pregnancy Pro
- What to Expect Pregnancy Tracker
- My Pregnancy and Baby Today by BabyCenter
- Ovia Pregnancy



- Pregnancy+
- I'm Expecting Pregnancy App

I took this step to ensure that the apps were not outliers with small user populations, and that they did clearly engage in the core exchange practices I was studying (*i.e.*, either requiring or requesting that users to disclose some form of personal information to use the service). For example, all the pregnancy apps request that the user disclose her due date in order to provide timely content. However, each app allowed users to contribute many additional types of data about their pregnancies, and some had specific features, such as sharing milestones to social media platforms, that encouraged user disclosure of personal information outside the app.

To participate in the study, participants had to have used at least one of these seven apps for a minimum of six months. In cases where a participant used multiple apps, I asked her to identify the app she used the most often, and made that app the focus of the interview.

### 3.1.1.2 Direct-to-Consumer Genetic Testing (DTCGT)

At the time I began the project, within the DTCGT service category there were only four commercial companies offering this service: 23andMe, Ancestry.com, FamilyTreeDNA.com, and National Geographic's Genographic Project. I conducted informal pilot interviews with users of these various services, and I ultimately decided on narrowing the scope of the DTCGT study to 23andMe. This was a difficult decision, as doing so made the study dependent on the design and strategic decisions of a single company. However, my pilot interviews were helpful in discovering that users of Ancestry, FamilyTreeDNA, and the Genographic Project were primarily interested in tracing their ancestral roots, while 23andMe users were often interested in both their ancestry as well as the genetic health information that only 23andMe currently offers to consumers. Because ancestral data most commonly involves the deceased, privacy concerns with ancestral data were limited (as I quickly realized, most of the dead have little privacy interest in their ancestral data). However, the medical conditions that the 23andMe DNA analysis exposes did carry with them both current and future privacy concerns for respondents (as well as, potentially, both current living and future relatives). Thus, after the pilot interviews I chose to only recruit 23andMe users for interviews.

### 3.1.1.3 23andMe Company Background

23andMe was founded in 2006 and launched its testing service a year later. The company's growth has been rapid as of late; the number of customers passed the 100K threshold in 2011, the 1MM threshold in 2015, and over 2MM in 2016. Based in the Bay Area (Mountain View), one of the company's co-founders was married to a Google co-founder, and Google was an original investor in the company. As such, several participants were aware of an existing relationship between the two companies (though most couldn't accurately recall the details). Some local connections also existed among participants—one participant noted that a friend worked for the company; another participant, a former Google employee, said that Google employees were offered a discount to sign

up with the service while she was employed there. However, I did not disqualify participants based on these relationships as I did not think they presented any risk to my research goals.

23andMe has had a tumultuous relationship with the Food and Drug Administration, the federal regulatory body that oversees the company, regarding how they provide results to customers about inherited diseases. After providing detailed results to consumers regarding their probability of developing specific inherited diseases, the company was required in 2013 to stop providing consumers with inherited disease results, limiting them to only providing ancestral data.[93] In 2015, the company was allowed to provide inherited health results again, albeit in a more restricted form than previously, and for fewer genetic conditions. In April 2017, the FDA gave the company explicit approval to provide users with genetic risk data for ten specific diseases.[113] The 2013 cease-and-desist was widely covered by the media and 23andMe communicated with its existing users about the change. As such, I asked all 23andMe users if they were familiar with the event and how it both impacted them directly as well as whether it had any impact on their relationship with the company.

#### **3.1.1.4 Genetic Testing Process**

In order to obtain DNA sequencing results from 23andMe, customers purchase a DNA collection kit (most recently priced at \$99-\$199, depending on when it was purchased). The primary exchange in the 23andMe users' relationship is initiated during the sign up phase through the submission of a DNA sample to the company—a vial of one's saliva. After receiving a testing kit in the mail, new customers are asked to sign up for an online account and fill out an online survey disclosing detailed information about one's health history and family background. Though users are not required to disclose any additional information after this exchange, they are incentivized to continue both their relationship with the company and additional disclosure through: ongoing survey participation, new research studies, and if the participant has opted into the company's genealogy platform, potentially new family member matches.

Approximately six weeks after they return their sample, customers are contacted with their results via email, which they view on the company's website. Customers are shown information about their inherited traits and diseases, as well as their ethnic composition. If one wishes to share their genetic information on the company's platform to identify potential relatives, they can opt-in to do so using the company's DNA Relative Finder service; they can also opt-in to compare results with immediate family members if they know any to have also submitted a sample. 23andMe also partners with research universities, pharmaceutical companies, and non-profit organizations and has received federal grants from the National Institutes of Health to use their customers' genetic data for research projects. The research is based on a combination of DNA and self-reported data; when 23andMe customers opt-in to the research component, they are asked to complete short, discrete surveys ('Quick Surveys') on an ongoing basis. Their answers, along with their anonymized DNA samples, are made available to the company's research partners for associative analysis. The company's research studies are overseen by an independent Institutional Review Board, and customers can withdraw their consent at any time (though not from studies in which they have already participated). If a user elects to opt-in to contributing to research studies on the

platform, as the majority of these participants did, then they are actually involved in two exchanges: a primary direct exchange with the company that directly benefits themselves, and a secondary indirect generalized exchange with researchers that indirectly benefits them.

## 3.2 Participant Recruitment

I conducted twenty interviews over the course of five months. Participants were recruited from two local online platforms, each requiring a free account: the Berkeley Parents Network newsletter<sup>1</sup>, and the NextDoor.com website within the city of Berkeley. Recruitment ads were posted multiple times on each network over the five-month period. I used a convenience sample, though I made an effort to diversify within the sample with regard to age, ethnicity and gender<sup>2</sup>. Potential participants were screened with a questionnaire that asked them to provide their: age; gender; education level; ethnic group affiliation; primary search engine and length of use; and contact and scheduling information. Pregnancy app users were asked to indicate which app(s) they used (from my list of prequalifying apps) and for how long. 23andMe users were asked how long they had been using the service, and which specific service features they used (*e.g.*, health and/or ancestry). Approximately sixteen people replied to the ad for pregnancy app users, and twenty-six replied for the ad for 23andMe. The final respondent pool was highly educated (from two years of college to advanced degree holders), ranged from 18-29 to 60-69 years of age, and skewed female due to half the sample consisting of pregnant or recently postpartum women.

I contacted participants via email if they fit the study criteria, and I made an effort to find a mutually convenient location and time to meet (on campus, coffee shops, and even participants' homes). Participants were paid \$30 via an anonymous gift card upon completion of the interview, which averaged about an hour. No participants withdrew from the study after agreeing to the interview.

Apps Used	Ovia (5), What to Expect (3), BabyCenter (4)
Time Used	Under 1 year (7), Over 1 year (3)
Ethnicity	White (6), Asian (3), Other (1)
Education	College (2), some graduate school (1), master's (5), adv. degree (2)
Age	18-29 (2), 30-39 (7), 40-49 (1)

Table 3.1: Summary of Mobile Pregnancy Tracking App Participants

<sup>1</sup>[www.berkeleyparentsnetwork.org](http://www.berkeleyparentsnetwork.org)

<sup>2</sup>All pregnancy app participants identified as female

Time Used	Over 5 years (5), Under 5 years (5)
Ethnicity	White (8), Asian (1), Hispanic (1)
Education	Some college (1), 4 year college (3), some graduate school (3), master's degree (2), adv. degree (1)
Age	18-29 (3), 30-39 (3), 50-59 (1), 60-69 (3)
Gender	Female (6), Male (4)

Table 3.2: Summary of 23andMe Participants

The 23andMe respondents, as they were not limited to childbearing years like the pregnancy app users, varied more widely in age, with six under the age of forty and four over the age of fifty (I could not locate a respondent in the 40-49 age bracket).

### 3.2.1 Search Interviews

I did not recruit participants on the basis of search engine use, but instead asked them to identify their primary search engine and length of use during the screening process. During the interview intake, I verified their choice of search engine by asking how they typically accessed it (desktop, mobile) and confirming that they were making a deliberate choice to use it (versus simply using the default on whatever platform or browser in use). Nineteen of twenty respondents identified Google as their primary search engine (the one exception was a Yahoo! user), with the majority having used it for ten or more years. Proportionally, this skewed sample is roughly in line with global search engine usage numbers, which show Google as having a global market share of over 80 percent as of July 2017.<sup>3</sup>

Coincidentally, one of the 23andMe interviewees happened to be a former Google employee, so I did not ask her about her search engine use as she made it clear that she possessed both insider knowledge of Google Search and was favorably biased towards the company. Thus, the search portion of the qualitative analysis consists of nineteen participants instead of twenty.

### 3.2.2 Interview Instrument

I developed the interview instrument over many months, starting first with identifying the theoretical concepts from SET and privacy that I wished to examine. The instrument began as an outline

<sup>3</sup>According to NetMarketShare, Google controlled 83% of the global search market in July 2017. <https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0>. StatCounter marked the US share at 87% as of July 2017. <http://gs.statcounter.com/search-engine-market-share/all/united-states-of-america>. Both links visited August 18, 2017.

Search Engine	Participants	Time Used
Google	18	15+ years (4), 10+ years (11), 5+ years (3)
Yahoo!	1	15+ years (1)

Table 3.3: Search Engine Participant Summary

of concepts and questions I was interested in exploring with participants: relationships; privacy; trust; assurance structures; and the development of the contexts in which I wanted to ground my inquiry.

The instrument underwent several revisions, as I worked on translating it from a set of concepts into both open-ended and specific questions. The final instrument was organized around the following themes:

- Context-specific questions (pregnancy, DTCGT, search queries) that asked about specific areas related to each context (*e.g.*, “Why did you decide to track your pregnancy?,” “How did you decide on 23andMe?”);
- SET theoretical concepts: questions exploring features of the relationships, assurance structures, benefits, trust, risks, and power;
- Privacy theoretical concepts: mechanics of personal disclosure, assessments of privacy harms and threats (*e.g.*, tracking, targeting, embarrassment, exposure), the participants’ object of privacy concern (*e.g.*, insurance companies? Friends and family?), assessments of information sensitivity, risks.

Within each thematic section, there were one to three mandatory questions I asked of every participant. Depending on a participant’s answers, I chose follow-up questions from a list I had prepared for each section. I organized the interviews as follows: context specific questions (pregnancy apps, 23andMe), relationships, free vs. paid dimensions, personal information disclosure and trade-offs, privacy/trust/risk, assurance structures and uncertainty, and closing with search query questions (which included repeats of both the aforementioned context-specific questions as well as the mandatory questions from the previous sections).

### 3.2.3 Contextual Questions

I started every interview with a lead question about the primary context, and then transitioned to a series follow up questions about how and why they chose their service, their goals, their general impressions of the company or service, and any impressions they had about the app or website specifically. I also asked context-specific questions, such as to whom, when, and how the pregnancy app users revealed their pregnancies, and whether the DTCGT users had any other forms

of genetic testing. My goal was to have the participants begin the interview by focusing on the app and their decision-making process before shifting to questions about their relationship with the app. I also wanted to document the participants' general thoughts and impressions about the app or service and before I began asking them more specific questions about their relationship.

**Pregnancy Context:**

- Lead Question: Let's start first with why you decided to track your pregnancy/period, and what you wanted to accomplish. Can you describe your decision-making process, and how you chose the app you use?
- What made you decide to track your pregnancy?
  - How did you choose this app?
- When and how did you disclose your pregnancy? (a) immediate family b) friends c) others (co-workers, etc.) (*Phone, in person, email, social media?*)
- When did you first hear about them, and what was your first impression?
  - What were your impressions based on?
- How did their website or app make you feel about them?
- What did you like/dislike about it?

**DTCGT Context:**

- Lead Question: Let's start first with how you came to use genetic testing, and why you wanted to have it done. Can you describe to me your decision-making process, and how you chose the service you used?
- What made you decide to have genetic testing performed?
- When did you first hear about 23andMe, and what was your first impression?
  - What were your impressions based on?
- How did their website or app make you feel about them?
- What did you like/dislike about it?
- Have you ever had any other form of genetic testing?
  - Did you consider any alternatives, such as getting tested through your doctor by a testing company?

Because I anticipated that all of the participants had likely been using a search engine for many years, I did not ask them extensive background questions during the interviews. Instead, as part of the written intake survey I asked respondents to answer the following questions:

**Search Context:**

- Name of primary search engine used and estimate of length of use
- Where do you access search the most often? For instance, on a smartphone? A tablet? A desktop computer?
- For the search engine you use the most often:
  - Is this a search engine you choose, or was it the default on your device/browser?
  - Do you have an account with this search engine (do you login to use it)?
  - Are there searches you would not perform on this search engine?

I reviewed the respondents' search answers prior to beginning the search portion of the interview, and based on their answers asked them additional questions as needed. My goal was to make sure we were discussing the search engine that they both used the most often and intentionally chose to use, not just whatever was the default on their phone or in their desktop browser. Thus, if there was a relationship to explore, it would be with the search engine they intentionally used. After reviewing the intake information, I conducted the remainder of the search engine portion of the interview following the outline of the pregnancy app/genetic testing contexts, asking the same core questions for the search portion. While this portion of the study was not intended to be a study of Google specifically, as noted in the methods section, 19 of the 20 participants listed Google as their primary search engine. Thus, many of the quotes reference Google directly in the discussion.

### 3.2.4 SET Thematic Questions

The goal of this portion of the interview was to use theoretical concepts from social exchange theory as a basis for exploring how participants perceived their disclosure relationship with the company (for their use of pregnancy apps or genetic testing, as well as their primary search engine). I was curious: how would the participants describe their relationship with the company? Would those relationships more closely resemble a negotiated exchange, or a relationship based on reciprocity? I used this portion of the interview to probe these concepts in depth.

I began the SET thematic section by asking respondents to describe the nature of their relationship with the company. It is this exchange relation between actors that comprises the primary unit of analysis in SET. I wanted to document how they described the core exchange: what were they giving the company, what were they receiving in turn? I then asked about structural features of the exchange: fairness (how fair they felt the exchange was to them as compared to the company); the benefits for each member of the exchange (and what benefit they thought they gained); aspects of trust (whether they trusted the company); and power (who they thought had more power in the

relationship). I also asked questions intended to probe whether the relationship featured aspects of negotiated or reciprocal relationships, and what assurance structures or institutions they relied on, if any, in their decision to disclose their personal information the company.

### 3.2.4.1 Relationships

To explore the participants' relationships with their companies, I asked every participant the following primary questions, with a list of potential follow-ups based on the participant's responses (not all participants were asked each follow-up question):

- Describe for me the nature of your relationship with Company X.
- For example, what do they offer you, and what do you offer them?

In addition to the primary question ("Let's focus on what you give them. What do you get out of your relationship?"), I drew from a list of potential follow-up questions based on their initial response:

- Can you describe for me what you give them?
- What do you get out of the relationship?

Almost none of the participants struggled with the lead question—most were able grasp immediately what I meant by asking about their relationship, and the few who asked for clarification understood once I added "what do you offer you, and what do you offer them."

### 3.2.4.2 Fairness and Benefits

After asking participants what they felt they offered the company and what they received in return, I focused more specifically on the benefits of the exchange and how fair they thought the exchange was. Did the participant feel she was treated fairly? Did she feel as if she had more or less power in the relationship as compared to the company, specifically in terms of access to comparable alternatives, or whether she could exit the relationship?

#### **Free products/apps (pregnancy apps, online search):**

- How fair is it that you get this product for free?
- What benefit is the company getting?
- How does getting this product for free make you feel towards the company?
- How obligated do you feel towards the company?
- Would you consider paying for [your app/search engine]?



**Paid product (genetic testing):**

- Would you have used the service if it were free?
- Was the price fair for what you provided and received in return?

**3.2.4.3 Reciprocity and Negotiation**

Based on participants' responses to the initial questions about their relationship and its benefits, I followed up with questions that probed for different features of reciprocal exchanges (affective bonds, depth of personal commitment, viewing the company more as a partner or person), negotiated exchanges (ease of exit, viewing the commitment based on terms/negotiation), and features related to power (whether they felt as if one of them had a power advantage, other options available to them, information asymmetries), and finally whether the relationships had evolved over time. The goal was to obtain a sense of whether these relationships had features that mapped to reciprocal exchanges or direct negotiated exchanges.

- Who benefits more from this relationship: you or them? Or do you feel it is equal?
- Do you think of the company as a partner or a friend?
- If the company/app was a person, how would you describe them? What is their personality like?
- How do they treat you? And how do you treat them?
- How easy or hard would it be for you to stop the relationship?
- Do you feel like equals? Or does one of you have more advantage/power?
- Do you feel as if you know as much about them as they know about you?
- Do you feel as if you have a personal commitment to the company?
- Is this a relationship you want to continue with in the future?
- Has anything changed in how you thought about the service after you started using it as compared to how you felt when you signed up/first started using?

**3.2.4.4 Assurances**

I asked the participants a series of open-ended questions intended to probe their thinking of assurance structures—what concepts of assurance did they have, if any? What types of assurances were they inspired to mention? I asked about their general level of comfort with using their service as a means to draw out this concept, with follow-ups asking about specific factors that might have offered them protection or a guarantee, as well as to whom they might turn if the company took an

action that angered them, was unfair, or violated the agreement they felt they had. Based on their initial response, I followed-up with more targeted questions, particularly if they expressed that they had no idea, or gave a suggestion in one broad category (*e.g.*, company reputation) but not others. The goal was to elicit as many potential assurances as possible without being overly leading. However, because I anticipated that most participants would likely not include legal assurances in their answers, I had a follow-up question to specifically probe whether participants were aware of any laws that applied to the information they disclosed. I also asked participants directly about privacy policies in order to determine whether they played any role in their decision-making process.

As another way to investigate assurance structures, I asked the participants the following question: “If [company/service/app] did something that you felt violated your agreement with them, angered you, or felt unfair or illegal, what would you do?” Given that assurances are a mechanism for engendering trust between two parties, I reasoned that asking participants who they would turn to if something went wrong with the exchange relationship might yield additional insights. And indeed, while the responses to my question regarding factors yielded responses that focused primarily on mechanisms for signaling trust, reputation (of the company directly, and social reputation), and design factors (both signalling trust and for allowing individual control over information), the responses to the violation question focused primarily on institutions: the company, platforms, third party agencies, government, and the legal system. Thus, assurances fell into one of two broad categories: mechanisms, and institutions.

**Lead Questions:**

- What made you comfortable/trust sharing your data with the company/app?
- Are there any other factors that made you comfortable using this company/app?

**Follow-Up Questions:**

- For example, is there anything specific that made you feel safe, or offered you protection or a guarantee?
- If [company/service/app] did something that you felt violated your “agreement” with them, angered you, or felt unfair or illegal, what would you do? Who would you complain to about [company/service/app]?
- Do you recall if you saw a privacy policy for [company/service/app]? IF YES: did you read it before you signed up?
- Were there any aspects of the policy that informed your decision to use the service?
- Do you know if there are any laws that protect the information you provided to [company/service/app]?

### 3.2.4.5 Personal Disclosure

I began this section asking every participant a question focusing on their personal disclosure to the company: What kind of personal information did you provide to the company/app? Based on their responses, I asked follow-up questions from the list of subtopics below. I composed these follow-ups using concepts from social exchange theory to inform my exploration into the relational aspects of personal disclosure. Again, not every follow-up question was asked of every participant.

#### Follow-Up Questions:

- Have you shared your information on the company's/app's platform? (if applicable)
- How has your interaction with the company affected your expectations?
- Has there been anything unexpected in your interactions with company/app?
- Is it a 1-to-1 relationship? Meaning, is information is going directly to company/app and not anywhere else?
- Do you feel as if there are aspects of the service that you can opt out of or just not use? If yes, what aspects?
- Do you think you can provide company/app with feedback or have them make a change that would better suit your needs?
- Have you ever contacted their customer service for any reason?

When I asked participants about their search queries, I used an abbreviated version of the questions listed above. After asking them what personal information they provided to their search engine, I asked the following search-specific questions:

- Is there any type of information you don't use search engines for?
- What do you think your search engine does with your search information?
- Have you ever looked to see if you can find a history of all of your past searches?

### 3.2.4.6 Risks and Trade-Offs

After discussing the personal information participants disclosed to the company or app, I transitioned to a set of questions probing issues of risks and trade-offs. Risk is an ever-present issue in both privacy and social exchange theory. Assessments of whom to trust, as well as the potential risks of disclosure are key aspects feeding individuals' privacy concerns and decisions. In social exchange theory, a key aspect underlying relationship formation is the reduction of risk and uncertainty that exchange relationships provide, and the trust required between exchange partners for

relationships to form and progress.<sup>4</sup> Questions investigating trade-offs were intended to identify participants' perceptions of the compromises they felt they made to accommodate the company or app. Finally, I asked contextually specific information disclosure questions to gauge participants' assessments of risk.

**Lead Question:** What are the risks to you with using company/app, if any? What is at stake for you?

**Follow-Up Questions:**

- Since you first signed up with company/app, do you trust them more or less now, or the same?
- What, if anything, would make you stop using company/app?
- Does how they handle your personal information meet your expectations?
- Who else do you trust with the type of information you provide to company/app?
- Do you know of any settings or controls that enable you to manage how your information is used? [Privacy settings, others]
- Are there any downsides to using company/app?
- Have you shared your results with your doctor or other health professional? [genetic testing]
- Would you want your health insurance company to have the results/your data? [genetic testing/pregnancy apps]
- Do you share more or less with the app than you do with your doctor/midwife? [pregnancy apps]

### 3.2.5 Interviews

The interviews typically lasted one hour and all interviews were audio recorded for transcription purposes. Participants were provided a CPHS approved consent letter and asked to sign a consent form. Each participant was assigned a participant number, which was recorded on the paper-based intake surveys. These surveys did not collect any identifiable information. Payment was made to participants upon completion of the interview. I conducted the majority of the interviews alone; for four interviews I was joined by a research assistant, Brandon Shalchi, who was included on my CPHS protocol.

---

<sup>4</sup>Some of the discussion of trust is captured in the section on Assurances.

### 3.2.6 Data Analysis

All of the interviews were transcribed by Landmark Associates and entered into the Atlas TI software program for coding purposes. I used structural coding as described by Saldaña, a coding practice that “both codes and initially categorizes the data corpus to examine comparable segments’ commonalities, differences, and relationships.”[100] I developed the coding scheme by mapping the core concepts of the interview protocol into families of specific codes. The first round of coding was performed by a research assistant, Rena Coen, using both a primary interview I coded myself as an exemplar and background readings I assigned to her to familiarize her with SET and qualitative coding.

Coding was a collaborative process. Research assistant Rena Coen coded nineteen of the twenty interviews, and noted questions within the interviews as she proceeded. I reviewed all of her codings, and we worked together to resolve questions or inconsistencies. We also developed emergent codes based on themes that arose during the coding process that the protocol-based codes did not capture. An example of this were specific assurances that participants suggested in the interviews. As a goal of this work was to investigate participants’ conceptualizations of assurance structures, we added codes for each type of assurance as they appeared in the transcripts.

After the coding was complete, I generated reports for each code from Atlas TI, which included specific quotes identified by participant number. I exported each report into a single document, and then made groupings of the reports based on themes (*e.g.*, all codes related to SET, all codes related to privacy, etc.).

I then assigned code groupings to my team of three research assistants (Rena Coen, Matthew Nagamine, and Qing Huang) to produce summaries capturing all of the themes present within each code, as well as tracking their relative frequency. We accomplished this by tallying the number of distinct participants explicitly tied to each theme. The goal was to attach relative weights to each theme in the analysis in order to highlight which themes were more widely expressed. Themes with less support, including an occasional outlier, were included in the analysis but were indicated as such, so that their influence would not be overstated. Given the size of this sample, it is important to be cautious in attaching too much emphasis to any single theme. Because it is an exploratory study, reporting themes or concepts that are less supported is useful for future research; a theme that fewer than half of the participants in a grouping discussed might be more widely supported in a larger sample, while a theme that is widely supported here could diminish in importance.

Finally, I reviewed each code summary (as well as completed a sizable number of the summaries myself), ensuring that the summaries I did not create were consonant with the data. I then used the summaries as the building blocks of my analysis of the data, comparing and contrasting the themes and their relative weights within the respondent pool.

### 3.2.7 Generalizability and Limitations

An important question with all qualitative research is its generalizability—to what extent can the findings be generalized to a larger population? This dissertation uses a purposive sample of twenty participants: ten pregnancy app users, and ten 23andMe users. All twenty of the participants

identified a primary search engine. While the selection of contexts was crucial for isolating an exchange relationship and investigating its form, the specific characteristics of each context should not be generalized across entire user populations. Meaning, for example, that the experiences of these twenty participants should not be construed as representative of the experiences of all U.S. users of search engines. But this limitation is directly linked to the intent of this study, which was not to perform case studies of how people use each of the three services, but instead to use services as a means to investigate the applicability of SET to explicit forms of exchange.

Further, the conclusions drawn in this dissertation might not be generalizable to other exchange relationships, even other direct negotiated binding exchanges. These three contexts were selected because the exchange of personal information for a service was explicit. However, other relationships in which the exchange of personal information isn't as explicit may not share the same features, or one might find that similarly explicit information exchanges in other contexts raise issues that did not arise in this study. As this work is exploratory, to the extent that it does raise compelling issues it opens doors for future inquiries of other relationship forms or contexts.

The sample itself also presents limitations. As discussed, this was a convenience sample, geographically limited to the city of Berkeley, CA. While I made every effort to recruit as diverse a sample as possible, ultimately it is reflective of the residents of this area and the specific membership of the websites and mailing lists I recruited from. That said, it is also limited to the user populations of the two primary contexts—pregnancy apps and 23andMe—which themselves are not nationally representative populations of the U.S., nor potentially of the services themselves. I do not have official demographic data for both contexts (other than the fact that pregnant women, and thus the majority of mobile pregnancy tracking app users, are female). Furthermore, we can reasonably conclude that usage constraints for these services, such as owning a smartphone for pregnancy apps, and the ability to afford the purchase price for 23andMe, also limits the scope of the population that could be included in this study.

Finally, unlike the pregnancy app participants, with whom I explored the concept of notification norms as a check against their app use, given the novelty of genetic testing no comparable norm exists for 23andMe users. Because this sample consists of participants who elected to sign up for genetic testing and does not include interviews with people who have not or would not use DTCGT, I do not have a simple way of addressing concerns that, for example, people who elect to participate in DTCGT are less concerned with aspects of privacy than those who would not.

## Chapter 4

# Qualitative Interview Findings

### 4.1 Overview

In this chapter I review and discuss my qualitative interview findings, using the following structure: in Section 4.2, I set the context for each of the service types by discussing why the participants elected to use these services, how they chose them, and their general impressions of the companies and services. In Section 4.3, I review the themes related to aspects of social exchange theory: the nature of the relationship between the participants and their apps or service, participant perceptions of fairness and benefits in the relationships, and the role of assurance structures. I discuss the themes in turn within the context of each service: pregnancy tracking, genetic testing, and internet search. In Section 4.4, I review the themes related to personal disclosure and privacy: the mechanics of the disclosure relationship and how it related to respondents' perceptions of privacy, the risks participants felt were posed by their disclosures, and the trade-offs participants made in order to use their app or service. Again, the discussions are organized by each service.

Several key findings emerged from these interviews:

- Participants' relationships with their service or app fit the model of a direct negotiated exchange. As such the relationships were: transactionally focused; provided a clearly understood benefit to the participant; and, relied on assurance structures to mitigate risk and uncertainty.
- The relationships showed evidence of power imbalances in favor of the companies. In some instances, these imbalances were mitigated by providing the participants options and flexibility in the disclosure relationship.
- Participants relied on both formal and informal assurances when disclosing to these services, though they overwhelmingly favored informal assurances. These included: company reputation (via direct and indirect social assurance), website or app visual design, information quality, and anonymity, among others. This finding suggests that policymakers or

researchers who focus only on formal assurance mechanisms, such as laws, terms of service agreements, or privacy policies, are considering limited data.

- Even when relying on assurance structures, participants used multiple strategies to reduce their exposure through disclosure in order to maintain their privacy expectations.
- Within the genetic testing context, participants were also simultaneously engaged in an indirect generalized exchange with the company through their contribution of their data to 23andMe's research program. Participants' contributions to this exchange deepened their commitment and motivation, with an effect of neutralizing the potential risks of genetic testing.
- The length of the relationship, the substance of the disclosures, and the frequency with which they occurred had a substantial impact on privacy concerns. Participants voiced the greatest concerns with their search queries, where their relationship with their search provider was the longest, they felt their aggregated search queries were the most sensitive data disclosed, and the accretion of the queries over time created a deeply personal portrait of who they were and what mattered to them. In contrast, despite the unique identifiability of DNA and its potential for exposing one's propensity to inherited disease, the genetic testing participants found it to be far less sensitive than their search queries.

I will review and discuss these findings at length, including the findings from the experimental survey, in Chapter 7.

## 4.2 Interviews: Setting the Context

This study consists of a set of purposive interviews with twenty participants across three contexts: pregnancy tracking applications, direct to consumer genetic testing (DTCGT), and internet search engines. Half of the interviewees used a pregnancy tracking app, half used the genetic testing service 23andMe, and all were asked about their use of internet search engines. As I review in detail in the methods section, these choices were strategic; my foremost goal was to focus on a direct exchange between an individual and a company where the outcome of the exchange consisted of trading one's personal information for access to a service. I identified the two primary contexts—pregnancy tracking and genetic testing—as fitting this criteria. Furthermore, the fact that one context featured free services (pregnancy tracking apps), and the other a paid service (23andMe), provided me the opportunity to explore whether the dimension of cost added any substantive differences between how participants valued the exchanges.

Although both of the primary contexts feature an upfront exchange of personal information for access to their services, customer motivations for using these services as well as the substance and amount of personal information disclosed differ substantially between them. Thus, I could not directly compare the two participant groups on these measures. In order to establish a baseline personal disclosure condition across the entire participant pool I decided to ask both groups about their relationship with their primary internet search engine. This strategy allowed me to both explore varying aspects of disclosure relationships and ask about a form of disclosure that differed



from the primary contexts, one that is accretive and drawn out over time and not focused on a primary transaction (*e.g.*, an initial sign up, when users typically fill out forms and disclose their email address and other personal information).

In this section, I discuss why the participants chose their pregnancy tracking apps and 23andMe's genetic testing service and their general impressions of the companies. I also review how their disclosures to these services compared to existing norms in similar contexts. For example, in the pregnancy tracking subsection, I discuss how these participants elected to reveal their pregnancies in order to establish a sense of the extent to which these participants conformed to existing notification norms, and whether their disclosure to an app conforms with these norms.

In the 23andMe subsection, I discuss participants' motivations for using the service, and their general impressions of the company. I highlight one of the motivating factors that proved to be important in shaping participants' relationships with 23andMe: the belief that their genetic testing results both benefited them and aided the company's scientific research. I then review the lack of extant disclosure norms with DTCGT, and the general risks associated with disclosing one's DNA.

Finally, in the search query subsection, I discuss a different set of issues related to the participants' long term use of search. Because search engine use is well-established among mature internet users (including this participant pool), I wanted to understand the mechanics and impact of long term disclosure with one's search engine, particularly a process that is slowly accretive over time.

## 4.2.1 Pregnancy Tracking Apps

### 4.2.1.1 Selecting Pregnancy Applications

The participants in this study found their apps from multiple sources: online searches, app store searches, referrals from friends, reviews from media sources, and through brand recognition. Most noted that they only used free apps and that their app's free status was a major factor in their deciding to download it. The majority of the participants downloaded a pregnancy tracking app in order to obtain basic pregnancy information, and some to use various tracking features (*e.g.*, weight gain, contraction timing). "Like news about the hospital recommended or the nutrition, which are important—or foods to avoid when you are pregnant." [P8]

Some had also used ovulation and menstrual cycle trackers prior to their pregnancies in order to facilitate getting pregnant. All the participants reported positive experiences with their apps. "I wanted to get pregnant and I'm a very calculated person and wanted just to use the tools that were available . . . I just felt really organized with all of my fertility data I was collecting." [P3]

### 4.2.1.2 Company and App Impressions

After discussing with participants their reasons for tracking their pregnancies, I asked them about the app(s) they chose and what their general impressions of the companies were. Their impressions were collectively positive, based principally on: the design and ease of use of the app; its information organization; friendliness of tone; low commitment for use (*e.g.*, minimal information

was required to start using them); community features; visual design; and information quality and relevance. The last two items in particular (visual design and information quality) signalled to participants that the apps were reputable through the perceived amount of investment the companies had put into them. Several women attested to the usefulness of their app for tracking information that they otherwise couldn't keep track of, as well as the quality and usefulness of the insights the apps provided.

“I thought it looked pretty easy and low commitment and I didn't really have to enter a lot of information, just my due date, so it was easy to get started—I could always just delete the app whenever I wanted.” [P6]

The participants that had criticisms were focused primarily on concerns with the app's content. Some participants used multiple apps, trading between them for different feature sets. Others switched and abandoned their original choices, due to dissatisfaction with features or content. In a few cases participants were concerned the amount of disclosure required: “I didn't [track in the Glow app] because I already did that in Ovia and then I felt like they wanted too much information or something. It was a little too—I didn't like the tone as much there or something.” [P5]

#### 4.2.1.3 Personal Disclosure and Revealing the Pregnancy

I asked participants how and when they revealed their pregnancies to their families and friends in order to get a sense of how closely this group adhered to traditional *notification norms*: as Dan Ryan defines them, notification norms are rules of information transmission that are motivated by role obligations and governed by social rules “that constrain who should be told what, when, and how.”[99] I documented this aspect of the participants' pregnancy experience in order to understand how closely this group hewed to existing notification norms. I wanted to establish whether or not they conformed to existing norms in order to understand how their disclosure to a pregnancy app might differ from how they disclosed their pregnancy to their families and friends, and to explore whether women who chose to disclose to a pregnancy app might be less concerned about their privacy than those who do not. For example, did the participants wait the traditional three months before they told anyone besides their partner or immediate family?<sup>1</sup> When they did disclose, whom did they tell first, and how? Did they follow conventional norms and disclose in person or by phone, or did they flout existing norms and disclose publicly via social media?

Arguably, traditional pregnancy notification norms are in flux given both the many affordances available today to broadcast one's pregnancy status, shifts in personal disclosure norms generally, and the effect of having multiple communication channels (*e.g.*, texting, posting to Facebook, etc.) by which to notify both close and distant (in both emotional and geographic contexts) family and friends. For example, some of the apps the participants used provide the ability to update one's connections on various social media channels with one's current pregnancy status. But if, for example, all of the participants in this study elected to reveal their pregnancies through their apps, or through a social media channel, this behavior could suggest that pregnancy app users were less constrained by traditional norms.

---

<sup>1</sup>Most women are advised to tell as few people as possible prior to reaching the three-month mark given the higher probability for miscarriage during that time period.

This was not the case—my interviews revealed that the pregnancy app users were still beholden to traditional notification norms. The majority of the participants made their initial reveal in person, although physical proximity to their loved ones was a determining factor for this mode of communication. Others also made the initial reveal (usually to parents and in-laws) by phone or email, and none of the participants used social media as a means to communicate their pregnancies to their closest loved ones. “I told my friends initially, just only work friends at first. Then it was kind of a secret from other people for a while. Then I finally went home and told my parents in person. Once they knew, then I felt like I could tell other people. For people that I’m really close to, I tried to do it in person.” [P6]

Several, in fact, refused to use social media in any form to announce their pregnancies: “Mostly in person. I’m not a really big fan of sharing my pregnancy and all on social and media and everything. Oh, I’m pregnant. I don’t think it’s a good way of telling people, so I mostly shared in person.” [P8]

For those who did eventually post their pregnancies to social media (usually Facebook), the reasons mentioned were to let faraway friends know the news and to avoid awkwardness in the future. “Towards the end—we actually very recently made my pregnancy Facebook official, so towards the end of the list, like, ‘Oh, I guess I should tell this person,’ I might have a texting conversation about it. Just like, ‘I want you to know before Facebook finds out.’ I guess I always knew I wanted it to go on Facebook, so my friends in far-reaching places who I wouldn’t necessarily tell could know without suddenly, if we were in contact one day, I don’t want them to be like, ‘Oh, you have a kid?’ It just feels awkward.” [P6]

## 4.2.2 Genetic Testing

### 4.2.2.1 Motivations for Genetic Testing

The participants who used genetic testing all reported signing up for 23andMe with the express purpose of obtaining genetic information about inherited diseases as well as personal ancestral information. Because 23andMe is at the time of writing the only DTCGT company that offers information about inherited diseases, this feature was generally the primary reason why participants chose this service over the handful of other competitors in this space, though some had also used some of the other ancestry-focused services. Participants described their impressions of the service in multiple ways, but all in positive terms: “fun,” “cool,” “unique,” “special,” “interesting,” to name the most frequently mentioned. Most cited the health and medical offerings as a primary reason for participating in genetic testing: wanting to know how their genetics could predict personal health conditions and risks for diseases. Two participants were specifically concerned with their carrier status of certain genes that could be passed down to their offspring, should they be carriers. Genetic testing gave these participants answers so that they could make the decision to have children or not.

Some participants cited gaps in ancestral memory for motivation: adoption in the family, divorce and being raised in a single-parent home, or being the only living family member. “With the speed of research and the advancement in technology, it’s always nice to have additional informa-

tion on your health, because you just never know. I don't mind sharing a little bit of myself out there if I get additional information. It's always nice to be a little preventive, because again, my family aren't the best at recording health history, so I have no idea what runs in our genetics or what doesn't." [P20] A few participants wanted to locate unknown living family members. "Now the reason I got into it, this is complex. My mother was adopted, so she never knew her parents, her real bio parents. My bio dad, and my mother divorced either just before, or just after I was born. I never knew anything about him. My mother would never talk about him, and even other members of the family that knew him would go, 'Oh, well, ask your mother.' Then, I'd say, 'Mom, can you tell me something about the——' 'Oh, well, I don't really—maybe another time.' Oh, God, yeah, so I thought, 'Well, I think it's time. I've always wanted to know something. I always had this sense that maybe I have brothers and sisters, or relatives out there that would help me figure out who the hell I am.' I went ahead and sprung whatever it was, \$190.00, and thought, 'You know this is a good—this'll be helpful.'" [P12]

While the primary motivation for signing up was to gain personal insights, many participants expressed their belief that their data was helping the company contribute to solving problems for the public good. Most of the respondents noted that their contribution to 23andMe's medical research influenced their decision to participate in genetic testing, as well as their selection of 23andMe over other genetic testing companies. "[Compared to Ancestry.com], 23andMe is more of just trying to understand the human genome more. It's more of big question things rather than let's find your lost grandparent. I think that the [company] icon points to that, too." [P13] Many also discussed the societal importance of genetics research and the necessity of large sampling to enable its success. "I think everyone should contribute their information to this database. That's the main thing. I understand that people have privacy concerns, but in this case I don't have any security or privacy concerns about 23andMe." [P17] I will discuss this aspect of the exchange in more detail in Section 4.3.6.

#### 4.2.2.2 Participant Impressions of 23andMe

The participants' impressions of the company were overwhelmingly positive. The most significant positive factor was the participants' belief that they were helping the company to solve problems for the public good. "What's nice is that from what I read, you're part of a bigger scientific study, right? In some ways it feels good to help further science research." [P20]

"It feels, I guess, more like a community garden where everybody just goes and they do their own thing, but it's all together...we're all cooperating, and we're all getting something out of it." [P1]

Other impressions were driven by participants' perceptions that the company was well-funded and well-established, as well as the visual design and good usability of the website. "It's fun. Yeah. It's bright, bright colors. It's, what, pink and green and fun? It's really not medical. They do a very good job of not seeming Big Brother-esque and making it like, 'Hey, find out more about yourself.'" [P19]

Nearly all of the participants were aware that 23andMe had been in regulatory trouble with the U.S. Food and Drug Administration (FDA), though most had only a vague understanding about

why. However, the FDA's intervention did not tarnish the company's reputation in most participants' view. In general, the respondents who were members prior to the intervention had a positive view of the company and a negative view of the intervention because it led to their having access to less information than they had previously. "I thought it was stupid. I liked having that access . . . [y]eah. I don't agree with that ruling. I think people should have access to their medical information.[P17]

#### 4.2.2.3 Complicating Social Norms

Unlike pregnancy, DTCGT is such a new phenomenon that there are arguably no extant notification norms—and minimal to any established social norms—regarding sharing one's genetic testing results. There is not a directly comparable experience with DTCGT as there is with pregnancy to assess the degree to which this group of participants compares against an established norm of disclosure. However, it would be wrong to conclude that voluntary 23andMe users are somehow significantly less concerned with privacy than the norm. There are several influential factors at play here—an expectation of anonymity on the service, the belief that contributing one's data aids scientific research, to name two—that I will discuss in more detail below. However, one factor that merits a quick discussion here is that of assessing risk. The potential risks posed by DTCGT that concern privacy advocates and researchers are not widely publicized, and the respondents in this study were generally unaware of them.[118] Additionally, the need to provide health practitioners with bodily samples is legitimate and normalized in the medical context, and despite the company not being a medical practitioner, the influence of the overriding health context, combined with the triviality of the sample itself (saliva), is one that did not trigger any sense of risk for these participants.

Interpersonal risks are also important. For example, the impact of an individual's choice to submit a DNA sample to 23andMe or a similar company is not limited to that individual; her DNA could be used to identify anyone genetically related to her, including yet to be conceived offspring.[92] Allowing a private company to act as a custodian for one's identifiable genetic material indefinitely (and theoretically, forever) poses multiple risks, including being subject to requests from law enforcement, theft by hackers, and the potential for undesired commercialization.[102] [61] While the identifiability of relatives through DNA may pose minimal risk in typical family relationships, stories continually emerge about the discoveries people made through DTCGT about relatives they knew nothing about, most frequently "paternity events:" discovering your father is not your genetic relative, or discovering half or full siblings one never knew existed. For example, one woman unexpectedly discovered she was fathered by a sperm donor, and had at least three half-siblings, after she shared her genetic profile on 23andMe's platform.[98] Some of the participants in this study shared similar stories, not all with positive outcomes—occasionally these discoveries lead to turbulence and strife in families as secrets emerge that the secret holders never intended to be known.

### 4.2.3 Search Engines

It is difficult to overstate the importance of search engines, but specifically Google, in mediating our relationship to the internet. For many, the internet is only experienced through one's primary search engine, both as a portal to the internet and an information locator; if your search engine can't find it for you, the content effectively does not exist.

Unlike with pregnancy tracking and genetic testing, all of the respondents had long term relationships with their search engine (the shortest was about five years) with continuous engagement. Because the relationships were well-established, and because using a search engine is an activity that is routine for all but the most novice internet users, I did not ask the participants the same set of context setting questions around how they came to use their search engine. As I describe in detail in the methods section, I collected details about the participants' search usage from an intake survey, verified that the search service they indicated was in fact their primary one, and explored the mechanics of how they engaged with it.

That engagement—the exchanging of search queries for results—is both accretive and routine, an action that becomes woven into the daily experience of using the internet. For example, when I asked participants if they knew if they were logged in when using search, only a few could definitively say yes or no; most weren't certain, and after asking several follow-up questions (*e.g.*, Do you also use Gmail? Are you using it in the same browser that you conduct a search query? Do you see your name or icon in the top right of the browser window?), we determined that most respondents were logged in (to Google) while searching. “That’s the thing. I don’t make a point to log out, so I guess I am logged in.” [P6]

“I don’t think I have to log in to use it. I’m sure it’s all connected. I’m probably already logged in to Gmail, so when I’m searching—so I know I can use it without logging in, but I know that when I am logged in, it’s all connected.” [P5]

How closely one's search queries are tracked hinges upon whether one is using their search engine while logged in (when queries can be uniquely identified) or not (queries might be coarsely identified or simply aggregated based on IP address and inferred demographics). This is an important distinction given that Google stores search queries *forever*. Only half of the Google-using participants were aware that Google keeps a personal search history for its registered users of everything they search for, in addition to other types of usage activity.<sup>2</sup> Participants' individual perceptions of their relationship were clearly affected by the extent to which they understood how their search queries were used. At the same time, there was not a direct relationship across all participants between the depth of their understanding of how search companies used their data and their acceptance and comfort with the exchange—some participants appeared aware of how their data was used but did not have misgivings about this relationship.

---

<sup>2</sup>Available at: <https://myactivity.google.com/myactivity> as of August 2017. Note that a few participants were confused about the difference between search history and browsing history.

### 4.2.3.1 Evolving Search Norms

Search engines have evolved from their initial purpose of simply delivering results in exchange for queries, and none potentially as much as Google, which since its inception in 1998 as an academic research project now accounts for most of the world's internet searches, along with a sweeping suite of software apps. As part of that evolution, Google moved from merely organizing the world's data to also collecting data about the individuals who use their services, starting with search queries. But even before privacy researchers and advocates began to voice concerns about the depth of the company's data collection about its users, one public incident revealed how well one's anonymized search queries could uniquely identify you: AOL's release of an anonymized search dataset in 2006 of 658,000 of its users, which after its release took a couple of *New York Times*' reporters only a few days to verify the identity of one of the users.[11]

Few users likely remember using Google when it was advertising-free; the company introduced ads (AdWords) alongside its search results in 2000, and the core model today remains fundamentally the same: based on one's search terms, the company delivers ads it deems relevant along with your search results, though the display of sponsored results has evolved.[74] However, this basic model—you express explicit interest in a topic, and a website shows you ads for it—continues to dominate how content is served across the internet, not just on search websites. As such, it is no surprise that this model is so widely accepted for serving content that the participants widely referenced it in their interviews. And to the extent that the search engines' business models reflect the straightforward exchange of search terms for ads, most participants had few concerns about it. However, this model is in flux, as online companies move away from direct associations based on behavior to algorithmic-based inferences.[116] Not surprisingly, anxiety emerged among participants in the cases where the exchange appeared to deviate from the explicit behavior-based model, such as when the company acted on inferences it made about one's activity. As one participant described: "I think [Google is] getting too much personal information these days. I think they track your search keywords. If you search for something once, it will always pop up on your screen. . . [a]nd they're getting so smart about it. If you search for plane tickets, they will tell you about rental car[s] and hotel[s], and all the other thing related to it as if they know who you are and what you're doing." [P7]

### 4.2.4 Summary

Across both primary contexts, participants selected either their pregnancy tracking app or elected to participate in genetic testing because they had specific information-centric goals: to gain insights into their pregnancies or genetic backgrounds, or to assist with managing their pregnancy or their health generally. Both groups' choices were influenced by factors such as company reputation, personal references, the visual design of the app or website, its perceived quality, and the ease of use, both to start the relationship and to maintain it.

With regards to personal disclosure, the pregnancy tracking app users' choices to disclose personal information to their app did not supplant or alter their adherence to existing pregnancy notification norms in their personal relationships. While the novelty of DTCGT makes it difficult

to generalize whether the genetic testing participants were inherently less concerned with their personal privacy by virtue of submitting to such a test, I will present findings later in this chapter that suggest that the participants were unaware of the potential privacy risks of these services, as well as the impact of participating in 23andMe's genetic research program on participants' commitment and motivation. Within the context of search queries, the duration of the participants' relationship with their primary search engine, as well as incremental changes in the model search companies follow for using their users' data, illuminate the growing concerns participants have with this disclosure relationship.

### **4.3 Social Exchange Theory and Relationships**

In this section I explore aspects of the exchange relationship using social exchange theory as my framework. I discuss the participants' perceptions of the relationship they had with their company, focusing on three distinct areas: the nature of the relationship itself; the participants' perceptions of fairness and benefits in the relationship; and the factors that contributed to trust in the relationship, including whether the participants relied on assurance structures (and if so, what factors mattered to them).

#### **4.3.1 The Nature of the Relationship**

My questions about nature of the relationship were intended to draw out insights about the core of the exchange: what did the participant believe she getting from the relationship, and what did she think she was giving up? What did she think the company gained from the relationship? Would she continue the relationship in the future? I asked the participants whether the relationship met their expectations in order to determine whether the relationship resembled a direct negotiated exchange (DNE, where the outcome of the exchange are defined as part of the negotiation), or a reciprocal exchange (where the outcome of the exchange are not negotiated and emergent). I also explored questions of cost: if the service was free, would the participants pay to use it? If, in the case of 23andMe, there was a cost attached to the service, would the participants use the service if it were free? I asked these questions because free online services are not free—users typically pay with their personal information. Would changing the cost proposition alter the terms of the relationship?

As we discussed these questions, other themes emerged: the impact of the duration of the relationship, and both the depth and substance of what was exchanged (pregnancy information, DNA, search queries). It was clear after conducting the interviews that these relationships fit the form of direct negotiated exchanges rather than reciprocal exchanges. The core features of reciprocal relationships—uncertainty regarding exchange outcomes, affective bonds—were generally absent in these discussions. Despite there not having been active negotiations between the participants and the companies in a classical sense—no one sits down with a representative from the company to jointly negotiate the terms of disclosure—they are negotiated in the sense that one partner (in these contexts, the company) offers the terms and the other partner (the individual) accepts or declines. While many of us may find this form of negotiation to be unsatisfying, given that the terms



are essentially ‘take-it-or-leave-it,’ it is reflective, I would argue, of both the power dynamics and existing social norms in these types of relationships.

Unlike with reciprocal exchanges, the outcomes of these exchanges were known in advance insofar that the individual generally understood what she would receive from the company in exchange for her disclosure. Of course, one of the compelling issues within privacy research is the fact that many online companies engage in exchange relationships with third parties using the data obtained from these first party exchange relationships. While many of the participants voiced concerns about these exchanges, I determined that these third party relationships were out of scope when assessing the form of the primary exchange relationship. Many users are often unaware (or don’t understand) that these secondary exchanges are even occurring, and while I did ask participants about these issues in order to understand how it influenced their disclosure calculations, I did not factor whether or not a company was engaging in secondary exchanges as part of determining the form.

### 4.3.2 Pregnancy Tracking

When asked to consider what they offered the company, several pregnancy app participants were initially flummoxed by this question as they hadn’t considered it explicitly. One participant, who professed that she had little experience with apps and was a newer smartphone user, had no idea. Most offered guesses related to advertising and company reputation: “I don’t actually know what they’re getting out of it. I hadn’t thought about it. I think by probably having their numbers they could tell people that the ads—that they have more potential viewers. That’s a good sell for them. I haven’t really clicked on anything, so I don’t know if they’re upset with me.” [P10]

“What do they get from me? I think that [they] advertise themselves, right? If more people start to use it, so they’ll get a good reputation and recommendation from every person so that they can build a really good reputation in the course of the time so they can attract more customers.” [P8]

Other participants were familiar with the free app model: offering one’s personal information in exchange for having to view ads: “That’s how they make their money. It’s a free app. They sell to advertisers. There’s a big market for that. I understand that. I don’t really like that, but that doesn’t mean that it’s gonna ever go away.” [P2]

“I am offering them data, and essentially I’m a statistic for their advertising numbers, when they sell their ads.” [P11]

Notably, no other form of exchange model was suggested by the participants—if the women had a model in mind at all, it was expressed in terms of advertising. When I asked the participants whether they would consider paying for their app, most of the participants offered that they would have been willing to pay a modest amount; several noted that they would expect less or no advertising at all in exchange for payment. Of the participants who were unwilling to pay for their app, most responded that it was too easy to get the information they needed elsewhere for free, even if it was less convenient to do so. “[T]he way that I use it, I don’t see any benefit to paying for that information. . . on the one hand, I think maybe it’s not fair. They’re providing the service and I’m not paying for it. That’s their business model. It’s not really my problem.” [P2]

For most pregnancy app users, their relationship to their app(s) was often slightly shorter than the duration of their pregnancy unless their app offered infant tracking features (none of the apps offered fertility/pre-pregnancy tracking). Thus, the length of the relationship with the app was discrete and typically ended with childbirth or soon after (a normal term pregnancy is nine months (40 weeks)). Most of the women I spoke with were currently pregnant for the first time or had just given birth to their first child; one participant had had two children, the youngest of whom was over a year old, and she used different apps during each pregnancy. While participants interacted frequently with their apps through their pregnancies, and some participants provided the apps with additional information beyond the initial exchange at sign-up, for most participants the salience of the app lasted only as long as the pregnancy itself. Some of the app companies offered additional products (apps, books, web content) that a few participants had transitioned to using or had expressed intent to do so after the birth of their infants.

None of the women had any reservations about continuing the relationship into the future if they had a reason to use the app. While all the participants had a positive impression of their particular app, only one respondent articulated this view specifically as a relationship: “I guess I felt like, “Hey, I’m grateful that this exists to figure this out for me.” I really like them. I liked the information that they have. Was it a super cozy warm loving relationship? I guess it was like good friends or something, like a supportive relationship in that way.” [P5]

Overall, participants did not have personal or deeply affective relationships with the companies who powered their pregnancy tracking apps. Few felt any sense of obligation to the companies, even when their apps and associated content were free: “No I feel like it’s any other web page. You don’t pay for web pages online that gives you as much information as this one with the [app]. . . it’s just maybe that it’s more personalized to you. I never feel like I owe anything to anyone online. If it’s not some—I don’t know—person who has their own business and they’re trying to—I don’t know. . . it just seems like a big corporation.” [P9]

Others suggested that their minimal investment in the product, accompanied with minimal switching costs, contributed to their lack of a sense of obligation or any affective bonds. For the one company that had a brand presence beyond the app (What to Expect, which began as a book and has grown into a series of books, a website, and the app), a couple of participants did express some loyalty based on their use of affiliated products. Finally, two participants noted that not only was it the company’s choice to make the product free, the product itself was not really “free” given their exposure to ads.

When asked if they would opt to continue their relationship with the company in the future, the majority of participants said that they would be open to doing so, but their answers focused on the temporary nature of their pregnancy status and were qualified based on whether they might use it for a future pregnancy, or whether the app had tracking features for newborns.<sup>3</sup>

---

<sup>3</sup>Most of the apps did not include newborn tracking features. Some companies offered additional apps that one could download after the baby was born, and most of the participants noted that if they decided they wanted to track their infant, they would likely start by downloading the company’s infant app.

### 4.3.3 Genetic Testing

In discussing what they gained from the company, the genetic testing participants spoke of the insights they gained about themselves, their genealogical heritage, heretofore unknown relatives they were connected to, and their genetic health information, which was of particular value to those concerned about inherited conditions. “[T]hey have told me a little bit about my body and my ancestry and my markers for certain conditions. I feel like that was almost a gift. They gave me something. They told me something about myself, and I essentially gave them very little. . . I just hope that it can be used for something good.”[P19] Several of the participants also emphasized that what they exchanged with 23andMe was, from their perspective, just data, and not personal. As one participant phrased it, “It’s not personal what I’m giving them at all. It’s just like little pieces of numbers and letters.”[P14] One participant joked that the data he provided wasn’t enough for the company to make a physical clone of him, and thus the risks to him were minimal.

Only one participant explicitly described his relationship with the company as a relationship, but his description aptly summarized how most of the participants characterized the company: “It’s a business customer relationship. We really don’t have the warm fuzzies going on, but they keep me up to date when they’ve got new information, or need new information. I think it’s fine.” [P12]

Overall, respondents reported getting what they anticipated from 23andMe: the exchange conformed to their expectations. The unexpected outcomes participants did report were focused on their genetic reporting results themselves not conforming to pre-held expectations, or on disagreements with the company’s methods for correlating self-reported survey results with genetic tests. Most were impressed by what they learned about themselves in exchange for the cost and a vial of spit: “I didn’t really know what to expect in terms of the value that I’d get back, but what I did get back was cool enough that I—yeah, it was worth 100 bucks and the sample.” [P15]

When discussing their expectations, respondents referenced their ability to exert control over their data (*e.g.*, consenting to research studies or sharing data on the ancestry platform were optional), their expectation that their data was anonymized when analyzed or used for research, the lack of any negative consequences or intrusions (thus far), and the belief that their data was being used for public good. “I feel like they’ve given me a lot of personal agency over opting in to things, that I get to choose what I wanna share, what I wanna tell them.”[P19] Although half of the participants used the service for over five years, most noted that they did not login very often, and if they did, it was principally in response to email outreach by the company or for the entertainment value of completing additional surveys. Thus, while it is possible that a relationship with 23andMe could be limited to a user’s initial sign-up, sample submission, and receipt of their DNA sequencing results, users could choose to both continue the relationship and provide the company with additional personal information beyond the primary exchange transaction. Nearly all of these participants opted to continue the exchange and none expressed any plans to discontinue it. However, it appears that the core of these relationships were focused on the initial transaction, and that subsequent exchanges were far less substantive and impactful.

The genetic testing participants had positive experiences but did not describe their relationships with the company in personal terms. This appears to be due to the fact that the relationship was primarily centered on the initial transaction. As one participant described: “I don’t even feel

like there's a relationship because it's like, how would I continue it or not continue it? I can't take back my DNA, and not that I would. I wouldn't, and so it's really just do you log on and see what they've discovered recently or not?" [P1] At the same time, the generalized exchange that 23andMe enabled appeared to deepen participants' commitment. Commitment is typically a feature of reciprocal exchange relationships, a byproduct of the affective bonds that can build between exchange partners. Here it provided intrinsic motivation to the company's customers to justify the disclosure of their personal data that exceeded their direct self-interest. As long as 23andMe continues to frame practice of using its customers' data for research that appears to have an explicit public benefit rather than for profit, even if the company does profit from it, they will continue to benefit from this commitment.<sup>4</sup>

#### 4.3.4 Search Queries

"Yeah, so Google—basically, my relationship to the Internet is Google." [P16]

As discussed earlier, the participants' relationships with the internet at large are primarily mediated by their search engine, which nineteen of the twenty participants identified as Google. When asked directly what they offered their search engine, all of the participants had ideas about how their search engine company used their search data, from the abstract to the highly specific (and with the exception of the few who said they didn't know or weren't sure, the suggestions were generally accurate). "Well, they target ads, which I block. I'm sure that data gets sold. I don't know. [Google] make[s] a boatload of money. I never really thought about it." [P2] Some participants were hesitant to describe what they had with Google as a relationship. "I don't think most people do care about whatever the relationship might be. They just want the data. I don't think people really think of it in terms of a relationship. I certainly don't. It's just a tool. I have a closer relationship with my toolbag than I do with [Google]." [P12]

Overall, the Google users had a positive impression of the relationship, based primarily on the quality of the services they used and the overall ease of use. No one expressed any immediate expectations of terminating the relationship. Similar to the other companies in this study, the participants did not express many affective bonds towards the company; their descriptions focused on the utility of search and its usefulness in information seeking. None of the participants reported having a singularly negative experience with Google (such as an account breach).

Some of the most interesting conversations in this study resulted from asking the participants whether they would be willing to pay to use a search engine: participants were split roughly in half between those who were willing and those who were not. For the respondents who used Google, this was a difficult question to answer if they were users of other Google services; the fact that search is one part of a multi-faceted Google experience made it difficult for some to think about search in isolation (and most didn't want consider giving up a service like Gmail). Of those who

---

<sup>4</sup>Consider, as a contrast, the animosity some feel towards Facebook, which collects data that could be used for similar commercial purposes, but does not frame the contributions individuals make as contributing to the public good. Nor can Facebook claim that their primary use of their users' data is for the benefit of all Facebook users or the public in general.

were unwilling to pay, or were uncertain, the reasons they gave were varied: a general expectation that all content on the internet should be free; frustration that Google in particular already made ample amounts of money; and, the most mentioned—that while they were less desirable, there were still other search options available (though, as one participant phrased it, “If you made me use Yahoo! I would probably be sad” [P10]). Thus, despite Google’s effective monopoly on the search market, some participants still did feel they had access to alternatives, albeit inferior ones.

Of those who *were* willing to pay to use online search, they articulated specific benefits they assumed would accompany the cost: a gain in privacy (specifically not being tracked), efficiency, and more reliable or unfiltered search results. As one participant put it, “There would have to be a really specific value proposition that I got out of it because there are so many search options out there that are free. I think either if I had serious concerns about privacy or there was significantly more value in the search, there were more features and it could get deeper into—I guess get better results, or get more results out of protected areas of the Internet, I would.” [P11] However, most of the participants mentioned that the long history of the service being offered for free would make paying for search hard to swallow: “If Google were, all of a sudden, one day, like “We’re gonna charge for this,” people would switch to a different service that was free. We’ve been conditioned as a society to think it’s free.” [P16]

The nature of the participants’ relationship with their search engine, but specifically Google, was driven primarily by utility, quality, and clear expectations of how their search queries were used by the company. In contrast to the other services, the participants’ relationship with Google revealed different dynamics because the relationship was typically much longer and encompassed a greater depth and frequency of exchange. While I did not ask participants to self-report how frequently they used Google for search, it is safe to assume that their individual transactions far outweighed their uses of pregnancy tracking apps and 23andMe, and certainly spanned a much longer overall duration. As later sections will demonstrate, this long and evolving history adds a greater complexity to respondents’ perceptions. Power dynamics are a far more prominent feature of the search relationships in this study than with the other services, particularly switching costs and resignation in the face of a practical monopoly.

### 4.3.5 Summary

In exploring the nature of the relationships across the three contexts, what was first apparent was that there were relationships—and while the participants did not routinely describe them using that specific term, their responses to the primary question demonstrated that the concept resonated. The relationships across all three contexts share the features of direct negotiated exchanges, rather than direct reciprocal exchanges. Unlike relationships based on reciprocity, which rely on non-negotiated and open-ended exchange, negotiated exchanges have defined expectations and outcomes. Further, the relationships lacked other features of reciprocal relationships, such as affective bonds. Instead, participants described their willingness to exit the relationship should better alternatives arise. The participants viewed the services with which they interacted largely as tools or utilities, and most were cognizant that the transaction required them to contribute information in order to gain value from the exchange. This was a useful insight, as prior to conducting the

interviews, I hypothesized that the relationships could be based on reciprocity, particularly the free services, and that the users might conceptualize the company as a human-like actor. That wasn't the case.

The relationships were characterized by both low friction and low terms of commitment for the participants. Even the 23andMe experience, which had the most upfront friction in terms of requiring participants to obtain a physical kit and then wait six weeks for the results, was based on a minimally invasive procedure—spitting into a tube. But it is precisely the ease of the exchange that both enables it and allows it to flourish. These aspects were most apparent with search, which is so simple to use that it fades into the background of one's online experience, becoming nearly invisible. As such, it is used so frequently, and over such an extensive duration, that the exchange of information builds and builds slowly and invisibly. The utility of the exchange, and its unbound- edness, enables a relationship that allows the search engine to compile an extensive history of these transactions.

Another emergent theme was the extent to which individual expectations were bounded by participants' mental models of the mechanics of the exchange itself. Insofar as the information exchange is direct and follows an obvious logical model regarding how the exchange partner (the company) uses the data either for providing the service or targeting advertisements, the participants were generally accepting of the terms of the exchange. Additionally, possessing some amount of negotiative power in the form of controlling individual disclosure—how much, to whom, when— provided participants with a degree of agency. When the company introduces uncertainty that violates this model, concern rises, specifically when the uses of the participants' information do not adhere to their mental models, such as distributing information to third parties, or using it to make inferences about the participants that go beyond its original collection context.

An interesting feature of these relationships is the existence of *secondary relationships* in these exchanges that both simultaneously co-exist with the primary direct relationship and are based on it. The pregnancy tracking apps in this study are a typical example of such a relationship, where a company uses the output of the exchange with the individual (the individual's personal information) to engage in a secondary relationship with another partner, such as a data aggregator or advertising partner. This secondary relationship consists of the access, sale, renting, or trade of the individual's data directly to a third party.<sup>5</sup> This relationship generally violates the expectations of the individuals in the primary exchange relationship, who have mental models that explain the advertising based on the direct exchange relationship, but don't encompass the secondary relationship. Given the dissatisfaction that customers often express about these secondary relationships, it is not surprising that they are typically not well-disclosed to customers, and many are unaware that they even exist.

Compellingly, 23andMe uses this secondary relationship in an explicitly public way. As discussed earlier, 23andMe asks its customers to opt-in to its research studies, a step that is a prominently disclosed on the website. Because the company makes explicit assurances about how it

---

<sup>5</sup>Of the six pregnancy tracking apps used by the participants (I Am Expecting, Pregnancy+, Ovia, What To Expect, Glow, BabyBump Pregnancy), only one, Glow, stated that they did not sell or rent their customers' personal information to third parties. All privacy policies were reviewed in August 2017.

uses its customers' genetic data, as well as the emphasis on using the research for public benefit, among the participants in this study the secondary relationship (an indirect generalized exchange) bolstered commitment and motivation. The participants believed that they were both receiving a direct benefit (from the primary exchange) as well as contributing to an indirect benefit (the potential to help themselves or others through discoveries made through 23andMe sponsored research using their DNA and survey responses). Notably, the framing of this appeal matters; the company's website highlights the benefits of their research to public health, but never mentions how the company benefits.[52]

The participants' relationships with Google exhibited features that fit the model of a direct negotiated relationship but were also largely absent in discussions of the other two services. Issues related to power, specifically: service lock-in, switching costs, information asymmetries, and a lack of alternatives, were key features of the relationship. These features are likely present due to the duration of the relationship participants have had with Google, as well as the company's functional monopoly over the search market. For example, the information asymmetry between individuals and Google is something that accumulates over time; if a new Google user was asked to disclose upfront the depth and breadth of personal information that Google has collected about a user who has used their search at least weekly for the past fifteen years, there would likely be considerable resistance.

At the same time, Google's role as a core mediator of the informational internet cannot be ignored. For comparison, 23andMe is the only direct to consumer genetic testing service that offers its particular suite of services today, but other options do exist, albeit with a different feature set. Relationships with 23andMe may also exhibit lock-in and switching costs if more options arise for utilizing one's sequenced DNA. But today, obtaining insight into one's genome is entirely optional. For most of us, accessing the internet is not optional—nor is the need to use a search engine to navigate it. By that measure, Google's position in the network of information intermediaries gives it a power advantage that no individuals can challenge.

#### **4.3.6 Benefits and Fairness**

I investigated the participants' perceptions of who benefited from these exchanges and how fair they believed them to be. A core assumption of SET is the expectation that "actors behave in ways that increase outcomes they positively value and decrease outcomes they negatively value." [80] Benefits in a direct exchange relationship flow bilaterally, and I wanted to observe the participants' perceptions of how fair/beneficial they thought the relationship was to them, as well as how they thought the companies benefited. I was also interested in asking the participants about issues of benefits and fairness as I hypothesized that their answers would provide insight into a related issue: the balance of power in the relationship. Did the participants believe they had access to resources or exchange partners who could provide them with comparable benefits? Did they believe they had exit power, or the capacity to leave the relationship? I believed obtaining insight into these dynamics would provide additional context to the disclosure relationship.

Accordingly, all the participants reported benefiting from their relationship, though opinions regarding how the companies benefited varied, particularly by context. Within the search engine

and pregnancy app contexts company benefits were tied directly to monetization via advertising. While the genetic testing participants understood 23andMe benefited from direct revenue received from its customers, their discussions also included a different benefit—the gains to both the company as well as the participants from the participants’ generalized contributions to the genetic research pool. Opinions about fairness and power also varied; while all of the participants believed their exchange with their company was fair, this was tempered in turn by the extent to which participants felt restricted by switching costs and access to comparable alternatives, particularly for Google users.

#### 4.3.6.1 Pregnancy Tracking

“I think the users benefit more. I don’t think they’re really asking for much from the users.” [P10]

The pregnancy tracking participants articulated many benefits they received from their apps, primarily: salient and timely information, ties to other pregnant women through community building tools (*e.g.*, message boards), and a means to track and organize their pregnancy-related data. Overall, the pregnancy app users viewed their exchange as both beneficial and fair; most felt that what they received was far in excess of what the company had asked of them. This view appeared to be based on the fact that engaging with the apps typically required little personal disclosure beyond one’s due date and an email address. “I think it’s probably equal. I just think about it because my value to them is in aggregate, so I know that my individual value to them is pretty low. In the end, I may get more value out of it because it’s not about me, to them, but about thousands of me.” [P11]

Several participants echoed this theme of their personal information likely being of little specific value to the company, but valuable in aggregate or as part of a larger demographic category (*e.g.*, a married pregnant woman aged 30-40 in the San Francisco Bay Area). Complementing the low barrier to entry, there was correspondingly a low barrier to exit, and several participants had in fact switched apps during their pregnancies for various reasons, or noted that if the app did something they weren’t happy with they would just delete it from their phone. Overall, given the numerous access to alternatives that participants had, as well as the relatively low value they placed on the information they provided the companies, none of the participants articulated any concerns with or perceptions of power imbalances.

This unity of factors: quality of the app, salience of content, low barrier to entry/exit, and the ability to control the depth of engagement contributed to these participants feeling as if not only was their exchange fair—to many, they felt as if they had received a greater benefit from the relationship than the company had. Importantly, though, this perception rests upon several key assumptions by the participants, namely: there was low risk of compromise or exposure of their data; that the apps were only engaged in the types of information exchanges (*e.g.*, targeted ads based on higher level demographic and/or geographic categories) the participants assumed were occurring; and that the advertising they were receiving was both obvious and avoidable. To the extent that each of these companies is in fact engaged in straightforward advertising practices and



competent security practices, these assumptions are reasonable, though there is evidence that not all of them are.[15]

#### 4.3.6.2 Genetic Testing

“It’s not really my information that gives them the power. It’s the information of everybody’s.” [P13]

The respondents felt they obtained a considerable benefit from their genetic test primarily due to its novelty and uniqueness—the combination of the results that 23andMe provides can’t be obtained elsewhere in 2018. Further, respondents found their personal results both interesting and salient, especially those respondents who joined before the FDA intervention, when more specific disease information was available. Participants discussed the benefits of sharing information in terms of its personal utility, such as the enjoyment that they got from answering survey questions, gaining information that previously only their doctors knew about them, and the ability to provide helpful information about their genetic predispositions to their doctors.

While self-knowledge was both the primary motivator and benefit for participants, the key secondary motivator and benefit was their contribution to an *indirect generalized exchange*—the belief by most participants that while they were gaining something for themselves, they were also making a contribution to health research that may or may not benefit them directly but had potential value to the public at large. One participant summarized his belief on this topic: “I don’t know that I offer them anything. They’ve already got my money, of course, and sometimes they have questions that pertain to some of their medical research, and when they do, I’m fine with answering those. I think it’s good on a more global, rather than just me specifically. I think it’s really good to have the kind of genetic information databases that they have been accumulating. I think it’s good for the medical researchers, among other things. I think it’s good for people to know that, “Yeah, you’ve got X number of people have that same potential for problems, or the same potential for success. . . I think in any medical study, the larger the sampling, the more accurate the results are gonna be, so I’m just very willing to participate in that. Whether I get any personal benefit out of it or not remains to be seen. I think it’s a good thing for people to participate in this kind of thing for the future, really.” [P12]

This generalized exchange was also explicitly mentioned as a key benefit for the company (after direct revenues from testing kit sales), but not cynically. Most respondents appeared to earnestly believe in the company’s mission, and understood that the value and potential of 23andMe’s service increases as the company expands its database. “The data I get is probably worth more than my one particular sample of DNA. To them, an aggregate sample of the population is way more valuable than just my one result. . . they get this large, pool of data that’s worth way more than what my results would have even been. Then you can start to draw some interesting things about the population. The real value comes from that.” [P16]

“I also feel like I’m contributing to a greater good. Being able to do research across a number of individuals. I think that’s important, because especially with women, you know? We don’t have

a lot of medical information about women because so much research was done on men. I figure I'm going to put my genes out there." [P17]

While most participants characterized the exchange as equal, given that participants paid for the service and received a product in return, several did reflect on the fact that giving up their DNA was likely a non-revocable act, as the previous quote alludes to.<sup>6</sup> Participant 16 noted that while she felt that she could engage with the company if needed, she didn't think it was possible to have her DNA removed from the service: "I feel like I could give them feedback. Whether I could make them, let's say, get rid of my data, and erase it from their server, and dispose of whatever copies of my material they have, I don't think they would do that." [P16] It is notable that the majority of these reflections did not lead to a critical assessment of the relationship or the power balance between the service and the participants. I will discuss in later sections the effect that risk and privacy factors may have had on these views, but an important factor to discuss here that is suggested by this quote ("I could give them feedback") is that of choice. Not only had participants chosen to participate in the service, to date 23andMe has made both contributing to research as well as sharing one's data using the DNA Relative Finder an affirmative opt-in.

Further, the company allows its customers to withdraw consent from participation in future research at any time (though they cannot withdraw previous research participation, even if they cancel their account), and allows them to opt-in to biobanking one's saliva sample.<sup>7</sup> Because the most critical aspects of the service are all opt-in, participants did not feel coerced or taken advantage of in their decision to disclose to the company. "Yeah. I feel like nothing's been forced on me. I feel like everything is opting in, even from the beginning, when it's like, 'Do you want your vial of spit to be used for research?' I really didn't care what they did with that vial of spit, but I liked that I felt like I had some control over it." [P19] While this lack of coercion did not translate into the creation of affective bonds directly with 23andMe, it did lead to highly positive assessments of the experience and at least a willingness to continue with the relationship under its existing terms.

Based on participant responses, in some interviews I posed additional questions about the use of their data by 23andMe's research partners, such as pharmaceutical companies, where the participant's data could help to develop a future drug and reap profits for both companies but yield no direct benefit to the participant. These participants continued to maintain that as long as someone could benefit from the drug that this example of use of their data was acceptable. The commitment to this narrative of contributing to public benefit was strong, and even suggesting scenarios of 23andMe directly profiting from individual data didn't alter this commitment.<sup>8</sup> As one noted in response to my question asking whether they cared if their data was used to benefit pharmaceutical research, "The reason I filled out that kit—even though there are other cool aspects to it, I wanted

---

<sup>6</sup>It appears from 23andMe's Terms of Service Agreement (visited 8/15/17) that if you use the service but do not opt into the research and biobanking aspects, you may be able to request deletion of your account and fully revoke your genetic information from the service. However, I am unable to test this proposition.

<sup>7</sup><https://www.23andme.com/about/biobanking/>; visited 8/15/17.

<sup>8</sup>I must note that I didn't expect this line of questioning to challenge any existing views (nor was I trying to), particularly since this is a participant pool of willing customers. An interesting contrast might be to recruit a pool of people who would never elect to genetic testing to understand what the concerns are.

to find out more about my genetics. Once I got that payoff of ‘Oh, I can log in and see what percent Neanderthal I am,’ once I got that, I don’t really care what they do with the rest of my data.” [P19]

“I never thought of that. Do I have a concern? Well, let’s say that a pharmaceutical company wanted to do some research on what kinds of genetic traits are found in, I don’t know. I mean, I think that’s a good thing. Right? It benefits us. I mean they’re making money, but at the end of the day, we’re gonna benefit from that.” [P17]

When I asked participants if they would have used this product if it were free, most unreservedly said yes. Only two were skeptical of the idea of a genetic testing product being offered for no charge, with one explaining that his skepticism might be ameliorated if the company had a good reputation: “I would question how much they were actually gonna do on my spit sample for nothin’. I mean, where would they get the money to store it, and run tests on it, and all that stuff, and build their website, and answer my emails? No, I probably wouldn’t do a free service. Unless, I mean, if it were something like National Geographic, or 23andMe offering a free service, somebody that I had a high opinion of, or knew about, yeah, sure. I’d do it.” [P17]

In sum, the combination of a highly unique and personalized product that is both painless and affordable to obtain, along with the option to contribute to scientific research that “sometime in the future you or your family may benefit indirectly from” created both a powerful incentive and perceived benefit for participants.<sup>9</sup> As one participant aptly summarized: “I feel like I’m not really offering them very much because there’s so many other people doing it, but I know if no one did it, then, it wouldn’t be good. Yeah, I guess, I’m offering them something good, but it’s just very painless for me to contribute it, so it’s no sacrifice.” [P14] Participants also believed that the exchange was fair, and that the company could not fully benefit unless it continued to increase its scale and recruit more users. I will explore more fully in other sections some of the additional expectations that undergird these beliefs, but I should note here that the company’s extant reputation, lack of any security or privacy crises to date, and most participants’ lack of reference for potential risk also played a role in this perception of fairness.

#### 4.3.6.3 Search Queries

The participants described their access to and use of a search engine as bringing them clear and difficult to replace benefits. Simplicity, ease of use, quality, and speed were the benefits most frequently mentioned. Importantly, most found it difficult to limit their reflections to just the search engine component—most used other products (with Google, specifically: Gmail, Calendar, Maps, etc.) and found value not just from search but from the suite of services available to them and how they worked together. Most articulated the company’s benefit in terms of its ability to target advertisements or to sell access to data, underscoring a general awareness that respondents’ search terms were a source of revenue for search companies, as compared to their queries as source for improving product quality or contributing to a generalized pool of knowledge. One participant described it a fair trade: “They’re getting data and being able to sell the data. I try not to give them any more than absolutely necessary. I know they’re collecting everything. I figure that’s

---

<sup>9</sup><https://www.23andme.com/about/consent/>; visited 8/15/17.

okay. I haven't seen anything sinister that I would worry about. I think it's kind of a fair trade on a certain level." [P12] Others described the relationship as more beneficial to the company than the individual users, focusing on the scale of the information the company is able to collect: "I feel like [Google] probably benefit[s] more than I do. . . I just think that they've been doing this for so long that they're really leveraging this information and the bulk number of—the massive amounts of information that they have to just get a lot of lucrative deals or to create new products or do business development that is really valuable to them. For me, I am getting a benefit, but for the scale in which they operate, and all of the me's out there, I think that they're really getting a huge benefit." [P10]

Most participants found the basic exchange of search queries for results to be fair, given the value of the benefits they received. Participants articulated two primary concerns about fairness, which were related to the longevity of their relationship with their search provider: changes in the 'objectivity' of search results, and the growing information asymmetry between themselves and Google.

Some participants complained that as long-time Google users, they felt their search results over time were becoming less objective and showed greater influence by advertisers. "I think the quality of the search results has dramatically dropped, so that makes it less of a fair deal for me. I feel like I have to work harder to get impartial search results than I did even a year ago." [P18] At the same time, most participants didn't see other search engines as providing comparable quality, which contributed to an unwillingness to abandon Google specifically.<sup>10</sup> The quality of Google's products was mentioned repeatedly as a key factor in participants' ongoing commitment to the company, as well as multiple participants' usage (and lock-in) with other Google services, such as Gmail and Calendar. One participant aptly described the relationship between ads and search quality: "I think overall [the relationship] is really positive. They've created this world where people are more okay with sharing their information and I can ignore the ads. I don't look at ads in my inbox or on the sidebar. I've never clicked on anything there. On search, I always skip the first three ads, the sponsored stuff, so I think they are transparent in that way. I'm sure there's a lot of stuff that they collect that isn't so transparent, but I still think of them as the better search engine. I would never use Bing. It's kind of like not wanting to use Yahoo! now or graduating from Hotmail. Overall, I think it's Google. They know what they're doing, for better or worse." [P5]

But as another participant observed, the quality of the search results was beside the point given that not just Google but the entire commercial internet was complicit in tracking user actions and using that data to companies' advantage: "I don't think it's fair, because they're tracking everything that we do online, and they're selling that information to advertisers, but again, they have a monopoly where you don't have any other option, because if you go to any other website,

---

<sup>10</sup>The one non-Google user in this participant pool expressed positive opinions about her exchange of queries for search results to Yahoo! (and also had positive comments about Google; she used both search engines, but claimed to use Yahoo! more often). Her bigger concern was about Facebook: "Facebook wants your entire—they want to control your emotional and social interactions with people. Google, you're searching for a very specific thing, and they give you a result for that. Whereas Facebook has to generate exactly what to show you, and what not to show you, which is, perhaps, a more powerful decision that they have to make, in terms of how they affect what you see." [P16]

it's the same thing." [P20]

For other participants, their assessment of fairness was based more on the recognition of the breadth and depth of the information collected about them over the years, and how this affected their experience: "I feel like it's getting a little too personal. I don't remember any specific example, but I remember being shocked a few times, as if they already know you, they're watching you. They never forget. That's the other thing." [P7]

Unlike the other two contexts in this study, with search queries most participants' concerns around fairness emerged from a sense of being known too individually, and not just as an abstract or demographic category. Again, as long as the model of tracking and targeting remains coarse, and Google's actions conform to existing mental models, some participants are fine with this form of exchange: "I'm definitely happy with using Google and definitely prefer it over other search engines. I'm sure other people don't like this, but I like that, because I'm signed in already, I can easily send emails and have it synced together. I don't really mind that it's tracking me to suggest ads or anything like that. From Google's end, I'm sure they're collecting all sorts of super valuable data about what young women like to search. I know they're benefiting from it, but I don't really care. I don't feel like the information that they're getting—it's fine since I'm being offered the service, it's worth it." [P6]

However, several of the conversations took a different turn when I probed the questions related to power, such as whether one could leave the relationship (access to alternatives), whether participants felt they knew as much about the company as the company knew about them, and whether the relationship had changed over time. Participant 19 offered an excellent summary that characterized several participants' perspective on the state of the relationship: "I am committed to them because I think they are the best search engine out there and I think that the suite of services that they offer is quite good, but if something better comes along, I don't have a huge commitment to them if there's a better product out there. Although it would be really annoying to start new with a new email address and all of that stuff." [P19]

It is the last portion of her quote that illustrates the dilemma several participants struggled with when I asked them if they could terminate the relationship: their depth of lock-in with Google's suite of services. The switching costs associated with abandoning the entire platform made it unthinkable (or at least daunting) for many to consider ending the relationship. While switching costs dominated the discussions of leaving the relationship, over half of the participants also mentioned a sense of *resignation* related to their access to alternatives—specifically, that while there were alternatives to Google in theory, in practice the alternatives were either unacceptable (because the quality was inferior) or other services were complicit in engaging in the same information gathering practices that participants disliked.

"I don't know what the alternatives are really, so I think I feel like whatever the search engine providers are doing, they're all likely doing the same thing, so it's like choosing one versus another, and I trust them more than—Firefox, I even feel like gets a lot of revenue from Google for having Google search—so they're all kind of interconnected." [P5]

Another participant discussed how contemplating a decision to stop using Google was much more difficult than choosing to boycott a company by refusing to buy a specific product: "It's because it covers all parts of your life. It covers everything from finding directions, to finding in

the middle of the night your baby is sick, and you just wanna look something up or finding a phone number and an address, or a recipe, anything. It's just everything. It sounds really scary when you say it like that, but it really is everything. It's like a god somehow, right?" [P9]

Others noted that the growing information asymmetry between them and the company was an issue of concern: "I used to really like them and think they were great, and they were doing wonderful things, and I like the library thing, scanning the books. Now I just think they're creepy. What's their mission? Do no evil or something? You think that's great. Then you realize the scope and scale of what they're doing is just bigger, and bigger, and then I don't know, you hear more about Silicon Valley culture and the bro-ness and everything. It just seems like that's not always that great. Then you just kind of start seeing and hearing stories about them scanning your Gmail, and profiling you with different marketing things, and yeah. It's like, why do they want this information? Not necessarily to help society, it's more capitalistic." [P10]

#### 4.3.6.4 Summary

Across all three contexts, the participants reported benefitting from their exchanges. And in turn, most had a clear sense of what benefit they thought the company gained from them. Given this strong articulation of benefit, it is not surprising that many felt that their relationships were generally fair. However, a sense of benefit was not the sole factor contributing to judgments of fairness within each context, and not all participants found the relationships to be entirely fair.

The pregnancy tracking app users articulated a strong sense of receiving a lopsided benefit—to the point where some felt as if they'd received something for nothing. As all of the apps in this study were free, in effect they had, which was reinforced by how little information the participants were required to directly divulge in order to use their apps. At the same time, these free apps are typically dependent on in-app display advertising as well as facilitating third party tracking, either by using a third-party software package in the app, or by selling customer profile data to advertisers or data aggregators. While most of the pregnancy app users thought their exchange was fair, at the same time some expressed concern about being tracked, especially those who had sought to control knowledge of their pregnancy by others. Thus, perceptions of fairness appear tenuous among this group; had the participants been required to divulge more information, or if the third party information sharing/selling practices of the apps had been more clearly disclosed, perceptions might have shifted negatively.

The 23andMe users felt strongly that their exchange of DNA and money for a personalized genetic analysis both provided them a clear benefit and was fair to them. But a strongly influential aspect was the indirect generalized exchange that participants contributed to by allowing their DNA to be used by researchers. Additionally, the knowledge by some participants that increasing contributions to 23andMe's DNA database would improve results for all users impacted participants' assessments. The combination of both obtaining a unique personal benefit while also contributing to the public good is powerful and has strong effects on how participants assessed the service.

Finally, the use of search engines, and specifically Google, revealed the most mixed responses from participants across the three contexts. While the participants gained immense benefits from

using search, their assessment of fairness varied, and appears based on two aspects: the length of the relationship, and Google's practical monopoly over the search marketplace. Participants were also concerned with a perceived lack of 'objectivity' in search results, the breadth and depth of personal information Google has collected about them, and relatedly, a sense of being known too individually by the company.

Unfortunately, this data doesn't provide an answer as to how closely perceptions of benefit and fairness are tied, though we can assume that the less benefit people report a relationship provides, the less fair they are likely to say it is. But as the participants' assessments of their relationship with Google demonstrate, one can be in a relationship where one's benefits outweigh concerns about fairness, but at the same time one can still characterize the relationship, or at least aspects of it, as unfair. And the concerns about fairness are indicative of issues of power, specifically power as characterized by Emerson's power-dependence theory: access to alternatives, and the exercise of power by one exchange partner over another. Here, the concerns are related to Google's functional monopoly, and their exercise of power by altering search results and targeting its users using their own information in ways they either did not anticipate or approve of.

### 4.3.7 Assurances and Trust

Exchange relationships are characterized by risk and uncertainty, and the negotiative aspect of direct negotiated exchange is meant to reduce risk and uncertainty by making the outcome of the exchange binding for both parties. *Assurance structures* also mitigate risk and uncertainty in negotiated exchanges by providing an external or third party mechanism to assure the exchange. According to Molm, they are "[m]echanisms that provide assurance include legal or normative authorities that impose sanctions for violations of agreements or failure to fulfill one's obligations, guarantees such as collateral that protect against loss, warranties that assure certain standards of quality, and so forth." [82] As such, negotiated exchange relationships are marked by less reliance on trust because assurances supplant the need to build trustworthy relationships.

I was particularly interested in the participants' views of assurances in these relationships, as the literature I discovered did not appear to survey the public to discover what consumers thought of as assurances. As I note the methods section, however, based on how participants responded to my questions, I did explicitly ask them about institutional assurances if their answers focused exclusively on mechanisms (as they most often did).

#### 4.3.7.1 Pregnancy Tracking

When asked what factors made them comfortable using the app they chose, most participants mentioned the app's ratings and prominence in the app store they used (Apple's and Google's), personal recommendations made by friends or the media (*e.g.*, lifestyle bloggers), or the fact that they had disclosed so little personal information to the company. Others assumed that they were anonymous to the app, and over half thought a law protected the information they disclosed to the app. Finally, the visual design of the apps and the quality of the content provided also acted as assurances.

App stores act as mediators for consumers searching for apps, providing validation and assurance to customers that the apps listed have been reviewed for adherence to store policies (and thus, conveying a marker of quality in some form). “It seems more legit. It does seem good that it’s in a space. You’re not just downloading it from some random place, and that there were good reviews. Yeah, I think I did look to see if there was any IT (information technology) problems with it. That’s annoying. You don’t want to get an app and it doesn’t work. It seems like it should be vetted if it’s in the Google Play Store, but maybe it’s not.” [P10]

The influence of an app store is threefold. First, Apple and Google previously had divergent approaches to app reviewing, with Apple enforcing strict content and security policies through a human vetting process, and Google only subjecting apps to an automated security review. Since 2015, Google also began human reviewing apps for compliance with store policies, adding an assurance that some smartphone users erroneously assumed existed previously.[59] Thus, appearance of an app in both app stores now correctly implies a level of assurance that may not exist if an app is obtained elsewhere.<sup>11</sup> Second, each platform’s listing of the app within its different categories (*e.g.*, “featured apps” or “popular apps”) reflects an additional dimension of assurance, even if the mechanism behind those categories isn’t always transparent. Third, the ratings of the apps themselves by customers provide another dimension of assurance—a form of social reputation (indirect social assurance) that adds to or even supplants the platform assurances.

Personal recommendations are a form of direct social assurance—a suggestion from a trusted friend, family member, or acquaintance. As one participant described, “You have a baby, and if you don’t want to publish anything about your baby, you should be careful with privacy regulations and everything. Yeah, I’ve just gone through [the privacy policy], but I just had a quick look. Also, I trusted my friend’s word. They told me that they are not disclosing anything. That was one thing that made me okay.” [P8]

As was true in other dimensions of the relationship, the fact that the pregnancy tracking apps in this sample required so little information to use was in itself a form of assurance—that the app wasn’t out to extract a maximal amount of disclosure from the user, and thus could be trusted. “Not entering my health habits makes me feel safe doing it, since they only have information that really, realistically, anybody could look up, except maybe the due date. That makes me feel better about giving them access.” [P2]

“I wasn’t suspicious and I was just comfortable, maybe that I didn’t have to answer those questions, all those questions, and I could still use it.” [P3]

In addition to the minimal disclosure requirements, several of the participants assumed that either they were anonymous to the app, or that when they posted to community forums or answered surveys that they were anonymous in those contexts, which made them comfortable disclosing to the app. “I don’t mind that they’re benefiting from my data, as long as it’s not directly traceable to me, as long as I’m providing anonymous data, then it doesn’t bother me.” [P6]

As predicted, none of the participants in this group suggested a legal assurance structure on their own. When I asked them specifically if there might be a law that protected their information,

---

<sup>11</sup>Note that this is only possible for Android apps; Apple iOS requires that all apps on non-jailbroken devices be obtained only through the App Store.



four of the participants said yes and two said there might be; of this group, four suggested that HIPAA (the Health Insurance Portability and Accountability Act) might offer them protection. In a sense this was a trick question, as there are no laws that protect the information that an app user would disclose to a pregnancy tracking app, even if the information was health related.

“I don’t know about the laws. I would hope there are some. I imagine they’re very rudimentary and there’s a lot of gray.” [P3]

“I think there’s a like a medical privacy law. I’m pretty sure—and I don’t know.” [P5]

“I imagine there are. I don’t really know what the extent covers, and that’s embarrassing because I’ve definitely been trained on HIPAA before, which potentially could be sharing health information.” [P6]

That around half of the participants thought that their pregnancy information might be protected if disclosed to a non-HIPAA entity demonstrates that even if they are not top of mind and are incorrect, some participants were referencing legal assurances at least indirectly as part of their decision to disclose.

Finally, the visual design and the information quality of the app also worked as an assurance to define the app as trustworthy and unlikely to be engaged in shady or illegal practices. Comments about the design reflected an expectation of professionalism as evinced through the quality of the visual aspects of the app, while comments on the information quality focused on the tone of the content, how comprehensive or factually focused the content was on pregnancy development, and whether the content had been physician-reviewed. These elements worked together for many participants to provide assurance that the company was a professional operation and not merely a random individual app developer with no adherence to standards.

#### **4.3.7.2 Institutional Assurances**

When considering to whom they might complain or what action they might take, most participants mentioned they might begin with the company directly, or the app store. Others noted they might simply delete the app from their phones. Despite the (incorrect) assumptions made by several regarding the applicability of HIPAA, only one participant mentioned “the government” and class action lawsuits as possible avenues to pursue. A few suggested they might try to publicly shame the company through social media. Interestingly, nearly all the participants evinced uncertainty about whom they would contact and how they would do so, with two participants suggesting that they either had no idea or weren’t likely to complain. The uncertainty and reluctance appeared to be tied to several themes: expectations around the lack of responsiveness for free apps; the minimal amount of information disclosed to the apps; and, given that, difficulty in imagining what harm the app might do to them individually.

“I would probably contact [the company]. . . I might write a bad review on the Google Play Store, maybe just uninstall it. I can’t really imagine them making me that angry, though.” [P10]

“It would have to be really egregious for me to complain, I think. I would probably delete the app and write a bad review in the app store. If I felt it was really egregious, I would probably complain to the app store, or to Apple. If I felt—if I felt like the content, less related to the

functionality of the app, but that the content wasn't good, I probably would complain to Ovia, like go to their website and complain to them." [P11]

### 4.3.7.3 Genetic Testing

When asked what factors made them comfortable using 23andMe, the visual design of the website and its information presentation were the most cited. Seven of the participants commented on the website's design choices, calling it "bright" and "fun" and nonthreatening—"not seeming Big Brother-esque." [P19] One participant stressed the minimalist design of the logo and the professional presentation of the test kit, expressing that the test kit felt "approachable." These comments align with extant research on the website credibility and trust factors. The presentation of information on the site also acted as an assurance. Participants cited the well-organized, clear, professional, and clean presentation of the company's website and the information they provide about their intentions for the use of their customers' DNA as improving their trust in the company.

Assurance based on personal relationships was another factor, with four participants having received a personal recommendation from a family member or friend. In one case, a close relative bought kits for everyone in the family: "She was so enthusiastic about it and about what it could show us that—she was really excited, and I think that enthusiasm rubbed off on me a little bit." [P19] Two participants mentioned the relationship between Google and 23andMe as an assurance, citing their belief that Google was trustworthy as the basis for their similar assessment of 23andMe. Others also mentioned that they thought 23andMe had a good reputation, which provided a foundation for fostering trust.

The ability to exert control over personal disclosure acted as an assurance, albeit in different senses and interpretations of the word. Participants cited the option to share their DNA matches how they choose, or to not share at all. "I feel like nothing's been forced on me. I feel like everything is opting in, even from the beginning, when it's like, "Do you want your vial of spit to be used for research?" I really didn't care what they did with that vial of spit, but I liked that I felt like I had some control over it." [P19] Others deliberately limited the information they shared with the service as a strategy for limiting personal risk. One specifically noted that he believed that the information that 23andMe gathered was limited in comparison to comprehensive medically administered genetic tests, making participating in 23andMe's services less risky. Others expressed their reliance on the company's data anonymization policies, assuming that it would be difficult or impossible for anyone else to trace their genetic information back to them individually.

To date, 23andMe's has required opt-in consent for any disclosures of one's DNA-related data. While I have not conducted a heuristic evaluation of the platform's disclosure mechanisms, based on the participants' descriptions the consent process appears to be generally well-understood and salient; none of the participants reported any confusion or concerns about it, and all appeared to understand (at least, to their own satisfaction) what they were opting-in to. When any uncertainty was discussed, it was a result of their recollecting a process from memory that they engaged with years earlier.

Formal policies and laws also played a role, perhaps more with 23andMe than most sites given the novelty of the product. Over half of the participants mentioned reading or skimming the com-

pany's privacy policy. One mentioned reading the policy because of a specific issue that was of concern to other participants as well: "My only concern, again, would have been that some of this information might have been shared at some point with an insurer that I might be trying to get insured by." [P18] When asked, half of the participants responded that they believed there was a law that protected their data, though only one knew specifically about GINA (two others incorrectly thought HIPPA provided protection). The other half were uncertain or didn't know: "It was an act of blind faith, really. Unlike me. I was just fascinated by the whole concept. Is there a law?"[P18]

#### 4.3.7.4 Institutional Assurances

When asked to whom they might complain if they had a negative experience, half of the participants mentioned the company directly. Others said they expected that there was government oversight, but only one mentioned an agency (the FDA) by name. Several participants said they would take to social media platforms to harm the company's reputation, and others mentioned relying on the legal system, such as through a class action lawsuit. "Well, if it were really, truly, large, devastating stuff that warranted a class action lawsuit type of thing, where it's like, "They need to change privacy laws because this is so bad," then, yeah, I'd want to get in on that little dance of "legal battle." [P16]

#### 4.3.7.5 Search Queries

"Google shapes how you see the world. It basically has become the foundation of the modern Internet." [P16]

When asked what factors made them comfortable using search, participants' responses to this question garnered a diverse set of responses. Google's popularity and ubiquity was one of the most mentioned: "All my friends use Google, it's just everywhere and everyone uses it so it seems familiar and safe." [P14] Relatedly, several specifically mentioned Google's positive public reputation as a free search service provider, with some contrasting this with Microsoft's motives and reputation as a for-sale software provider. "I think the main thing is I feel like despite some things that I disagree with with the way they operate Google, I feel that they have a pretty high degree of integrity compared to other commercial companies that preceded them, and that came after them. For example, letting people use that for free, and then charging the advertisers. We get a really quality product for nothing. I think there's a lot of integrity. I mean that was a conscious choice on their part. They could have hooked us all on Google, and then started charging us for their search engine, or to use one of their other products." [P17]

Others were made comfortable by the company's business model because of its perceived transparency. "It's definitely gotten more sophisticated over the years where—I think when it first started, maybe there just wasn't awareness that all this data was being collected, but now it's super obvious because the ads are tailored and you're like, "Oh, that's 'cause I looked at this thing earlier." Anyone could figure out that their data is being collected in some way or another, whereas at least initially, that wasn't as obvious." [P6] Again, this rationale rests upon the participants' assumptions that the company's information collection and advertising practices continue to follow

the cause and effect model where ads are displayed based on one's recent search and browsing history.

Some participants based their comfort around the assumption that if Google did something wrong (either deliberately or mistakenly), the company's gigantic user population would both expose any flaws and ensure things were fixed promptly given the leverage they have. Others relied upon an assumed and expected anonymity in the company's database (specifically to third parties, such as advertisers), "I think the fact that [personal information] is, I believe, to be largely used in aggregate. I know that there's targeting, but I know that that's all done by machines, and that there isn't—it's not specifically targeting me, but it's targeting people with these profiles and types of behaviors, and that, on the other side, the people that are paying to do the targeting, again, aren't getting the personal information." [P11]

When I asked participants if they thought if there might be a law that protects the information Google collects from them, only a few thought there might be: one had a vague sense there might be one (there is not), and another incorrectly assumed that the EU's Right to be Forgotten law was also applicable in the U.S. Most either weren't sure or didn't think there were laws that protected their search queries or other personal information. As one vaguely noted: "I remember a privacy law that got passed a few years ago, but I don't know if it specifically protects search queries." [P17] However, several participants did mention that they recalled viewing (mostly skimming rather than reading) Google's privacy policy; this appeared to be a recent phenomenon instigated by Google itself via emails or other notifications as a result of having made changes to its policy, as none of the participants could recall reviewing the policy when they first began using Google.

#### 4.3.7.6 Institutional Assurances

When asked whom they might complain to if they had a negative experience, half of the respondents said they would contact the company directly. Four respondents, reflecting the influence of Silicon Valley throughout the Bay Area, specifically mentioned having friends or other contacts at Google whom they would attempt to work through with the expectation that otherwise they wouldn't expect a response: "I just had a friend who just got employed by Google. Maybe I can ask them how to get a faster route to that." [P9]

Others mentioned contacting the government, specifically the Federal Communications Commission (FCC), the Federal Trade Commission (FTC), and their congressperson. "I don't think I would get a response from Google. I would probably complain to a friend, or friends who work at Google, and ask them who I should—if they know who I should talk to, not in terms of at Google, but in terms of if there's a regulator of privacy. It's not the Better Business Bureau. It's not the FCC, I don't think." [P11]

Several suggestions were made for airing grievances: posting to social media platforms, consumer advocacy organizations (*e.g.*, Better Business Bureau, Consumer's Union), and suing the company directly. Still others noted they wouldn't complain at all because it would be futile: "No. It would be much more difficult to mount any certain criticism or argument against an organization as disbursed and as powerful as Yahoo! or Google. I probably wouldn't even try." [P18]

“I mean, if there was an effective way to voice my complaint, sure, but then again, I think there are other individuals out there who are much smarter than I am who are having similar complaints and concerns. Like Google said, ‘You have an option not to use us. You’re free to use other options.’”[P20]

#### 4.3.7.7 Summary

Having ascertained that the relationships in this study are direct negotiated exchanges, this inquiry provides additional support that the relationships rely on assurances to facilitate exchange. The participants volunteered multiple forms of structural and institutional assurances that they relied upon when deciding to disclose.

It is important to note that these qualitative findings do not provide a measure of how crucial any one assurance is, or a sense of how they interrelate. Nor do they allow us to speculate how the absence of an assurance affects individual decision-making. This analysis is limited to examining participants’ perceptions of assurances, and does not provide insight as to their relative importance in the relationships, or whether they function independently or are interdependent.

What is notable about the structural assurances that participants identified is the range of informal, or ‘soft’, assurances. These included: reputation (direct and indirect social assurance), visual design, information/content quality, anonymity, negotiation in disclosure (*e.g.*, opt-in, minimal disclosure requirements), and adherence of business practices to existing mental models (specifically advertising practices). In addition, as I will discuss in the next section, expectations of anonymity, as well as assumptions that the companies were generally more interested in aggregate analyses of data across multiple users rather than of specific individual users, also acted as an assurance. Formal binding assurances—privacy policies and laws—were not as frequently suggested, and sometimes the participants mistakenly assumed laws existed where none did. This finding suggests that when policymakers and researchers limit their focus to formal assurance mechanisms, such as laws, terms of service agreements, or privacy policies, they are getting an incomplete picture of what people rely upon to engage in these relationships. It also potentially tests the limits of formal definitions of negotiated binding exchange against empirical observations outside a lab.

In terms of institutional assurances, the participants did rely, to some extent, on traditional forms of power such as government regulation and the legal system. But these mechanisms were often viewed as slow, ineffective, or out of reach; for example, the few who discussed suing a company, whether directly or as part of a class-action lawsuit, framed it as either an action of last resort, or as something they didn’t expect would provide them much direct benefit. These participants’ comments suggest a view of government and legal institutions as being slow to respond, somewhat ineffectual, and difficult for individuals to access and navigate, due to complexity or expense. In contrast, participants generally viewed public shaming via the media—specifically social media platforms, not traditional print media—as more apt to get results, and quickly. When I questioned the respondents in more depth about these responses, they expressed the expectation that the companies cared deeply about their public reputations, and that an upswell of bad online media could cause substantial harm. Consumer revolt, then, was seen as a more effective short-term tool than government intervention. It further highlights another dynamic of these discussions—that the

participants generally did not view themselves as having substantial power as individuals in their relationships with these companies, beyond perhaps the power to exit, which in itself was limited in its effects. But with the potential for amplification that social media (and potentially, consumer focused complaint websites) provides, one voice can join with many. It is this coalescence that the participants felt provided them with leverage, as well as, perhaps, the experience of having taken action directly, rather than delegating to an institution.

## **4.4 Privacy: Personal Disclosure, Risk, and Trade-Offs**

After focusing on the participants' relationship with their company or app, I transitioned to a series of questions that asked first about the types of personal information they had disclosed to the company, and then to questions about risks and trade-offs. In keeping with critiques that argue that asking interviewees explicit questions that may prime privacy concerns, I did not use the word 'privacy' in this line of questioning.[20] My goal was to elicit participant perspectives about these themes without explicitly driving the conversation into privacy issues. However, several participants introduced the topic by name on their own accord.

### **4.4.1 Personal Disclosure**

Across the three contexts, the substance and amount of personal information the respondents disclosed varied dramatically. The pregnancy tracking app users disclosed the least, and all respondents reported disclosing the most through their search queries due to the long duration and high frequency of those exchanges. Note that this quantification is reflective of the participants' perspectives; some critics might reasonably argue that, depending on the context, the single disclosure of one's identifiable DNA to a company outweighs fifteen years of aggregated search queries. Yet it was the aggregation of one's search queries, in both the breadth and depth of the information they contained, that generated the most concern among these participants.

Although the respondents' disclosure experiences were overwhelmingly positive, many still reported multiple ways in which they constrained their disclosures to these companies: providing the bare minimum of personal information; neglecting to opt-in to public sharing; and using browser-based strategies, such as private windows, logging out, and clearing cookies. It was this negotiative power—the ability to exert some control over disclosure—that gave the participants comfort with the exchange despite many legitimate concerns. And, inversely, the things that made the participants the most anxious were the areas in which they felt they had the least: third party sharing, automated inferences, and tracking across websites.

#### **4.4.1.1 Pregnancy Tracking**

As discussed earlier, the majority of the pregnancy app users deliberately provided as little personal information to their apps as possible. While nearly all the participants provided their email address and due date to their app/service in order to access the key functionality, there was little consistency around the other types of information participants provided. The range varied; depending

upon each participant's specific need for their app, some disclosed personal health tracking details (weight, height, medical info), tracked doctor's visits, filled out survey questions, and contributed to community forums/message boards (anonymously or with a self-selected user name). As one participant described, "There's a section within the app, where there are questions that they ask about you, and show you how other people have answered that. That's primarily what I've done. I haven't uploaded pictures or inputted milestones or anything like that. It was interesting to go through, especially early on. There were a bunch of quiz questions, and seeing what other people responded to. They also have my email address, and at least my first name." [P11]

Not all of the apps had features allowing users to post pregnancy updates directly from the app to one's various social media accounts, but most of the participants were not interested in using this type of feature, though several had observed friends and acquaintances on Facebook doing so. Overall, there was a conscious choice among seven of the participants to keep their pregnancy information off of social media for a variety of privacy reasons, but foremost: not wanting to openly confront and discuss fertility issues, and a general disinclination to share intimate pregnancy details (and answer intrusive questions). "I've been very public about my life on social media in one way, but when this came along I changed mentality a bit. I don't wanna put [out] too much [sic] baby pictures." [P9]

When asked about the negotiative aspects of the relationship, none of the participants reported any negative experiences related to disclosure, nor any need to contact the company. None reported any experiences that didn't align with their expectations. But when asked whether they thought they were engaged in a one-to-one relationship with their app, some participants specifically mentioned that determining that their app did not share their personal information with advertisers or social media companies was an important aspect of their decision-making process: "They do not disclose any information. It is one of the important things that I care [about]. [D]isclosure was important criteria for me to continue to use the app." [P8] The other respondents were either uncertain or knew that their app shared information about them with advertisers, though many expressed uncertainty about how exactly this happened. "I don't really know, but I doubt it's a one-on-one relationship. I know there are some thing somewhere where you can check if you want the offers or not, and I've been thinking of unchecking those, but of course the fact that it's checked means that it is being shared with companies. I don't know. I just feel like I'm suddenly on more lists than I thought I'd be. . . so I'm not 100 percent sure who is the responsible party, but I didn't really sign up for that many things initially, so I think there's at least a high likelihood that it could've been BabyCenter." [P6]

#### 4.4.1.2 Pregnancy and Advertising

A feature of being pregnant in the U.S. in the 21st century is that often the internet appears to know you are pregnant before most of your friends and family do. In 2012, *New York Times* journalist Charles Duhigg wrote a now infamous story of how the retailer Target claimed to be able to predict its customers' pregnancies potentially before the women themselves were aware.[35] In 2014, Professor Janet Vertesi went to great lengths to conceal her pregnancy from the internet, requiring her to conduct herself online as if she were a criminal attempting to evade surveillance.[114] I

was interested discovering whether any of the participants shared the experience of the Internet marketing machine ascertaining their pregnancies before they had made the decision to disclose. If they had, what was their reaction was to this phenomenon? In most cases, it is difficult to trace back who or what exactly in the marketing machinery ‘decides’ you are pregnant, but it is likely that downloading a pregnancy app and creating an account with the app/company/website will invoke the process. With this group, most had the experience of having pregnancy-related marketing information (direct emails and website ads) targeted at them early in their pregnancies (typically before the three-month mark). Most were unsure when, where and how their status leaked. “It’s really scary, because sometimes I’ll go online and Amazon has diapers on sale on the sideline of my Gmail. It’s like, “How do you know? I have no idea [how the Internet found out], but I’m sure it’s based on my research on diaper brands.” [P20]

Of those participants who had experienced aggressive pregnancy targeted advertising early in their pregnancies, all were annoyed or disturbed by it, and some suspected their apps might have been the source of the reveal. Others conjectured that their searches for pregnancy information or products likely began the process. Most were annoyed at how early in their pregnancies they were targeted for unrequested marketing, typically before they had affirmatively informed their loved ones. Some used strategies (such as utilizing fake email addresses) to circumvent the deluge. “I know how some of this works. I haven’t gotten too many—when I noticed the targeted ads was when I really started actually looking at products online, more than anything else. I know my mom downloaded the What to Expect app, and I think she created an account there, and immediately started getting diaper ads in her email. I did not create an account with What to Expect. I just used the app, without an account. . . I purposely did not do searches related to pregnancy, when I was signed in to my Google account. I would sign out, or I would go into incognito mode.” [P11]

#### 4.4.1.3 Genetic Testing

It was a given that all of the genetic testing participants shared their DNA with 23andMe, as well as their contact information (name, address, email) in order to purchase and receive their testing kit. Thus, I asked the participants about other forms of disclosure, specifically whether they: completed the intake health profile; opted into research studies (and thus voluntarily filled out additional “quick surveys,” which is the primary method for research participation)<sup>12</sup>; and shared their genetic profile on the company’s DNA Relative Finder service. Most, if not all, reported having completed the health profile (no one recalled skipping it, but as several participants signed up five or more years ago they weren’t certain). Six of the participants said they had opted into 23andMe’s research platform, and the other three either did not opt-in or were uncertain. Most participated in the quick surveys, though one participant specifically said she refused to take them because, unlike the genetic analysis, they were correlational and thus not scientifically sound: “I felt like, if I answered them, I would be contributing to bad science, and I felt a responsibility to the world, the community. You’re answering these questions that don’t make any sense, and. . . they’re

---

<sup>12</sup>23andMe’s primary research method is to use self-reported survey questions to associate genetic data with participant responses. At the time of writing, the company also had three disease research programs in progress, which require a diagnosis to join. See <https://www.23andme.com/research/> for more information.



going to carry weight, and people are gonna think, “Oh, if I have this gene, I cry easily,” and that’s ridiculous.” [P1] Four participants did specifically cite their privacy concerns as a reason that they limited their sharing with the company.

Only one participant did not choose share his data on the DNA Relative Finder platform. The other nine had, with a wide diversity of experiences and outcomes, including: comparing data with known relatives (like a parent or sibling); finding distant or unknown relatives (including one participant who was adopted and knew nothing about his genetic background); and, exploring unresolved questions around their paternity.

Almost all of the participants expressed little concern about giving the company their personally identifiable DNA (potentially, for eternity), in part due to the minimal effort required to submit a sample. The test requires filling a vial with one’s saliva, which is easy to produce and infinitely renewable. “I felt like what I gave them was next to nothing, although a lot of people I talked to felt like I gave away the most dangerous information I could’ve given away.” [P1] The company presently highlights the ease of creating a sample on their website as such: “It’s just spit. No blood. No needles.”<sup>13</sup>

Overall, because the participants had highly positive and stable relationships with the company, even considering the earlier discussion of the FDA intervention, little or nothing occurred prior to our discussions to challenge participants’ core motivations or cause them to question their choice to share their data with 23andMe. As discussed previously, the fact that all DNA-based sharing on the platform requires affirmative opt-in drove a high degree of comfort with the process for participants: “I think by opting in to so many things, I felt like they weren’t gonna do a little bait-and-switch, like that if I opted in to this thing, all of a sudden they were gonna do something else. I just remember opting in to so many things that it seemed like everything was very specific. I felt like it was relatively risk-free for what I did opt in for because I thought that there wouldn’t be a huge—like that they wouldn’t do something else with my data.” [P19] Further, the information requested, while extensive, did not violate the participants’ sense of contextual integrity as it appeared to respect the normative boundaries of the service (*e.g.*, related to one’s health).

In probing participants’ expectations of whether the relationship was a one-to-one relationship, several of the participants said they believed that some amount of their personal information was kept between them and the company, but what exactly that information was varied. Some didn’t think that their data was being shared with marketers; one noted that if he discovered that his information was shared with commercial entities he would be very uncomfortable and would want to end the relationship, though he worried that there was no way to take back his information: “No, well God, I hope not. If they did [share], then I think I’d drop [them]. Of course, they have my information already, so it just wouldn’t make a whole lot of sense to drop them. I would be very much uncomfortable with that if they were sharing it [for] commercial [purposes].” [P12] Others who didn’t believe that their personal information was being shared suspected the company was making money off of selling aggregated genetic information, or that insurance companies were funding some of the company’s research. One didn’t believe her information was being shared at all, saying that she trusted the company: “I had a level of trust in their integrity... I don’t think I

<sup>13</sup><https://www.23andme.com/howitworks/>; visited Sept. 19, 2017.

had any misgivings at all.” [P18]

Interestingly, most of the participants said that they didn’t consider their DNA to be particularly sensitive personal information. Of course, given that all of these participants were voluntary users of 23andMe, this is not a surprising finding in itself, though participants’ rationales for why their DNA wasn’t sensitive varied. Some suggested the company was getting useful information from aggregated information and that the individual profiles weren’t exhaustive or even necessarily accurate. One noted that your DNA doesn’t catalogue your past transgressions—highlighting the idea that while DNA may express something about you, it doesn’t capture who you *really* are: an amalgam of your thoughts, emotions, and actions. Another participant echoed the idea that genetic information was a descriptive but not holistic view of her, adding that she was unconcerned about keeping her genetic information private because her DNA wouldn’t grant someone access to the things she did consider to be private: “That’s almost like a science fiction movie a little bit. I’ve given someone my genetic information. I don’t think I ever really thought of it that way, that somebody has a little vial of [my] genetic information sitting somewhere. I don’t think I’ve ever really looked at it that way because that genetic information, while it does show characteristics about me, it doesn’t show me. I feel like there are enough safety—safety might not be the right word. Security. There are enough links that it’s not like by getting that data, by getting my sample, they can also get my bank account information and my address and all those things. I feel like there is enough of a separation that my genetic information is just one in a million samples. While it describes me, it doesn’t describe the things that I try to keep personal to myself. I see it as being one small part of a really large sample. I leave DNA traces on everything I touch. I guess I don’t see it as that. . . personal.” [P19]

#### 4.4.1.4 Search Queries

Compared with the other two contexts, the participants expressed the most concerns with their personal disclosures when discussing search queries. When I asked what they thought they disclosed through their queries, the most common answer was: “everything.” Following that response was a lengthy mix of information types: one’s location history, shopping activity, general interests, personal demographics, name and address, health status, and more. A few 23andMe participants explicitly compared what Google knew about them to their genetic disclosures, suggesting that Google knew far more: “Oh my gosh, Google probably knows more about me than 23andMe does, it’s pretty creepy. They help me find everything that I need: recipes, academic papers, anything. Obscure things, common things, helps me spell things, and in turn it knows all of my everything. Except for my genome, I guess. Maybe it has inferred huge chunks of it . . . What other personal information do they have? Everything. They know where I live, they know my credit card number, they know what kind of hair I have, they know everything. I can’t think of anything that they don’t know honestly except for yeah, my DNA, I guess. You have to really dig down deep to find something that Google doesn’t know about me.” [P14]

“A ton, yeah. Name. Same as the 23andMe list. Yeah, except the genome and more. Probably a lot more too. When I’m sick. When I’m not sick. They can learn a lot, again, if they’re paying

attention. If they specifically wanna know about me, they could probably find out a lot. En masse, maybe not.” [P15]

Unlike the other services in this study, which are narrow in the types of data they collect (*e.g.*, your DNA, your due date, your contact information), search interfaces are unbounded, inviting all manner of inquiries. One of the key differences between what the participants thought about what they disclosed through search queries as compared to the other services was *qualitative*—search queries potentially revealed their thoughts and emotions, as compared to data about them, which was seen as factual (and not as private). In that sense, both the power of search and its inherent risk is its universal generality. Search becomes a repository of literally everything, and the repurposing of one’s queries easily violates contextual integrity given the narrow context under which the information is provided. While Google’s ability to track this information appeared to be a source of concern for many of the participants, there were only a few who stated that there were searches they would not perform on Google or Yahoo! due to tracking concerns.<sup>14</sup> Over half of the participants verified that they were usually logged in to their Google or Yahoo! accounts while searching, and only about a quarter of them were aware that Google stores their search history (and that they could make changes to it).<sup>15</sup> Most reported using a strategy to mitigate their disclosure, such as: erasing their search history (both through their Google account settings as well as one’s browser history), using private browsing features, clearing cookies, logging out of their Google account before searching, and using Google’s account privacy settings. “I do erase my search history and I do browse in incognito mode often as a default for me. I just don’t like having the traces of things that I’ve looked at and it does feel a little safer to me for some reason.” [P5]

As described previously, most of the participants had a (mostly accurate) theory about how the company used their search queries, but from an information science perspective, what is interesting about their answers is the absence of focus on the mechanics of information search and retrieval (*e.g.*, one’s queries being used to improve search engine quality) and instead on their use for efficient advertising and improved individual targeting. While Google is respected for the quality of its search, at the same time their advertising and targeting abilities appear to be highly resonant with the public. This focus may be due, in part, to the fact that it is precisely these abilities that inspire concern; some of the participants discussed specifically how search-based tracking made them feel “creepy,” “scary,” and “sketchy.” “It’s really scary, because sometimes I’ll go online and Amazon has diapers on sale on the sideline of my Gmail. It’s like, ‘How do you know?’ I didn’t physically search for it. A certain brand will pop up more. For example, Huggies would always pop up around my search thing and in my junk mail and spam. I have no idea [how they found out], but I’m sure it’s based on my research on diaper brand, how many diapers, how do you stock up on diapers?” [P20]

One participant, even though she disliked targeting and tracking, explained that she understood it was the result of algorithm and not initiated directly by people watching her: “I think the fact

---

<sup>14</sup>A potential oversight with this question was assuming that participants knew that it was possible to use an alternative search engine that would allow them to conduct a sensitive search anonymously, such as DuckDuckGo. I did not ask each participant if they knew how they could search anonymously. Also, we must allow the possibility that some of the participants simply did not have search queries they were concerned with anyone knowing about.

<sup>15</sup>A Google user’s search history is available at:<https://myactivity.google.com>

that it is, I believe, to be largely used in aggregate. I know that there's targeting, but I know that that's all done by machines, and that there isn't—it's not specifically targeting me, but it's targeting people with these profiles and types of behaviors, and that, on the other side, the people that are paying to do the targeting, again, aren't getting the personal information." [P11] The basis for her and other participants' assurance was derived from an expectation of anonymity—that despite the deeply personal nature of the information they are disclosing to the company, it is the fact of their abstraction, and their value as part of a demographic rather than as an individual, that protects their privacy.

One of the other relational elements that differentiates the search context from the other two in this study is the duration of the relationship—well over ten years for most of the participants. The breadth and depth of one's disclosures to a search engine through search queries over such a long duration is unlike almost any other relationship any of us have with a digital service. Left unmodified, it is accretive and all-encompassing, including both the many trivial items we search for everyday as well as the things we don't want others to see, or that taken out of context may portray us in a bad light. One participant described her distress at first encountering her Google search history, which caused her to modify her account settings: "I also realize that by giving them search queries, I am giving them data about myself. I remember it might have been two or three years ago, one of my friends was like, 'Oh, you can log in to this website or log in to your Google account and see all of your search history forever.' Forever-forever, like I logged in and saw searches I did twelve years ago. That was horrifying because I totally understand that I am giving them that data, but your search history is almost even more personal than your genetic information sometimes. I hide my search history from my husband sometimes. I don't want him seeing what I'm searching for, like 'crazy itchy rash.' He doesn't need to know that. When I realized that there was something linking my email address and all of those searches that I had made throughout college, all these different times, that was alarming to me. That was like, 'Wow.' I didn't realize all of those were collected and archived in that way, so that was frightening." [P19]

Another unique aspect of the long term disclosure relationship with search engines are the changes that have occurred over time. With Google specifically, there has been a long and complex evolution from its launch in the late 1990s as a company that offered a single service and little in the way of tracking its users to the behemoth it is today. While the core interaction has remained remarkably static (inputting a query into the search box), what has changed dramatically is what the company does with those queries, along with the scope of other information it collects and the range of additional services it offers. From a long-term user's perspective, it can be difficult to recall or disambiguate many of these changes as they have been both incremental and opaque. And of course, our own lives are not static—the content of our searches changes as we do, and thus one's search history can be a living document of not only our day-to-day lives but also our individual evolutions as people and the things that matter to us. One participant, who is between 18-30 years old and has used Google for over ten years, described her experience with this evolution: "I would say, I know that there are people out there who keep their privacy a lot more private. Maybe because I started using the internet when I was really young and it just gradually got more and more invasive. At the beginning the internet was so much more basic, you didn't do as much with it, it was really slow, and it didn't have the power that it does now. Then slowly month by month,

it like—it's like when you drop a frog in the boiling water and it tries to escape, versus just slowly cooking the frog." [P14]

#### 4.4.1.5 Summary

From the participants' perspectives, the data they disclosed across these three contexts can be ordered from pregnancy tracking apps as the least sensitive to search queries as the most sensitive, with DNA samples falling in the middle. Despite having the potential to store a range of body and health related data, the pregnancy app participants disclosed very little to their apps, and what they did disclose most considered of minimal value. 23andMe users provided the company with a sample that is uniquely and immutably identifiable, but the participants generally considered it less privacy concerning than the aggregated content of their search queries, which they and the pregnancy app users both characterized as the most deeply personal data in this study. From the perspective of a privacy researcher, on its face this conclusion appears paradoxical. While a collection search queries can potentially identify an individual (and they certainly have before)[11], one's DNA will, in theory, always identify you, forever, and potentially many of your relatives as well. What could pose a greater risk to one's privacy, if one of the key attributes of privacy is one's ability to remain unidentified, or anonymous?

However, the privacy concerns that participants shared were not about being individually identified in these contexts—in general, the information collected did not violate norms of contextual integrity—but instead about being the *object of inferences* and, in turn, being *targeted by those inferences*. What made search queries deeply private but not one's DNA is that search queries contain, among many other things, a map to your thoughts and emotions, your health concerns, your embarrassing questions, your travel plans—just to name a few. Search queries are multidimensional (unless one is a deliberately parsimonious and narrowly focused search user) and are a compass maybe not to one's soul, but at least for many, to both what they presently value as well as what they have valued over time. Most of the participants recognized the aggregation of this information made them vulnerable—not simply to exposure or embarrassment, but to manipulation, that Google's inferences about these queries could allow the company to target them or filter their access to information in ways that undermined their personal agency. While only a few participants suggested specific actions the company was currently taking that unnerved them, there appeared to be a broader sense that Google engaged (or could) in actions that were more sophisticated than the simple cause and effect scenario of searching for a specific pair of shoes, and then seeing those same shoes in ads around the internet.

While the participants' privacy concerns were not focused on remaining anonymous to the company, this is not to say that anonymity was unimportant. In fact, the participants had a nearly universal expectation that these companies would use the data they collected in ways that provides them anonymity, particularly in secondary exchanges. This expectation of anonymity is built upon the assumption that the companies find their primary value in these databases of users in the aggregate, and that any individual contribution each user makes is miniscule. There was also a reliance upon *abstraction*, an expectation that no individualized surveillance was occurring, at least by other people, and that algorithms were responsible for any inferences made by these companies. Several

participants expressed confidence in the fact that they believed computers, not people, were analyzing them. Whether the participants expected true anonymity—meaning, that the data used for analysis was scrubbed of any personally identifiable information—or practical obscurity, meaning that the transaction costs of identifying them among millions of other users provided them with *de facto* anonymity—is beyond the scope of this analysis. But what is clear is that the participants did not expect to be identified and targeted on an individual basis, nor they did not expect the companies to engage in secondary relationships that did the same. These expectations also acted as an assurance structure to the participants in that their assumptions of anonymity and aggregate analysis minimized their risk of disclosure to the companies.

The participants also used multiple strategies across all three contexts to minimize their disclosures to these companies, but the one utilized most often was their ability to negotiate the terms of their disclosure through both optional and opt-in disclosures. This should perhaps come as no surprise, given that since the advent of the commercial internet there have been debates on the ethics and merits of opting out of disclosure versus opting in. Control over one's personal information is often offered as an ideal, without any additional analysis as to what the effect of control is beyond personal autonomy. In these discussions, the participants expressed feeling the relationships were fair, beneficial and more balanced when they had negotiative power. At the same time, undisclosed or poorly disclosed secondary relationships that did not respect the participants' mental models or disclosure preferences posed them concern.

## 4.5 Risks and Trade-Offs

In order to understand the full calculus of disclosure in these relationships, I asked the participants about the risks they felt were attached to these disclosures, and what they felt they were trading off in order to use these services. The participants articulated a variety of risks posed by their disclosure relationships. Most of these risks were directly posed by the companies and were based on contextually violating actions, such as sharing or selling the respondents' personal information to third parties. Governments and hackers also posed risks through the discriminatory or unauthorized use of personal information. The participants also identified social risks posed by unauthorized disclosures, such as embarrassment or exposure of private information.

When asked to consider what they were giving up—trading off—to use these services, the participants (not surprisingly, since they were all current users of these services) felt that the trade-offs were minimal or were more than justified than the benefits they gained. That calculus seemed unlikely to change absent a profoundly negative experience that creates greater liabilities than benefits.

### 4.5.1 Pregnancy Tracking

The pregnancy tracking participants identified few salient risks associated with their disclosures. Due to the low sensitivity of the information they disclosed, none had investigated the privacy settings on their respective apps, and few were concerned with disclosing the same information to

their health care providers or insurance companies. Because the trade-offs inherent to using the apps were the same as using other online services, their use of pregnancy tracking apps didn't stand out.

#### 4.5.1.1 Risks

Because they hadn't given up much personal information to use their pregnancy tracking apps, the risks to privacy the participants identified were primarily driven by use rather than disclosure: secondary uses of information; data sharing with third parties; profiling that could lead to unwanted advertising and tracking; and, data breaches. "They probably have my name. Just having my name and my email it seems like that can kind of lead to other sorts of tracking. It is kind of creepy. I really am creeped out by those ads that follow you around everywhere." [P10]

At the same time, the other half of the participants considered their disclosures to pose low risk, again primarily because they hadn't shared much information, and also because what they did share they didn't consider sensitive. Some noted that they didn't see themselves as a person of interest to anyone; that their data had low or no individual value and was only valuable in the aggregate. "I guess in some ways I feel like the information isn't—maybe this is weird, but it isn't that secretive. I certainly trust any care providers or I probably wouldn't mind sharing it with friends if they were actually interested. I don't know... 'do you wanna' know when we had sex this month?' I suppose it's not that weirdly confidential, but maybe that's strange. When I hear myself say that, it's like, 'No, it is very confidential. It's all about your body,' but it doesn't feel that way, probably because a lot of people talk about conception and fertility and all that stuff, so maybe it's not a taboo subject or something. I'm not scared to share it." [P5]

Multiple participants' comments suggested that they saw themselves as having agency or responsibility for their disclosures, in terms of the resulting risks being their fault or responsibility. "I know anytime I give an app access to my data, probably there's—I mean there's a chance that it could be tracking some of my activity when it's open or selling my information to advertisers. Nothing beyond that. Just because I know better than to enter my personal medical information into a free app." [P2]

Participants' evaluation of risk are also tied to other assumptions: that the tracking that occurs is based on information that they have personally disclosed, rather than tied to their activity or other usage-based or aggregated data; that the tracking is anonymous, or at least, not individually targeted but rather demographically. None of the participants had noticed or investigated whether their app had any privacy controls.

#### 4.5.1.2 Specialized Sharing

I asked participants if they shared more or less with the app than their doctor or midwife, and whether they would be willing to share the same data with their insurance company as they had with their app. The doctor/midwife question was intended to elicit whether the participants thought the disclosure they were practicing with their app was substantively different than with their health care

provider.<sup>16</sup> I asked about health insurance providers because of the concerns that many Americans have regarding health insurers using knowledge about their health to discriminate against them, even after the passage of law prohibiting this type of discrimination.<sup>17</sup> In response to the first question, because most of the participants shared little information with the app, they generally had not disclosed more to the app than to their health care professional. The one exception to this was the few participants who had posted to their app's message boards or community forums, as they were seeking assistance with questions they did not want to discuss with their doctors. Several also said that they had sought out more information from the app than from their doctors—primarily advice or knowledge that they either didn't have enough time to ask about during a typical doctor's appointment, or it felt too ancillary to discuss with her or him.

With respect to their insurance companies, most of the participants were not interested in disclosing the same information to them as they had their apps, though nearly all the participants were unconcerned about any specific risk posed by their access. Because all of the participants were receiving or had received prenatal care throughout their pregnancies, they assumed their insurance companies already knew they were pregnant, and thus most did not view the companies' knowledge as a risk. As one participant described, she was unconcerned with this risk with the qualification that she assumed her data was anonymized: "I wouldn't have thought, 'Oh no, I shouldn't share this cuz it like God forbid, it got out to my insurance company.' I can't even just think of an example of something I would share that insurance company wouldn't like, but I wouldn't let that stop me. When I think about how much sensitive stuff I put on, then yeah, nothing. That wouldn't have ever my mind to not post because of the anonymity." [P4]

#### 4.5.1.3 Trade-Offs

Pregnancy app users articulated few trade-offs in their use of pregnancy tracking apps. This appears to be based on two factors: first, since the overall risk was low, the participants didn't feel as if they were jeopardizing much in deciding to use these apps. Second, the trade-offs participants made were aligned with the trade-offs they were already making in the course of using online services. As one participant said, "With Ovia, you had to create an account, in order to use it. That's why I did do a little bit more research on it. I think, at some point, I made the assumption that, okay, they're a publishing company, so they're selling advertising and, to some extent, user data. It's interesting, going through this, thinking about how much more I value the privacy of that information, and yet, I didn't dig super deep. There's just such a norm, that you don't read the privacy policies, and that there's an underlying trade-

One participant volunteered that she did not understand the fundamentals of the online information ecosystem, and thus couldn't assess what she was giving up. Another said she accepted the tracking she assumed was occurring because the app enabled the building of community around

---

<sup>16</sup>Note that all of the women I interviewed had an established health care relationship with a professional that was overseeing their pregnancy.

<sup>17</sup>While I conducted these interviews after the passage of the Affordable Care Act, which provides protection against discrimination on the basis of pre-existing conditions, enough enmity remains among the public towards insurers that I expected that some participants would still express concern about their access.



pregnancy issues, which might not happen if the app required payment. Four participants said that they would be happy to pay for an app; one in order to avoid the tradeoff of poor security, and the others to avoid tracking and advertising.

## 4.5.2 Genetic Testing

Overall, 23andMe participants identified few salient risks, evinced high trust and confidence in the company, and identified few trade-offs they felt they had to make in order to use the service. As discussed earlier, because all DNA disclosures require an affirmative opt-in rather than an opt-out, there was little focus by the participants on privacy controls, though some participants did report deliberately disclosing less information than asked. Discussions about specialized sharing identified medical insurance companies as a primary threat despite the existence of the Genetic Information Non-Discrimination Act (GINA).

### 4.5.2.1 Risks

When asked, almost all of the participants mentioned potential risks with genetic testing, though none appeared to believe these risks were either immediate or salient to them. None mentioned any concern with their DNA being used by any entity other than 23andMe to identify them. One participant (P15), who had a PhD in bioinformatics, brought a technically informed analysis to his assessment of risk: “You could one day think, ‘Maybe someone will come up with a way to target you.’ It’s really far-fetched given today’s technology, but with the whole CRISPR/Cas9 targeting enzymes, I’m not sure what havoc could be wreaked in the future. There’s some bio privacy concerns, but I don’t know. Right now, it just seems so far-fetched that I’m not worried at all.” [P15]

The most suggested risk was that of insurance companies acquiring one’s genetic information and using the data to discriminate, despite the existence of GINA. Others mentioned government control: one, as a Medicare enrollee, was mildly concerned with the government gaining access to his DNA (particularly after the Trump Administration took office, which occurred in the middle of my interview schedule), and another referenced actions taken by Germany in the 1930s, noting that our government could potentially use the data for future racial discrimination. Futuristic fears around cloning and personally identified targeting were suggested, with one participant specifically concerned with genetic data feeding into a government controlled social credit system like that being developed today in China. Some felt their anonymity was at risk, but potentially mitigated by data aggregation. One participant believed that already there was no realistic guarantee of anonymity, given that genetic testing threatened closed adoptions and could reveal previously unknown half-siblings. Another participant mentioned the potential for family conflict that genetic testing can raise, as well as gaining the knowledge that you may have a genetic disease. “No, to me there’s no risk. People who have histories that they might not know about. That’s probably the people that are at risk the most. Emotional risk or maybe that one test that does tell you that you’re more likely to get cancer or something. Probably those are the kinds of risks that you have to face,

but it's kind of important. You should probably know those things about yourself anyway, so to me, no." [P13]

#### 4.5.2.2 Specialized Sharing

As with pregnancy tracking apps, I asked the 23andMe users if they would share their results with their physician and with their health insurance company in order to explore if any different risks emerged from these two specialized audiences. None reported any concerns with sharing with a physician or health professional, though only a few had done so. As noted above, many did have concerns about the potential for discrimination arising from an insurance company having access to this information. It is interesting to note that as part of these discussions, a few participants volunteered to me that their genetic testing results were "normal" (contained no markers for common genetic diseases), but had the testing revealed an abnormality, such as a fatal genetic disease, they acknowledged they might feel differently about the level of risk the testing posed. Again, Participant 15, the bioinformatics PhD, specified his concern: "I mean given that I don't have one of these 100 percent mutations carrier status or whatever, I'd say then, yeah, probably, at this moment [he would be OK sharing his results with his insurance company]. They would probably [laughter] give me a discount if I didn't have those. There could be a risk if I—I didn't think about it too thoroughly, but there could've been something in my report where it was 100 percent have something that's gonna manifest when you're 45. Yeah, or some manifesting thing, right? That might have been bad, right? Then the insurance company or someone could interpret that and screw you, right? Yeah, I didn't have anything like that, so then I go back to the whole, 'It's too complicated.'" [P15]

Another participant identified the privilege that his normal testing result provided him: "No, I really wouldn't [share with my insurance company]. I'm lucky enough, I guess, that I don't have anything that would directly increase insurance rates. I could imagine if there was someone who had cystic fibrosis, or something, that could be devastating, because then—it's pretty easy for an insurance company to figure that out, if a person has cystic fibrosis, right? I'm in a weird privileged position there. It's kinda hard to have a viewpoint on it. I can imagine if you have to spend all this money on care, and then they go and increase your rates also, it's a little messed up." [P16]

#### 4.5.2.3 Trade-Offs

Overall, participants suggested very few trade-offs they felt they were making to use 23andMe (one participant said coming up with enough spit to fill the sample vial was the worst thing that had happened to him). Several participants emphasized that any potential risk of contributing their DNA to the company was more than outweighed by the direct benefits they received in turn, as well as the potential of their contribution to research: "I'm not bothered by it 'cause I don't feel that I'm being used. I think it's science. They're trying to gather information, and that's okay. I'm happy to contribute to that. I paid \$100.00 or \$90.00 at the time. Doesn't bother me at all. I got some test results that no doctor would ever give me, probably wouldn't've shared with me if they had ordered." [P18]

“I think it’s a pretty good trade. They’re getting data and being able to sell the data. I try not to give them any more than absolutely necessary. I know they’re collecting everything. I figure that’s okay. I haven’t seen anything sinister that I would worry about. I think it’s kind of a fair trade on a certain level.” [P12]

“In the one-to-one exchange, I think, I get a fair trade; but, because they’re a company that’s operating on this big scale—they get this large, pool of data that’s worth way more than what my results would have even been. Then you can start to draw some interesting things about the population. The real value comes from that.” [P16]

Participant 15, the bioinformatics PhD, argued for less anonymization in the company’s pool of data in order to maximize the benefits for the majority. At the same time, he also acknowledged that those with genetic diseases might be negatively affected by this change and potentially require some protection: “I don’t know why people are so uptight. I don’t know why they have to blind all these genetic studies. It just seems like the studies would probably get a lot more out of them if all the metadata about the people went along with them ’cause they’re finding diseases that only affect certain subpopulations or drugs that only work with certain subpopulations. By the time they mix up all the data and pass around the metadata—it’s diluted. I don’t know. It just seems like there’s tremendous value to not [anonymizing] it, and I haven’t seen too many examples where people got screwed because of it, other than these hundred percent mutations that I imagine you wouldn’t want everyone to know about. I hope that people can get a lot of value out of these going forward. Yeah. The more open and the more people do it, the more we’re gonna find out. I think it’ll just greatly outweigh—all those benefits will greatly outweigh the drawbacks.” [P15]

### 4.5.3 Search Queries

While the participants voiced a range of concerns and risks with the aggregation of their search queries, none felt personally threatened to the point of discontinuing their use of search services (or they utilized specific strategies to minimize disclosure, such as using private browsing). Several drew upon the ‘nothing to hide’ argument to assert that they didn’t have search histories that would make them targets of any nefarious forces. And while most recognized they were making trade-offs for the convenience of using search engines, none were sufficiently threatened or dissatisfied enough to alter their behavior.

#### 4.5.3.1 Risks

When asked what risks were present when using their search engine, the participants suggested a broad list. Nearly all of the participants suggested one, including most of the pregnancy app users, who by comparison had difficulty identifying risks in their app usage context. At the same time, it is important to note that these concerns were not preventing these participants from continuing to use Google or Yahoo!. Some were concerned with the potential for their information to be stolen or compromised by hackers (including the impact of one’s search history being made public). Others believed that there were potential risks if their queries were to find their way into the hands of the government. A few were concerned with the commercial uses of their personal information,

specifically that their queries were being used to manipulate them for the purposes of serving ads and selling products. As one participant described, “I think there’s a lot of risk for all of us, because when a private company who knows your color preference, when you are most likely to buy, or what you need to buy, or let’s say I’m expecting. I’m researching a lot of baby things. For the next twenty years, they’re gonna continue to sell me baby things as my baby ages. That’s scary, because you’re constantly having someone sell you things. I think that’s really something—on NPR, they said subconsciously we don’t think about it. We just do it. These companies probably know me more than I know myself.” [P20]

Others risks mentioned included: individualized information discrimination (*e.g.*, filter bubbles), employer and/or medical insurance company access and discrimination, risks of others (especially people they know) gaining access to one’s search history, and a sense that the use of one’s queries to generate targeted advertisements constitutes a loss of privacy. “I’m sure there are [risks]. I don’t know, if there was some kind of dictatorship situation and they wanted to find all of the people with a certain characteristic and then eliminate them. That would probably be bad to have so much information. If someone in power had access to Google’s information and they wanted to find all the people who read a certain negative article about Donald Trump or something and then punish them, then that would yeah—they got me. Because it’s associated with your personal contacts, who you email the most they know who my mom is, they know who my dad is, they know where I live.” [P14]

“I wouldn’t want my friends to be able to watch what I’m doing. A lot of that goes to the inference stuff but also just general privacy. Now we’re getting into just general privacy. I just don’t want people to be in my business all the time and be in my head.” [P15]

Some participants commented that the personal risk to them of using search engines was either extremely low or simply theoretical. One participant specifically trusted Google not to do any evil. One discussed that her use to the company as a source of revenue (through product advertising) presented a risk to her wallet, but ultimately Google’s targeting was helpful to her: “I guess for me as a person I don’t think there’s necessarily a ton of risks out there. It’s more just I’m helping them with their marketing and business development. I think for me personally I don’t feel like they’re gonna do anything [to me]. I guess at my own peril I’ll see ads of things that I might want to buy, because they nail me. That’s helpful, because it kind of whittles down my searching.” [P10]

#### 4.5.3.2 Nothing To Hide

To put the participants’ articulation of risk into perspective, it is important to discuss that at the same time many of them articulated specific concerns, over half of the participants felt as if there was little risk that they would ever be personally targeted. Overall, these participants expressed the belief that their search data did not contain any information that could be harmful to them personally. Most believed that they, individually, were not at risk of being targeted because there’s simply no reason for them to be targeted among the many users of Google search. “Maybe because everyone using it. I’m not that important for them to do anything about it.” [P7] Three did not believe that their search data was important or extremely personal, and two were reassured by the fact that their search queries are algorithmically processed. One was more concerned with the

social risk of their search history being exposed rather than any threat from Google. An exchange with P16 provides an illustration of these issues:

**P16:** Realistically, there isn't a specific person at these companies who is like, "I'm gonna look at what Johnny looked up on this day," or whatever. It's an AI [artificial intelligence] algorithm that's sorting through all stuff. Now you get into weird things like, "Okay, if AI becomes conscious," and stuff like that. It's weird, but we're not close to that. You're telling this computer; the computer doesn't judge you; there isn't someone sitting behind another computer that's like, "Ah, that's what you're looking at!" You've got this interesting one-way relationship.

**Jen:** Does this feed into what makes you comfortable using search engines, or just Google in general, example? You're the one needle in the very giant haystack.

**P16:** Yeah, absolutely, and knowing that probably there are a billion people who search for just as obscure things as I do. I like to think I'm special, but, you know, [laughs] 'aggregate.' The people probably search for even more bizarre stuff than I would ever type into Google. Generally, I think, my searches are pretty tame. I can imagine it can say a lot about a person. It can also be misleading. I might search for something, having a certain context in my mind; and then this algorithm might skew it, and take some other meaning from it, because I'm not necessarily typing everything out, right? I could type in something that could be interpreted differently, even given the context of what I've searched before. Again, it's not like there's a certain person who is judging me for any of that stuff.

As P16's excerpt implies, one of the motivating features of the nothing to hide argument is the individual's assumption that he or she is not engaged in any activity that could draw the attention of law enforcement or other authorities. Another participant, who had one family member who had worked for the National Security Agency (NSA), and another in the federal executive branch in cybersecurity with indirect exposure to the national security apparatus, expressed a similar sentiment: "The NSA's not trying to profit off of me. The NSA, its broad mission is to keep me safe. Maybe because of the environment in which I was raised, [I know] they don't care about me. I've got nothing to hide from the NSA, so I don't care what they—they're not gonna be reading my Gmail conversations because they don't really care what we're having for dinner on Tuesday. They don't care." [P19]

As I will analyze in more depth later, the nothing to hide argument rests upon several interwoven assumptions: the obscurity provided by very large datasets; the privilege of knowing or assuming that one's beliefs, interests, or identity are not of interest to those who might cause you harm; and, the sense of privacy gleaned from a computer, rather than a human, reviewing one's data.

### 4.5.3.3 Trade-Offs

Despite the concerns participants did have—being tracked or profiled; receiving biased or manipulated search results; the asymmetry between individuals and Google specifically with regards to one’s personal information—they presently do not outweigh the benefits participants gained by using Google. The benefits participants listed—convenience, information quality, gains in personal efficiency and productivity—were valuable and not easily gained from other sources. Which, of course, evokes another shared concern: that Google effectively is a monopoly, and that no realistic alternatives exist for either its search service or its suite of services.

Based on participants’ comments, it appears that Google will remain their default search engine as long as they continue to deliver a quality product, do not experience a high profile privacy breach or other trust-jeopardizing event, and stay on the side of finding useful ways to mine its users’ personal information without treading into overtly creepy waters. The efficiency gains for individual knowledge seeking and access are too great to ignore: “I guess I realize [in] the cost benefit analysis that there is definitely some potential risk, but there are benefits. It’s just part of the modern society we live in, that if I just stopped using it, it would be—I don’t know. I am very efficient with it. I am a very good researcher. I feel like it’s a lot easier. I guess I’m old enough to know when we had card catalogs and used encyclopedias and stuff, that it is very convenient.” [P10]

At the same time, many of the participants are keenly aware of, and are not pleased with, the trade-offs they are making to use the service. Should an alternative arise that delivers similar quality without many of the downsides expressed by participants, in theory such an alternative would be appealing. Of course, one does exist—DuckDuckGo, a search engine premised on delivering results without tracking individuals—but the challenge for such a service is not simply delivering competitive search results. Rather, they must convince the public that the personalization of one’s online experience that Google provides, both through search as well as its other products, carries with it substantial negative costs, even when those costs are largely invisible and accretive. This is of course not their only challenge; one participant summarized the enormous hurdle of creating a product that achieves sufficient network effects to sustain ongoing improvements as thus: “I think, as an individual, what I gain back [from Google] is fantastic! When you use services like DuckDuckGo, the search results just are not as good. It’s because Google has been paid to make those awesome algorithms to give you the best results, because those algorithms change consumer behavior. There’s an incentive, then, to make a better one. It’s weird. Even though you have stuff like the NSA, who has all the data, right? Google, who maybe has most of the data, but not all of it, ‘cuz they can’t access certain things that the NSA does, Google has the better algorithms, because they’re being paid to make better algorithms. There’s this sort of weird ‘fair trade.’ You continuously get better and more improved services, in exchange for your data.” [P16]

### 4.5.3.4 Summary

The risks that the participants identified fall into three general (and neither exhaustive nor mutually exclusive) categories: targeting and manipulation, exposure and embarrassment, and violations of

contextual integrity.<sup>18</sup>

Concerns with targeting and manipulation encompassed uses of personal information to display ads, to sell products, to draw inferences and act on them, and the filtering or biasing of information. Exposure and embarrassment included revealing one's personal information to others, exposing embarrassing information, or having one's information compromised, particularly by hackers. Violations of contextual integrity included the uses of personal information by the exchange partner beyond the original intent of collection, particularly secondary uses such as sharing with an ad network.

Comparatively, the assessment of risks were lowest among the pregnancy app users, and were the highest and most varied when assessing search. With pregnancy tracking apps, the risks were not substantively different than other typical uses within the mobile app ecosystem. The participants did not identify anything extraordinary as compared to using other mobile apps. Many risks were identified by participants when discussing search queries, and it is notable that these were more focused on unauthorized access of one's search history, as compared to Google specifically causing harm (though there were concerns with Google as a bad actor). But the most interesting discussion revolves around 23andMe and the risks participants suggested.

Because direct to consumer genetic testing (DTCGT) is arguably still in its infancy, I would expect that participants would have limited knowledge of the possible privacy risks DTCGT poses. From a privacy perspective, the assignment of one's DNA to 23andMe raises many red flags. In particular, the potential secondary uses of 23andMe's database of DNA is likely a cause for greater concern than any direct action the company may take with its customers' data. These uses include: access by governments, law enforcement, or potential hackers, who arguably pose greater threats than the current marketing uses of the data. For comparison, social security numbers are difficult but not impossible to replace; one's DNA is immutable. Given that it is difficult to anticipate today how our DNA might be used in twenty or thirty years, allowing a private company to sequence and store this data in perpetuity poses intangible risks. While there is a compelling argument to be made for the need of widespread participation in DNA analysis for public health, 23andMe is a private company that makes no explicit guarantee that its data will be used to benefit either its contributors or the public broadly. However, it is the intangibility of the risks, and our human limitations to both anticipate long term consequences and engage in hyperbolic discounting that make it difficult for anyone, including privacy experts, to predict how these issues will ultimately play out. But perhaps it is the strength of the assurances present in the exchange relationships with 23andMe that mitigate enough of this uncertainty for participants to proceed with confidence.

With respect to trade-offs, across all three contexts the present benefits outweighed the aspects that participants felt they had to compromise on in order to use these services. Short of catastrophic failures on the part of these companies to protect their users' information, an alteration of this proposition would likely require changes in relationship terms that result in more benefit for the companies and less for the users. There are hints of issues among these participants—say, stricter disclosure requirements, default opt-ins for information sharing—that could tip the balance away

---

<sup>18</sup>There is potentially an argument to be made here that these all fall under violations of contextual integrity. I am separating them into three categories as I believe the other two categories delineate a more specific intent.

from a perception of mutual benefit and towards a growing sense of unfairness that the relationship benefits the company more than the individual. To some extent, trade-offs are impacted by present norms and practices, and a race to the bottom towards exchange terms that grossly benefit the company at the expense of the individual has the effect of benefiting the negotiation position of the companies.

## **4.6 Summary**

This chapter reviewed my findings from the twenty qualitative interviews I conducted across three information exchange contexts. In Chapter 7, I will discuss in depth the implications of the results and how they fit in with the larger focus of this dissertation.



# Chapter 5

## Experimental Methods

In this chapter I review the methods I used to conduct two online survey experiments. I discuss the rationale for the experiments, the study design, dependent and independent variables, development of the survey instrument, study procedures, and the analytic strategy.

### 5.1 Introduction

I created an experimental survey to measure the effects of negotiation and assurance structures on perceptions of trust, fairness, power, and privacy. I drew upon the framework of exchange (as defined by social exchange theory) to examine the relational aspects (and outcomes) of an information exchange between individuals and a company.

I operationalized social exchange in the experiment by creating a direct negotiated exchange to isolate these dynamics. The exchange occurs between two actors: an individual (the study participant) and a fictional company. It is information-based, consisting of the disclosure by the individual of personal information in order to obtain access to an information-based service offered by the fictional company. The individual's personal information flows to the company, and benefits based on this information flows back to the individual. I developed a vignette featuring a fictional company (the 'High Tech Device Company') selling a product (the WearMe wearable tracking device, similar to a wristwatch, for \$99) that requires an explicit disclosure of personal information by the customer in order for them to use and experience the benefits of the product.

In the first study (Study One), I isolate the negotiative aspect of the exchange. The negotiation process consists of the respondents' acceptance of the WearMe Device Terms of Use—a randomly assigned experimental manipulation statement that either presents an offer of optional negotiation based on a more equitable relationship ("You can choose the types of data about your body and your physical location the WearMe device will track."), or a mandatory, 'take-it-or-leave-it' form of negotiation based on a less equitable relationship that favors the company ("You must allow the WearMe to track all the data it requests about your body and your physical location."). In theory, exit (refusing to use the device or to disclose the requested information) is an option, though I did not present it in the study. I examined the effect of the random assignment on the key outcomes of

participant perceptions of trust, fairness, power, and privacy.

I modeled this exchange to maximize its ecological validity; when we choose to transact with a company, especially in the online context, we typically do not actively ‘negotiate’ in the manner traditionally implied by the word. Meaning, there is no haggling, offering/counter-offering, or even direct discussion with the company. Instead, we are usually presented with a Terms of Service agreement or similar document, which most consumers simply accept or sign with little scrutiny given that they cannot engage in actual or meaningful negotiation. These relationships are based on ‘take-it-or-leave-it’ terms: the only way an individual consumer can reject them is to vote with her feet (exit) and refuse to engage with the company, assuming they have options to obtain the resources they value. In the instances when consumers are given negotiation power, it is often in the form of exercising choice within boundaries created by the company, such as through customizing privacy settings, or by designating some or all of the information requested as optional. This is why I describe the relationships as “more” or “less” equitable in this study—the more equitable optional disclosure condition is still not, by definition, an equitable relationship.

Studies 2A and 2B build on Study One by using the two disclosure conditions as controls for an investigation of the effect of assurances. I introduced three assurances—one formal, two informal—and measured their effect on the key outcomes of participant perceptions of trust, fairness, power, and privacy as compared to the control conditions. Study 2A used the mandatory disclosure condition as the control, and then appended each assurance to the control statement for a total of four conditions. Study 2B followed the same design except the optional disclosure condition was used as the control. In both studies, I examined whether the assurances had a positive effect on the dependent variables as compared to the control conditions, with the assumption that the assurances would mediate some of the risk or uncertainty posed by the disclosure conditions.

The experiments were reviewed and approved by the U.C. Berkeley Office for the Protection of Human Subjects (#2015-12-8189).

## 5.2 Study One: Negotiated Relationships

Study One examines the effect of framing the negotiation in a direct negotiated exchange between an individual and a company actor as optional or mandatory. The purpose of the exchange is for an individual to obtain access to an information based service (a wearable device for tracking health and fitness) by providing the company with both personal information and payment of a fixed cost. The respondents were randomly assigned to one of the two negotiation conditions. In both cases, the company dictated the terms of participation and the respondents accept the terms.

This study builds on the 2012 study by Brandimarte, Acquisti and Loewenstein which established a “paradox of control” for privacy: as individual control over personal disclosure increases, one’s willingness to disclose personal information also increases, and inversely lower individual control results in less disclosure of personal information.[19] My Study One attempts to replicate their core finding by examining whether increased control over disclosure translates into increased willingness to disclose. However, this study focuses explicitly on the effects of power on personal disclosure as framed in a more equitable power relationship (optional disclosure) as compared to a

less equitable one (mandatory disclosure). In both studies, the recipient of the individual's disclosure of personal information is a company providing an information-intensive consumer service, but Study One differs from Brandimarte *et al* in that I evaluate the individual's personal disclosure in response to a manipulation of the terms of the exchange, rather than focusing on access and use by others. With this approach, I attempt to examine an individual's decision to disclose their personal information within the context of their relationship with recipient of the disclosure (the company) rather than as a decision made independent of this relationship. I also address a concern I have with Brandimarte *et al*: that the way in which release and access were operationalized in their study was ambiguous and left respondents to infer whether the action (publishing personal information) and the access (either a social network comprised of fellow students or faculty, or scientific researchers) had any direct consequence (as and such, any risk). In Study One I specify both the terms of release and access to avoid this ambiguity (in both cases, the information is limited to use by the fictional company).

As discussed earlier in this dissertation, I define the exchange of personal information by an individual to a company as a form of social exchange: a direct negotiated exchange (DNE). In DNEs, actors exchange resources (in this case, personal information for access to a service) through a process of negotiation that reduces the risk and uncertainty that comparatively accompanies reciprocal exchanges, which by definition have no prior agreements. The benefits of DNE are jointly negotiated and thus bilateral (though they can be unequal), and the actors are assumed to understand the terms of the exchange.[81] In this study, the negotiation process consists of the respondents' review and acceptance of the WearMe Device Terms of Use—a randomly assigned experimental manipulation statement that either presents an offer of fine grained negotiation (“You can choose the types of data about your body and your physical location the WearMe device will track.”), or a ‘take-it-or-leave-it’ negotiation (“You must allow the WearMe to track all the data it requests about your body and your physical location.”). This negotiation is intended to resemble the negotiation process that most consumers face when selecting a new product or service: their agreement (or acquiescence) to an offer made by the company, where the consumer's choice is to accept the offer as it stands or decline to use the product or service.<sup>1</sup>

SET provides a theoretical framework for examining the outcome of this negotiation by identifying the structural components that affect the exchange itself: trust and power. Trust is an emergent phenomenon that arises in response to uncertainty and risk, which Molm defines as the “expectations that an exchange partner will behave benignly, based on the attribution of positive dispositions and intentions to the partner in a situation of uncertainty and risk.”[78] Trust can be supplanted or bolstered by assurances that seek to provide external guarantees to the exchange in order to mitigate risk. Ultimately, without trust or assurances, exchange relationships cannot flourish. The same is true for disclosure relationships; without some form of trust in the exchange partner or an external assurance, most individuals would be reticent to disclose personal information in any relationship, let alone with unfamiliar partners.

---

<sup>1</sup>While the experiment proceeds with the direction that the respondent has chosen to accept the offer, as discussed earlier I include manipulation checks to establish whether a respondent would in fact use this device given the terms of the offer, as well as whether the respondent has used or would use a wearable device.

As discussed in depth in Chapter Two, in SET power is articulated in terms of dependence: “an actor is dependent on another to the extent that outcomes valued by the actor are contingent on exchange with the other.”[80] This study focuses on measuring respondents’ perception of power at the initiation of the relationship and their assessment of the company’s power over their disclosed data. Generally, an exchange partner can exert power over another when the first partner has more options for exchange than the second. According to Molm, “[i]nequality in exchange benefits arises from. . . unequal rates of exclusion from transactions, and unequal divisions of profit within transactions. The former drives the latter. More powerful actors benefit from a lower probability of exclusion and from the greater profit they receive when they are included.”[81]

Explicit examinations of power differentials between individuals and companies is an underexplored concept in privacy scholarship, yet one that is arguably salient when examining individual disclosure to companies or organizations. Extant studies have noted exercises of power indirectly, such as through the application of defaults that maximize personal disclosure[50], information asymmetry between individuals and companies[111], and the limitations of privacy policies as a device for communicating company privacy practices[57]. Most research (and theorizing) treat individuals and companies as if they are equal actors, when in terms of negotiation power, they are likely anything but. In many consumer contexts, individuals may have access to multiple alternatives such that they have substantive choices in deciding with whom to exchange their information for a service. But when the majority of information-based services follow the same consent model, the actual negotiation power of individuals in the marketplace is reduced to ‘take-it-or-leave-it’ as individuals have no opportunity for substantive negotiation. As Molm notes, “[i]n negotiated exchange, the response to exclusion increases power use by increasing the inequality of the negotiated agreement and increasing the powerful actor’s benefits.”[81] As long as individuals are unable to represent their interests in the negotiation process company actors can demand unequal agreements. Further, in information exchange relationships the power of the company relative to the individual may increase over time as the company collects greater amounts of information about them. This may have the effect of increasing the utility of the service to the consumer (e.g., personalizing a service such as Google to increase relevant search results) while at the same time increasing the individual’s switching costs.

### **5.3 Study Two: Assurances**

Study Two builds on Study One by using the two negotiation conditions as controls in two separate experiments with assurance structures. According to Molm, “[m]echanisms that provide assurance include legal or normative authorities that impose sanctions for violations of agreements or failure to fulfill one’s obligations, guarantees such as collateral that protect against loss, warranties that assure certain standards of quality, and so forth.”[82] The role of assurances in exchange relationships are to mitigate risk and uncertainty by providing a third party mechanism to assure the exchange. For example, escrow services are a form of assurance, where they act as a trusted custodian between two parties engaged in a financial transaction to ensure that both parties meet their specified financial obligations.

In this two-stage study I tested the effect of assurances against both of the negotiation conditions from Study One. Study 2A deployed the mandatory disclosure condition as the control, while Study 2B utilized the optional disclosure condition. In each study, respondents were randomly assigned to one of four conditions: the control condition or a condition paired with an assurance statement (mandatory disclosure + assurance, or optional disclosure + assurance). I focused on two general forms of assurance: formal and informal assurances. I define formal assurances as those provided by institutions such as government or other organizations, and include laws, regulations, industry standards, and similar. Formal assurances carry with them a form of sanctioning power or consequence, such as a civil or criminal violations or penalties. I define informal assurances as those that provide assurance without sanctioning power. This is a broad category, including social assurances (reputation, recommendations, etc.), and structural or design assurances, such as anonymity, obscurity, or design-based credibility.

## 5.4 Experimental Design

I designed a series of between-subjects experimental survey studies that manipulated the aspects of negotiation and types of assurance independent variables in order to examine their effects on respondents' perceptions of fairness, benefits, power, trust, and privacy in an exchange relationship. The studies presented all respondents with a single vignette (216 words in length). The experimental manipulation consisted of a text statement that was appended to the vignette. Respondents were randomly assigned to a condition upon commencing the study; their assignment condition dictated which manipulation statement they viewed. After reading the vignette and the manipulation statement, all respondents were given a single survey consisting of three parts: vignette-specific questions; general survey questions; and, demographic questions.

Similar to the qualitative interview study, I tested my hypotheses using a context that involved an explicit exchange of personal information with a company in order to access to a service. I created a fictional personal health and fitness wearable tracking device (reminiscent of a FitBit, which I called the "WearMe"), and a company, the "High Tech Device Company." I chose a health and fitness wearable as the context for the study for the following reasons: there was a clear disclosure of personal information required in order for the device to function; the personal tracking device category is a popular and growing area of consumer electronics but without dominance by a single company<sup>2</sup>; and, the premise of the exchange was likely widely understandable by most respondents. I was particularly concerned about the last point; I did not want to choose a technology that was too esoteric for most people to grasp, especially if they had never used such a device before. In order to provide as much clarity as I could, the vignette provided a high level description of how the wearable device worked. The survey also included questions to record whether: the respondents had ever used or were currently using any wearable devices; how likely they would

---

<sup>2</sup>As I discovered with search engines in the interview study, the search category is so utterly dominated by Google that a general study of search becomes Google-specific. While FitBit is one of the most popular wearable tracker brands, there are several competitors in the field and thus the category is not so severely circumscribed by the actions of a single company.

be to use *any* similar type of wearable device; and, how likely they would be to use *this* wearable device described in the study.

The structure of the experiment underwent several phases of design and testing. After choosing the topic of the vignette and developing an initial set of questions, I created a functional version in the survey platform Qualtrics. After an initial test and refinement among friends and family, I then recruited ten UC Berkeley School of Information master's students to participate individually in a talk-aloud study, in which I asked them to sit with me and take the survey while talking through their questions, concerns, and rationales for their answers. This phase was iterative; after completing a few, I made changes to the survey, presented it to new participants, and repeated the cycle. The master's students were paid \$20 for approximately thirty minutes of their time.

Based on the feedback I received, I revised the survey again and then conducted a formal pilot run with fifty participants on Mechanical Turk. Based on the data from the earlier tests, I calculated the average completion time to be around twenty minutes, and calculated a rate of \$3.00 per survey completion. All formal pilot and actual study participants were paid at this rate.

After additional revisions, I ran the first version of the study (Study One) in July 2017 on Prolific Academic, with 110 participants (approximately 50 per condition). However, after reviewing the data from that run, I identified a number of concerns: my participant pool was not large enough to observe an adequate effect size; one of my manipulation check questions as written was flawed; and the visual design I had implemented did not sufficiently draw attention to the experimental manipulation. After addressing these issues, I re-ran Study One on Prolific Academic in September 2017 with a pool of approximately 250 participants (approximately 125 per condition). An initial review of the data resolved my concerns.

Study Two tested the effect of assurances in two stages: Study 2A included the mandatory experimental condition as the control and then matched with three assurances. Study 2B ran the optional negotiation experimental condition as the control and also matched it with the same three assurances. Both studies were run (independently) in November 2017 with approximately 520 participants (125-130 per condition). I made some minor changes between Study One and Study Two: the recoding of several answer choices that were inadvertently reverse coded; I dropped a series of questions about risk and anonymity and replaced them with two questions focused on eliciting the participants' understanding of the term 'standard industry practices' in relation to data protection; I added two general privacy questions related to anonymity and the 'nothing to hide' concept. All other aspects of the survey remained the same as in Study One.

## 5.5 Study Assumptions

Several factors in the vignette scenario were fixed in order to aid in isolating the main effect of the manipulation and to avoid potential confounding variables: the type of data the device collected, the price of the device, its benefits, and the potential privacy risks it presented to users.

### 5.5.1 Data Collected

The information and data featured in this study consisted of identifiable and non-identifiable personal information, including information that could be used to infer one's identity if aggregated with other data points, as well as body data, and precise location data.

Data directly collected by the WearMe device consisted of:

- body data: heart rate, the number of steps the respondent walks or runs, body temperature, sleep cycles;
- location data: the respondent's precise physical location throughout the day;

Identifiable and non-identifiable information collected by the company, in the service of creating an account and purchasing the device, included:

- name
- age
- gender
- email address
- mobile phone number
- home address

**Device price:** In order to reduce ambiguity regarding the company's incentives for information collection, I formulated the exchange as a fixed-price trade by demonstrating that the business model was not based on the sale of personal information. In both studies the price was fixed at \$99.00. I selected this price after researching the current prices of similar wearable devices and obtaining feedback from pilot respondents who owned or were interested in purchasing wearable devices. My goal was to select a price that was plausible for the fictional product and wasn't either too high or too low, which might introduce assumptions or speculation that could confound the study. The price of the device was displayed in the manipulation statements as follows: "WearMe Price: \$99.99."

**Benefits:** The vignette described the respondents' exchange benefits as personalized health and wellness information obtained through the WearMe device and the accompanying mobile app. The company's benefits were implied as the income received through the sale of the device as well as access to the respondents' personal information to build its personalized recommendation product.

**Privacy Risk:** In an effort to mitigate any confounds posed by the risk the device posed to a user's privacy, I explicitly defined how user data was protected and addressed secondary uses:

*Data security:* I included a statement about the company's data protection policy that was deliberately general and emphasized that the company was following common practice ("HTDC collects

and stores your data in accordance with standard industry practices.”). In the pilot studies, some respondents raised questions about how the company protected their data when this statement wasn’t included. However, I did not want to ascribe specific practices to the company that could confound or interact with the assurance statements as independent variables, or the questions that ask about assurances.

*Secondary uses:* I included a statement clarifying that the company did not use any of the data it collected for advertising (“HTDC does not use the personal data the WearMe collects for advertising purposes.”). Without this, the pilot respondents expressed uncertainty when evaluating questions related to trust and privacy without having any sense of whether the company engaged in secondary use practices. In order to keep the vignette simple, and to avoid portraying the company as overly conscious about privacy, I did not include any additional secondary uses in the vignette.

## 5.6 Independent Variables

Two independent variables were manipulated in the studies the: *form of negotiation* for Study One, and *assurances* for Study Two.

### 5.6.1 Study One

Negotiation was operationalized as the respondent’s ability to control her disclosure of personal information to the company and the device. There were two conditions: a mandatory disclosure condition, and an optional disclosure condition. The conditions were appended to the end of the vignette and entitled “WearMe Terms of Use.”

- *Mandatory Disclosure Condition:* You must allow the WearMe to track all the data it requests about your body and your physical location.
- *Optional Disclosure Condition:* You can choose the types of data about your body and your physical location the WearMe device will track.

In pilot tests, I titled the manipulation statements as “WearMe Offer”, but changed the title after feedback from respondents that the term ‘offer’ was ambiguous.

### 5.6.2 Study Two

Assurances were operationalized as independent statements attached to the disclosure conditions. The statements were intended to mitigate the risk of disclosure.

- *Formal Assurance (Legal):* The data the WearMe collects from your body and about your location is protected by law.
- *Informal Assurance (Indirect Social):* Over a million people use the WearMe to improve their health and fitness.



- *Informal Assurance (Structural)*: Your data is anonymized when used for any purpose other than to provide you with WearMe recommendations.

The formal assurance specifies that the data the device about the participant's body and location is protected by law. It is unspecific about the law itself; based on my qualitative interviews, in cases where information was protected by law, most of the participants who were aware that salient laws existed could not specify whether the legislation was enacted at the state or federal level, or the protection provided. Assurance is provided through the implied sanctioning power held by a state or federal authority to punish the company should it violate the law.

One informal assurance is based on indirect social reputation—specifically, a marker of social reputation with an indirect relationship to the individual, as opposed to direct social reputation, such as a reference from a friend. In this study, I operationalize this assurance as a positive reputation based on the fact that many other people have chosen to use the product. Assurance is provided by the fact that because a large number of people have presumptively evaluated or experienced the product and found it credible or beneficial, the relative risk must be low.<sup>3</sup>

The other informal assurance is structural in nature, operationalized as privacy protection or obscurity through anonymity. Assurance is provided by an explicit design choice that affirmatively seeks to protect individual privacy and minimize disclosure risk to the individual.

## 5.7 Dependent Variables

There were four dependent variables measured in this study: respondents' perceptions of trust, power, fairness, and privacy. Trust, power, and fairness are theoretical components of SET and empirically measurable dimensions of an exchange relationship. Trust is also a core concept within privacy theories in that it is an antecedent to disclosure: a lack of trust undermines an individual's willingness to disclose. Privacy is a multidimensional concept, and the dimensions I measure here are *dimensions of protection* and *dimensions of provision* [84], specifically: the object of privacy (personal control), the target of privacy protection (personal information), and the mechanisms by which privacy is protected (as implemented by the company through policy and practices).

### 5.7.1 Trust

Trust between exchange partners facilitates the exchange itself, and its presence or absence may indicate the extent to which an exchange partner relies on assurances to engage in the relationship. Trust is measured here as both a general concept and as encapsulated trust (where the individual believes the company is acting in her own best interests).[51]

- *General trustworthiness*: Based on what I have read, I find the High-Tech Device Company to be trustworthy. (7 point Likert Agree/Disagree scale).

---

<sup>3</sup>This assurance was presented slightly differently than the other two in both studies. Because it was not a statement about the terms of the agreement, I included it directly above the "Terms of Use" in the survey. It also appeared above the Terms of Use in the split-screen layout.

- *Encapsulated trust*: Based on the information given in the scenario above, I trust the High-Tech Device Company with my data. (Data includes your body data, personal data, and location data.) (7 point Likert Agree/Disagree scale).

### 5.7.2 Power

Power is measured both as a general concept and as the individual's perceived control over the negotiation (negotiative power). The power one partner has over another in the relationship indicates the extent to which one partner controls access to alternatives for the other partner and their dependency on the exchange itself. Power is affected by one's position in an exchange network: even in situations where individuals do not trust their exchange partners and find the exchange unfair, they may find that they have no access to reasonable alternatives (other than the power to exit), or they are resigned to their status.[27]

- *General power*: In your opinion, who has more power in this relationship (exchanging my data for use of the WearMe device) – you, or the High Tech Device Company? (rating on a scale of 1-10, with 1= The company has all the power in the relationship; 5 = Neutral/equal power; 10 = I have all the power in the relationship; default slider position was 5).
- *Negotiative power*: In your opinion, who has more control over the terms of use in this relationship (exchanging my data for use of the WearMe device) – you, or the High Tech Device Company? (rating on a scale of 1-10, with 1= The company has complete control over the terms; 5 = Neutral/equal control; 10 = I have complete control over the terms; default slider position was 5).

### 5.7.3 Fairness

Fairness is measured by the respondent's assessment of the overall fairness of the exchange relationship, as well as the respondent's perception of who specifically gains more benefit from the exchange.

- *General fairness*: The relationship (exchanging my data for access to the WearMe device) I have with the High-Tech Device Company is fair to me. (7 point Likert Agree/Disagree scale).
- *Perception of benefit*: In your opinion, who benefits more from this relationship: you or the High-Tech Device Company? (rating on a scale of 1-10, with 1 = the relationship only benefits the company; 5 = we benefit equally; 10 = the relationship only benefits me; default slider position was 5).

### 5.7.4 Privacy

Privacy is measured here as *dimensions of protection* and *dimensions of provision*: the target of privacy protection (personal information), and the mechanisms by which it is provided (as personal control, and as practiced by the company).

- *Privacy as control*: Based on the information given in the scenario above, I am confident I would be able to control which data I disclose to the High-Tech Device Company (7 point Likert Agree/Disagree scale).
- *Privacy as a practice*: Based on the scenario and your own personal experiences, how would you rate the High Tech Device Company's privacy practices? (rating on scale of 1-10, with 1 = far below average, 5 = average, and 10 =far above average; default slider position was 5).
- *Privacy expectations*: Based on the information given in the scenario above, the way in which the HTDC collects and uses my data meets my privacy expectations. (7 point Likert Agree/Disagree scale).

## 5.8 Mediating Factors

I controlled for several potentially mediating factors in this study that attempt to capture the influence, if any, of individual characteristics and dispositions on participant responses. Details on the composition of scales can be found in Appendix A.

### 5.8.1 Demographic Factors

I collected the following standard demographic information from respondents:

- Current age (as ranges: 18-24, 25-34, etc.);
- Ethnic affiliation (as a single response, including a multi-racial response);
- Education (single response);
- Annual income (as a range: Between \$25K-\$49,999K, etc.);
- Gender (as a single response, including transgender, genderqueer, and a non-identification category).

I did not anticipate that the demographic factors would have significant effects on the dependent variables. Neither SET or privacy research suggest the salience of any specific demographics to their respective theories, and the privacy literature in particular has not yet identified any core demographics that are consistently predictive of privacy attitudes or expectations.

### 5.8.2 Trust and Caution Scale

Building on research by Cheshire *et al* examining perceptions of control over personal information online, I included Yamagishi's ten item trust and caution scale[120]. Cheshire *et al* included the trust and caution scale in their study as "indicators of a type of social intelligence—the propensity to trust others and be cautious in social interactions. . . generalized attitudes about trusting others in a variety of contexts are likely to associate with attitudes and behaviors in many offline or online interactions." [23] I also included it in order to assess what influence, if any, personal attitudes related to trust may have on the Trust dependent variable. Both scales run from low agreement (1) to high agreement (7); a higher score on the trust scale means the respondent's answers indicated a more trusting attitude, while a higher score on the caution scale means that the respondent's answers indicated a more cautious outlook.

### 5.8.3 Information Technology Knowledge Scale

The Information Technology Knowledge Scale (ITKS) I used was a revised version of the scale used by Cheshire *et al*, "designed to measure one's overall level of comfort and self-described knowledge about information technology" [23]. While not intended to assess whether a respondent legitimately understood the technology utilized in the vignette, it functions to provide a check on respondents' confidence in navigating complex technological systems. Consisting of four questions, answered on a seven point agree/disagree Likert scale, with a higher score indicating a higher self-rating of one's technical competency.

### 5.8.4 Privacy Scale

I developed a privacy scale for this study by selecting previously validated questions from surveys of information privacy that specifically related to disclosing personal information online or to a company or institution in an online context, as well as including two questions I wrote based on specific aspects of disclosure that I uncovered in the qualitative research phase.[73][112][70] I found I had to create my own scale because most existing scales either focused on areas not relevant to the study (*e.g.*, social media use, interpersonal privacy concerns) or included older questions focusing on specific forms of interactions that were no longer technically relevant. A legacy of Westin's privacy classifications (as discussed in Chapter 2) is the recognition that individuals do often hold privacy concerns that span a spectrum from greater to lesser relative to others. Thus, mapping privacy concerns across a multi-dimensional scale allows researchers to both gain a sense of how their participants rate relative to one another, and obtain an aggregate sense of the concerns of the respondent pool. Because I created my own scale I cannot assess the respondents relative to previously published studies on a scale basis (though this is possible on individual questions).

All questions were measured on a seven-point agree/disagree Likert scale, and after conducting principal component factor analysis to create the aggregate scale, the final scale runs from one (low concern with privacy) to seven (high concern).

## 5.9 Vignette-Specific Control Factors

In addition to the independent and dependent variables, I collected several vignette-specific factors in this study: respondents' judgments of sensitivity of the information types collected by the device; whether the respondent would disclose device data to their health insurance company; ratings of assurances (including the core assurances manipulated in the assurance scenario as well as additional assurances); perceptions of the form of the exchange relationship; the influence of anonymizing data on respondents' perceptions of risk; and device usage (whether the respondent currently uses a wearable device, and how likely the respondent would be to use the wearable device described here as well as any wearable device).

Analyzing these responses, I determined that in most instances, there were no significant differences in how the experimental groups answered these questions. Which, in retrospect, made sense—many of these questions were asking about preferences that were stable and not influenced by the experimental condition, rather than aspects that were directly related to the vignette. For example, questions asking respondents to rank the sensitivity of the data types collected by the device were not significant by experimental condition either in *t*-tests or in various regression analyses. In general, for most of these factors, it appeared that responses were not affected by condition or by the context of the vignette. Respondents ranked the sensitivity of location data as high ( $\mu$  of 4.44 out of 5, with 5 = extremely sensitive), and ranked their comfort with disclosing their location data to the company as low ( $\mu$  of 3.78, with 3 = somewhat disagree and 4 = neither agree nor disagree). *T*-tests of both of these measures against experimental condition were not significant; both groups were not comfortable disclosing location data, and this is likely a stable preference that is independent of the experimental condition or the vignette. Similarly, questions probing respondents' measurements of risk in specific scenarios where their data was sold, shared, or stolen (comparing identified data versus anonymized data) were also not significant by experimental condition; respondents likely already had stable preferences regarding the risk of data use by third parties.

The vignette-specific questions I ultimately included in the analysis were:

- Experimental condition;
- A measure of participants' comfort with disclosing their data to the company, based on the average of three questions specifying each data type collected in the vignette (personal data, body data, and location data);
- One question focusing on the ecological validity of the vignette: "If the WearMe Device was a real product, how likely would you be to use it?"; all responses measured on a 7-point likely/unlikely Likert scale.<sup>4</sup>

The disclosure comfort scale provides a vignette-specific composite privacy measure that averages a respondent's reported levels of comfort with disclosing each of the data types included

---

<sup>4</sup>The two other questions in this series, asking if the respondent would use any type of wearable device, and asking whether she had used/was using a wearable to track physical or emotional activity, were not significant when comparing groups, nor in initial regression analysis.

in the vignette. The Real Product measure controls for the effect of the respondents' willingness to use the WearMe device, working with the assumption that respondents who reported being less willing or unwilling to use the device may provide answers that are systematically different from those who were willing.

### 5.9.0.1 Scales

*Dependent Variables:* All of the dependent variable responses are coded from low to high agreement or rating.

- *Experimental Condition:* I created a dummy variable for the experimental condition, with mandatory disclosure as the baseline condition. Thus, the condition coefficient should be interpreted as affecting the disclosure optional group. In all cases, the coefficient and *t*-value are positive, as the dependent variable measures are scaled such that a higher agreement or rating of the dependent variable equals a higher value.
- *General Privacy scale:* A higher (positive) score indicates greater privacy concern.
- *ITKS scale:* A higher (positive) score indicates greater technical competency.
- *Trust scale:* A higher (positive) score indicates greater trust in others.
- *Caution scale:* A higher (positive) score indicates more caution towards others.
- *Comfort with Disclosure:* A higher (positive) score indicates greater comfort with disclosing the information specified in the vignette.
- *Real Product:* a dummy variable based on a Likert scale response to the question "If the WearMe Device was a real product, how likely would you be to use it?" Responses are scaled from 1 (extremely unlikely) to 7 (extremely likely); a higher score indicates greater likelihood of using the product. Responses were coded 1 if the response was 5 (slightly likely) or higher; 0 if the responses were 4 (neither likely nor unlikely) or lower.

## 5.10 Survey Instrument Development

The vignette and survey instruments underwent multiple revisions. Study One contained 96 questions: 49 vignette-specific questions; 42 general questions related to privacy, trust and caution, and technical ability; and five demographic questions. Studies 2A and 2B varied slightly from Study One after making a few revisions.

The vignette-specific questions included: questions that tested the effects of the independent variables on the key dependent variables; related questions that attempt to measure participants' views on the level of sensitivity of the information collected by the fictional device; and a set of questions designed as a manipulation check evaluating their views on the form of the relationship, their reliance on a range of assurances, their assessment of risk, and level of comfort with disclosing categories of data to the fictional company.

### 5.10.1 Recruitment and Study Procedures

All recruitment for the main studies took place on Prolific Academic (PA). PA is based in the United Kingdom, and at the time I ran each survey the pool exceeded 6,000 participants based in the United States, aged eighteen or older. I chose PA after an evaluation by Peer *et al*[90] demonstrated that PA participants showed greater naivete and higher quality task completion than Amazon Mechanical Turk participants. Additionally, PA cost substantially less than Mechanical Turk, and offered more transparency into the composition of their participant pool than did Amazon.

Similar to using Mechanical Turk, I placed a recruitment ad on PA's website describing the task, the estimated time for completion, and the amount paid. PA allows researchers to pre-screen potential participants; I limited my potential pool to those currently living in the United States, over the age of eighteen, and to those who had a previous study approval rate of at least 50 percent. I chose the 50 percent figure in order to try to eliminate lower quality participants without over-restricting the pool, as my primary concern was attracting a diverse pool of participants who were not experienced academic survey takers, a concern investigated with Peer *et al* about Mechanical Turk. After running the first study, I restricted participation in subsequent studies to participants who had not participated in any of my previous studies.

PA displays the study name, the researcher's name, a description of the study, the payment for the study and the hourly rate, the maximum allowed time and average completion time, and the available number of slots out of the total needed for study completion. I entitled my survey "University of California Berkeley—Wearable Device Survey," and allowed a maximum completion time of one hour. The following description of the study which appeared in the recruitment ad was approved by CPHS:

**Wearable Device Company and Product Survey—UC Berkeley**

In this study, we will ask you to read a description about a fictional company and one of its products, a new wearable device that helps people improve their health and fitness. You will then answer a series of questions about the company and the product, as well as some general questions about your internet experience and about your background. We expect this study to take approximately 20 minutes for you to complete. You will be paid U.S. \$3.00 upon completion of the survey.

Once a participant accepted the task, she clicked on a link in PA which took them directly to my survey on the Qualtrics platform (at the subdomain *berkeley.qualtrics.com*). The first screen that loaded upon arrival was a CPHS approved consent letter featuring the UC Berkeley official seal and the text of my consent document. The participant had to select a button stating "I consent - begin the study" to proceed. If the participant chose to withdraw at this stage, she could either close the browser or select the "I do not consent - I do not wish to participate" button, which displayed a thank you message, a note that the participant would not be paid, and a link back to the PA website.

If the participant consented, she was then shown an instruction page with the following text:

**Instructions:**

In this task, we ask you to imagine you have decided to purchase a wearable device

(similar to a wristwatch) to improve your health and fitness.

You will be presented with a description of the device and the company that produces it on the next page, as well as the terms you must agree to in order to use the device.

Then you will be presented with a series of questions to answer about your impressions of the device.

This task should take approximately 20 minutes to complete.

After clicking ‘Continue,’ the participant was then shown the vignette text. I randomly assigned the participant to an experimental condition at this stage by using custom Javascript to set an embedded variable named ‘condition’ that corresponded to the manipulation statement shown to the participant. For example, Study One had two conditions: mandatory disclosure and optional disclosure. This variable was stored in the survey database in a corresponding row for each respondent. The ‘Continue’ button was delayed by ten seconds on this page to ensure that participants took at least that much time to read the vignette text. All pages from the vignette page forward contained hidden timing questions in order to track how long a participant spent on each page, as well as to calculate the total duration of the survey. Outlier responses that were significantly shorter than the average were reviewed for quality purposes. All questions, with the exception of two open text responses questions, required a response to proceed.

The randomized manipulation statement was appended to the bottom of the vignette using custom Javascript, and the statement text was stored in an embedded variable that allowed me to access it throughout the duration of the survey.

One of my concerns with the visual aspect of the experimental design was ensuring that both the vignette text and manipulation statement were readily available to participants throughout the survey. There was no functionality within Qualtrics that could address this, so I engaged a software developer (my husband) to create custom Javascript that would present all vignette-specific questions on a split screen. The top of the split contained the vignette text and the manipulation statement; the bottom contained a maximum of three questions per screen. As noted earlier, in the initial version of Study One I placed the vignette text above the manipulation statement in the split screen, which meant that a participant had to scroll downward if she wanted to access the statement on question pages.

I was uncertain whether participants would recall the manipulation as they proceeded through the questions, and the preliminary results from Study One supported my concern as the experimental manipulation produced small effects that supported my hypothesis but that were not statistically significant. Thus, I chose to revise the layout by moving the manipulation statements above the vignette text in the split screen. This revision meant that the manipulation statements were always visible on every split screen page without scrolling. Additionally, I made minor design revisions to the vignette text as well to make it easier to read: placing each sentence on a new line, and putting the list of personal information collected into a bulleted list rather than as part of one long sentence. I then reran Study One with a larger respondent pool (250 instead of 100) and found that the small effects I had observed in the first one were now strongly significant.

After viewing the vignette and the manipulation statement, respondents proceeded through eleven screens of vignette specific questions. When participants reached the general questions



**WearMe wearable tracker -- The High-Tech Device Company (HTDC)**

The High-Tech Device Company (HTDC) is a world leader in consumer technology and wearable devices, building products powered by research and passion.

Our wearable tracker, the "WearMe", helps people live better by providing personalized insights into how they sleep, move, and feel.

Similar in size and shape to a wristwatch, the WearMe features five colors in three sporty and fashionable styles.

Our approach to lifestyle tracking is unique, relying on multiple data points to customize the WearMe's recommendations.

The WearMe tracks your:

- body data
  - heart rate
  - number of steps you walk or run
  - body temperature
  - sleep cycles)
- location data
  - precise physical location based on where you go throughout the day.

The WearMe then provides you with personalized recommendations for improving your health and fitness based on the tracking data collected from your WearMe Device and additional information about you that provide in your WearMe profile.

Our patented approach provides you with a health and fitness program tailored to your goals and needs.

You access your personalized recommendations using the WearMe mobile app or at the WearMe website.

This requires activating an account with your personal data:

- name
- age
- gender
- email address
- home address
- mobile phone number.

Accounts are subject to our Terms of Use.

We collect and store your data in accordance with standard industry practices.

**WearMe Terms of Use**

**WearMe Price: \$99.99**

You can choose the types of data about your body and your physical location the WearMe device will track.

<< Back

Continue >>

Figure 5.1: Vignette Text

---

**WearMe Terms of Use**

**WearMe Price: \$99.99**  
 You must allow the WearMe to track all the data it requests about your body and your physical location.

**WearMe wearable tracker -- The High-Tech Device Company (HTDC)**

The High-Tech Device Company (HTDC) is a world leader in consumer technology and wearable devices, building products powered by research and passion.

Our wearable tracker, the "WearMe", helps people live better by providing personalized insights into how they sleep, move, and feel.

Similar in size and shape to a wristwatch, the WearMe features five colors in three sporty and fashionable styles.

---

**As you answer the following questions imagine you have purchased and are using the WearMe device.**

**Feel free to refer back to the WearMe Terms of Use and the Company Overview (in the split screen above) as you need when answering.**

*Please indicate your level of agreement with this statement:*

I am comfortable with disclosing my personal data (name, gender, age, email address, home address, mobile phone number) to the High-Tech Device Company in order to use the WearMe device.

Strongly agree

Agree

Figure 5.2: Example of the split screen in the experimental survey

section, I removed the split screen. There were five screens for the general and demographic questions. Upon reaching the end of the survey, respondents were required to enter their PA participant ID into a text field, a standard practice for PA surveys. I recorded the ID as required by PA so that I could use it to verify survey completion and authorize payment. Upon entering the ID and hitting submit, respondents were redirected to a PA completion page so that they could verify their completion of the survey.

I reviewed respondent completions within three days of the completion of each study. In general, once published the studies took between two hours to two days to complete, depending upon the day of the week and time they were published. After verifying survey completion, and reviewing outlier response times to ensure that there were no obvious cheats (*e.g.*, answering all questions with the first response, or response times under five minutes), I approved participant payment, and closed the survey to additional responses on both the PA and Qualtrics platforms.

### 5.10.2 Study Two Procedures

Studies 2A and 2B proceeded exactly as Study One. Study 2A was run in October 2017 with 515 participants; Study 2B was run in November 2017 with 527 participants.

## 5.11 Analytic Strategy

After analyzing the data for direct effects (*t*-tests for Study One; ANOVAs and *t*-tests for Study Two), I used regression analysis in order to control for other factors that might contribute to the effects on the dependent variables. Hypothesis tests and other results are discussed in Chapter Five.

### 5.11.1 Study One

I ran Study One in September 2017. Two hundred and fifty respondents completed the experiment, with 122 (49%) randomly assigned to the optional disclosure condition, and 128 (51%) to the mandatory disclosure condition. The respondent pool skewed younger (84% under the age of 44), identified primarily as White/Caucasian (71%), was slightly more male (56% to 42% female), was predominantly college educated (32% had completed some college or an AA degree; 52% had completed college or graduate school), and poor to middle income (37% reported annual income under \$25,000/year; 39% reported annual income of between \$25K-\$75K).

In Study One, I discovered that the distribution of income in this sample was skewed toward the lower end of the scale, with 61 percent of the respondents reporting an annual income below \$50K. I applied a logarithmic transformation to the income variable before including it in the regressions to normalize the distribution.

#### 5.11.1.1 Stage 1: Bivariate Hypothesis Testing

I examined the main direct effect of the experimental condition by conducting a series of *t*-tests comparing the means between the two experimental groups against each of the dependent variables. Analysis controlling for other key factors occurred in Stage 2. The fairness, trust, and power dependent variables had at two conceptual measures: one general conceptual measure, and one specific operationalized measure, while the privacy dependent variable had three operationalized measures.

I then conducted a linear regression (OLS) against each dependent variable and the experimental condition. The results of these two statistics informed whether: a) I could accept or reject the hypotheses; b) whether, if conceptually appropriate, I could create combined single measures for the dependent variables. Combined measures were created by standardizing the scores for each measure and then computing the average. I standardized the measures because the scales used for the dependent variable measures were not uniform: some used the 1-7 agree/disagree scale, while others used a 1-10 scale rating perceptions of power, benefit, and control over the terms of use (1 favored the company; 10 favored the respondent).

I ran *t*-tests and calculated effect sizes (Cohen's *d*) for each individual dependent variable measure as well as the combined measures, and report the results in 6. Correlations for the general and specific measures, along with Cronbach's  $\alpha$  values, are also reported for each dependent variable.

### 5.11.1.2 Regression Analysis

I incorporated factors from three sources: traditional demographic factors; general factors measuring privacy attitudes and preferences, trust, caution, and personal facility with technology; and vignette-specific factors, such as the respondents' interest in using the WearMe device, and respondents' ratings of the sensitivity of specific types of information collected by the device. I created four regression models for each key outcome:

- Model One: Controls for factors related to trust and privacy: general privacy concern, trust, caution, and technical competency;
- Model Two: Controls for demographic factors;
- Model Three: Controls for vignette-specific factors;
- Model Four: Controls for significant factors from the previous three models.

I did not create a 'kitchen sink' model incorporating all of the factors because I did not have a theoretical basis for doing so.

## 5.11.2 Study Two

For Study 2A, five hundred and twenty-seven respondents completed the experiment in October 2017, with 130 (25%) randomly assigned to the mandatory disclosure condition, 397 distributed equally between the three assurance conditions (25% each). The respondent pool skewed younger (75% under the age of 44), identified primarily as White/Caucasian (75%), was nearly evenly divided on gender (50% female, 48% male), was predominantly college educated (28% had completed some college or an AA degree; 61% had completed college or graduate school), and poor to middle income (27% reported annual income under \$25,000/year; 43% reported annual income of between \$25K-\$75K).

For Study 2B, five hundred and fifteen respondents completed the experiment in November 2017, with 128 (25%) randomly assigned to the mandatory disclosure condition, 387 distributed equally between the three assurance conditions (25% each). The respondent pool skewed younger (85% under the age of 44), identified primarily as White/Caucasian (66%), was closely divided on gender (46% female, 51% male), was predominantly college educated (36% had completed some college or an AA degree; 50% had completed college or graduate school), and poor to middle income (29% reported annual income under \$25,000/year; 42% reported annual income of between \$25K-\$75K).

In both studies, I first tested for main effects using ANOVA. However, in both cases I was not able to support any of the hypotheses. I then tested the means of each condition against the control, and again found no significant results. These findings and an analysis of the factors that may have contributed to this study's lack of significant results are discussed in Chapter 6.

# Chapter 6

## Experimental Survey Findings

In this chapter I review the results of the experimental survey studies.

### 6.1 Study One Findings

#### 6.1.1 Hypothesis 1: Trust

**H1a: supported**

**H1b: supported**

A comparison of means between experimental groups supported both H1a and H1b. The mean value for general trust was statistically significantly higher in the optional disclosure condition than the mandatory disclosure condition ( $t = -1.71, p \leq .05$ , Cohen's  $d = .22$ ), as well as for encapsulated trust ( $t = -2.34, p \leq .05$ , Cohen's  $d = .30$ ). The optional disclosure group demonstrated both higher levels of general trust toward the company directly and trust towards the company's access to their personal data than the mandatory disclosure group. The means for both groups for the general trust measure were slightly above neutral (between 4 (Neither agree nor disagree) and 5 (Somewhat agree)), while the means for the encapsulated trust measure were lower, between 3 (Somewhat disagree) and below 5 (Somewhat agree). Because the two measures were highly correlated and had a high alpha value, I created a combined measure for these variables. As both measures were conducted on the same scale (1-7), I computed the combined measure by taking the average of both scores for each respondent. The mean value for the combined trust measure was also statistically significantly higher in the optional disclosure condition than the mandatory disclosure condition ( $t = -2.19, p \leq .05$ , Cohen's  $d = .27$ ).

#### 6.1.2 Hypothesis 2: Power

**H2a: supported**

**H2b: supported**

A comparison of means between experimental groups supported both H2a and H2b. The mean value for general power was statistically significantly higher in the optional disclosure condition

than the mandatory disclosure condition ( $t = -2.50, p \leq .01$ , Cohen's  $d = .32$ ), as well as for negotiative power ( $t = -2.14, p \leq .05$ , Cohen's  $d = .27$ ). The optional disclosure group reported greater individual power in the relationship and over the terms of the relationship than the mandatory disclosure group. The means for the general power measure for both groups were between 3 and 5 on a scale of 1-10, with each extreme representing total power over the relationship (1 = the company, 10 = the individual). Thus, both groups felt the balance of power in the relationship favored the company over the individual. The means for the negotiative power measure for both groups were between 3 and 4 on a scale of 1-10, and used the same scale as the general power measure. Similarly, both groups expressed that the company had more control over the terms than the individual. A regression analysis of each measure controlling for the experimental condition also supported the results of the  $t$ -test, with a  $p$ -value for general power of .013, and .034 for the negotiative power measure. The two measures were moderately correlated with one another (.65), and also had a high  $\alpha$  value (.78). Based on the strength of these two measures I combined them to create a single measure for power for subsequent regression models. As both measures were conducted on the same scale (1-10), I computed the combined measure by taking the average of both scores for each respondent. The mean value for the combined power measure was also statistically significantly higher in the optional disclosure condition than the mandatory disclosure condition ( $t = -2.56, p \leq .01$ , Cohen's  $d = .32$ ).

### 6.1.3 Hypothesis 3: Fairness

**H3a: supported**

**H3b: weakly supported**

A comparison of means between experimental groups supported H3a and weakly supported H3b. The mean value for general fairness was statistically significantly higher in the optional disclosure condition than the mandatory disclosure condition ( $t = -4.04, p \leq .001$ , Cohen's  $d = .51$ ), as well as weakly so for perceptions of benefit in the relationship ( $t = -1.55, p \leq .10$ , Cohen's  $d = .20$ ). The optional disclosure group both rated the relationship as fairer to them and as having more benefit to them than the mandatory disclosure group. The means for the two groups on the general fairness measure were between 3 (*Somewhat disagree*) and below 5 (*Somewhat agree*), and above 4 and below 5 on a scale from 1-10 for the perception of benefit measure. Both groups rated the company as gaining more benefit from the relationship than them as individuals. Because the perception of benefit measure was only weakly supported, I did not create a combined measure for it for later regression analysis.

### 6.1.4 Hypothesis 4: Privacy

**H4a: supported**

**H4b: supported**

**H4c: supported**

A comparison of means between experimental groups supported H4a, H4b, and H4c. The mean value for the disclosure measure was statistically significantly higher in the optional disclosure

condition than the mandatory disclosure condition ( $t = -7.35, p \leq .001$ , Cohen's  $d = .93$ ), for privacy practices ( $t = -3.68, p \leq .001$ , Cohen's  $d = .47$ ), and for privacy expectations ( $t = -3.01, p \leq .001$ , Cohen's  $d = .38$ ). Again, privacy was measured as one's perceived control over disclosure, as well as by having participants provide a rating of the company's privacy practices and whether the practices met their expectations. The optional disclosure group reported more control over disclosure, rated the company's privacy practices significantly higher, and rated the company data collection and usage as meeting their privacy expectations at a higher level. The difference between means was especially notable on the question of disclosure (controlling which data the respondent would disclose to the company)—the mean for the optional disclosure group was 1.72 greater than for the mandatory disclosure group, almost a full standard deviation for both means.

I created a combined measure for privacy based on two of the three measures: privacy expectations and privacy practices. Because these two measures used different scoring, the combined measure was calculated after standardizing scores for each variable and then taking the average of the two. I did not include disclosure in this combined measure as it is a theoretically distinct concept from the measures of privacy practices and expectations. The privacy expectations and practices measures had a moderate correlation of .59 and an  $\alpha$  value of .74.<sup>1</sup> The mean value for the combined privacy measure was also statistically significantly higher in the optional disclosure condition than the mandatory disclosure condition ( $t = -3.77, p \leq .001$ , Cohen's  $d = .48$ ).

## 6.2 Stage Two: Regression Analysis

After verifying the hypotheses, I performed a series of regression analyses in order to further explore the main effects while controlling for key factors that might otherwise explain the outcomes of interest. Because the various factors are measured on different scales, standardized coefficients are reported for all regressions.

### 6.2.1 Trust Measure

The combined Trust measure remained significant with the experimental condition across Models One and Two, but not Three and Four.

#### 6.2.1.1 Model One

In Model One, the Trust measure demonstrated a positive relationship with the optional disclosure experimental condition ( $coef = .13, p \leq .01$ ), technical knowledge ( $coef = .16, p \leq .01$ ), and a negative relationship to general privacy concern ( $coef = -.58, p \leq .001$ ). The trust scale was significant only in an individual regression with the experimental condition, and the caution scale was not significant in either an individual regression or in the complete model. In sum, greater trust towards how the company handles personal data was associated with greater technical competency,

---

<sup>1</sup>Note that adding the disclosure measure into a calculation of  $\alpha$  results in a modest increase, from .74 to .79.

Table 6.1: Trust Measure Combined

	Model 1		Model 2		Model 3		Model 4	
Condition	0.13**	(0.09)	0.13*	(0.12)	0.03	(0.07)	0.05	(0.07)
ITKS Scale	0.16**	(0.05)					0.01	(0.04)
Trust Scale	0.09	(0.05)					0.01	(0.04)
Caution Scale	0.10	(0.06)					0.04	(0.04)
Gen. Privacy Scale	-0.58***	(0.06)					-0.21***	(0.06)
Age			0.04	(0.06)			0.06	(0.03)
Eth. (0=White, 1=Non-White)			-0.02	(0.14)			-0.00	(0.08)
Gender (0=Male, 1=Female)			-0.11	(0.13)			-0.04	(0.07)
Income (log)			0.25***	(0.11)			0.02	(0.06)
Education			0.04	(0.07)			-0.02	(0.04)
Real Prod. (0 = Unlikely/Neut)					0.24***	(0.08)	0.20***	(0.08)
Comfort w/Disclosure					0.67***	(0.03)	0.57***	(0.03)
Observations	250		233		250		233	
Adjusted $R^2$	0.39		0.08		0.70		0.71	

Standardized beta coefficients; Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



and with lower overall privacy concern in the optional disclosure condition. The adjusted  $R^2$  value for this model was .39, demonstrating a strong fit.

#### 6.2.1.2 Model Two

Model Two demonstrated a poor fit (adjusted  $R^2=.08$ ); only the experimental condition ( $coef = .13$ ,  $p \leq .05$ ) and income were significant ( $coef = .25$ ,  $p \leq .001$ ). Greater trust towards the company was associated with higher income in the optional disclosure condition.

#### 6.2.1.3 Model Three

In Model Three the real product measure ( $coef = .24$ ,  $p \leq .001$ ) and comfort with disclosure ( $coef = .67$ ,  $p \leq .001$ ) factors were significant (adjusted  $R^2=.70$ ), but the experimental condition was not. Greater trust toward the company was associated with willingness to use the WearMe device and greater comfort with disclosing data to the company.

#### 6.2.1.4 Model Four

While Model Four had the best model fit (adjusted  $R^2=.71$ ), the experimental condition was not significant in this model. The significant covariates in Model Four were general privacy ( $coef = -.21$ ,  $p \leq .001$ ), the real product measure ( $coef = .20$ ,  $p \leq .001$ ), and comfort with disclosure ( $coef = .57$ ,  $p \leq .001$ ).

#### 6.2.1.5 Discussion

Looking across the four models, the coefficients demonstrate that the respondents' general orientation towards privacy, their vignette-specific comfort with disclosure and likelihood of using the wearable device provided the greatest contribution to respondents' assessment of trust. As their trust in the company increased, respondents' level of general privacy concern decreased, and their comfort with disclosing personal information to the company and likelihood of using the device increased. Technical competency also provided explanatory power in Models One and Four, though far less than the general privacy and disclosure comfort measures. Because one aspect of the trust measure focused on how the respondents believed the company would treat their personal data, the fact that a higher self-rating of technical competency was associated with greater trust suggests that these respondents may believe they have a greater understanding of how companies actually process personal data.

Despite the fact that their inclusion negates the effect of the experimental condition in Models Three and Four, the strong  $R^2$  value associated with the disclosure comfort and the real product measures suggest that overall comfort with the substance of the vignette and a willingness to use a wearable device contributed strongly to all respondents' perceptions of trust. Absent these factors, one's general privacy orientation remains the most substantive contributor, demonstrating that an openness to the type of data collection and use illustrated in this scenario is key; respondents who

were less comfortable with the disclosure scenario or less willing to use this type of device were not as trusting.

## 6.2.2 Power Measure

The combined power measure, created by combining the general power measure and the negotiative power measure, remained significant with experimental condition across Models One, Two, and Four, but not Three.

Table 6.2: Power Measure Combined

	Model 1		Model 2		Model 3		Model 4	
Condition	0.16**	(0.09)	0.16**	(0.11)	0.10	(0.10)	0.14**	(0.09)
ITKS Scale	-0.03	(0.05)					-0.12*	(0.05)
Trust Scale	0.22***	(0.05)					0.16*	(0.05)
Caution Scale	0.21***	(0.06)					0.16*	(0.06)
Gen. Privacy Scale	-0.51***	(0.06)					-0.35***	(0.08)
Age			-0.07	(0.05)			-0.07	(0.04)
Eth. (0=White, 1=Non-White)			-0.05	(0.13)			-0.05	(0.11)
Gender (0=Male, 1=Female)			-0.11	(0.12)			-0.08	(0.10)
Income (log)			0.31***	(0.10)			0.15*	(0.09)
Education			0.07	(0.06)			0.05	(0.05)
Real Prod. (0 = Unlikely/Neut)					0.21**	(0.12)	0.10	(0.12)
Comfort w/Disclosure					0.31***	(0.04)	0.13	(0.04)
Observations	250		233		250		233	
Adjusted $R^2$	0.35		0.15		0.23		0.40	

Standardized beta coefficients; Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 6.2.2.1 Model One

In Model One, the combined power measure demonstrated positive relationships with the experimental condition ( $coef = .16, p \leq .01$ ), trust ( $coef = .22, p \leq .001$ ) and caution scales ( $coef = .21, p \leq .001$ ), and a negative relationship with general privacy ( $coef = -.51, p \leq .001$ ). The technical competency scale was not significant. In sum, higher ratings of individual power and control by the respondents were associated with the optional disclosure condition, greater trust towards others, higher caution towards others, and with lower overall privacy concern. The adjusted  $R^2$  value for this model was .35, demonstrating a moderate fit.

### 6.2.2.2 Model Two

In Model Two, we see the continuation of the trend of none of the demographic covariates other than income having significance in the model. Higher ratings of power were associated with the experimental condition ( $coef = .16, p \leq .01$ ) and greater levels of income ( $coef = .31, p \leq .001$ ); the adjusted  $R^2$  value was .15, indicating a poor fit.

### 6.2.2.3 Model Three

In Model Three, the experimental condition was not significant, though the comfort with disclosure ( $coef = .31, p \leq .001$ ) and real product measures ( $coef = .21, p \leq .01$ ) were significant. The adjusted  $R^2$  value was .23, indicating a poor fit.

### 6.2.2.4 Model Four

The significant factors in Model Four were the experimental condition ( $coef = .14, p \leq .01$ ), technical competence ( $coef = -.12, p \leq .05$ ), trust score ( $coef = .16, p \leq .05$ ), caution score ( $coef = .16, p \leq .05$ ), and general privacy ( $coef = -.51, p \leq .001$ ), and income ( $coef = .15, p \leq .05$ ). (adjusted  $R^2 = .40$ ). Interestingly, technical competence was not significant in Model One, suggesting an interaction effect with one of the other factors in this model. It also has a negative relationship, indicating that higher perceptions of power are associated with lower self-ratings of technical competency. Because the real product and comfort with disclosure measures were not significant with the experimental condition in Model Three, I anticipated that they would void its significance in Model Four, but I was incorrect. But given that the adjusted  $R^2$  value for Model Three was only .23, these two covariates clearly contribute little to the overall variance.

### 6.2.2.5 Discussion

The experimental condition remained significant across all the models except the third—the vignette specific factors in Model Three appeared to have no bearing on respondents' ratings of power. Overall, higher ratings on the combined power measure were associated with the optional disclosure experimental condition, lower technical competency, greater trust, caution, and with lower general privacy concerns, as well as higher income—the only dependent variable for which

income remained significant. These associations were robust across Models One, Two, and Four. They are sensible on their face: while having greater negotiative power should increase one's sense of individual power, so too should higher income insofar as wealth enables individual power. The association between increased perceptions of power and lower technical competency is an interesting one, suggesting that those who feel more powerful may not understand how the technologies used in this scenario work.

These findings introduce an interesting relationship between the Trust and Caution scales: increased power was associated with both increased trust in others and increased caution towards others. At first glance, these findings seem incompatible—how could someone be both more trusting and more cautious towards others? However, this finding mirrors those of Cheshire *et al*, who found similarly in their study of online discretion and information control: “Our findings are consistent with related research that shows that trust and caution are independent dimensions of a similar concept. The distinctive pattern of higher trust and higher caution indicates a propensity to engage in risky and uncertain social interactions while sustaining forethought and discretion.”[23] Similarly here, greater individual power was associated both with greater trust towards others and greater caution. Because this study focuses on personal disclosure, the conclusion I draw from this finding is support for the argument that online disclosure isn't a behavior that can be stereotyped as inherently high risk. Given how widely people are required on a daily basis to affirmatively disclose some form of personal information in order to engage in routine transactions (*e.g.*, banking, reading the news, checking a child's school progress), being willing to engage in these forms of disclosure (as demonstrated by lower general privacy concern) requires a higher level of trust towards others' intentions (including corporations). At the same time, this openness does not imply either gullibility or a lack of discretion. That these two measures were associated with increased individual power also suggests the opposite—that those who felt less empowered were both less trusting and less cautious. While greater individual power was significantly associated with the optional disclosure condition, this finding implies that to the extent that individuals' ratings of trust and caution are related to the experimental manipulation, those who rated themselves as less powerful may also be less trusting of others, and less cautious towards them. This combination may suggest both lower trust as well as lower caution related to a lack of personal agency: feeling less powerful may not only be related to being less trusting towards others, the lack of agency might contribute to being less cautious—perhaps because if one lacks agency, one might believe that she can have little impact on the actions of others.<sup>2</sup>

### 6.2.3 Fairness Measure

The general fairness measure (how fair is the relationship to me) was significant with the experimental condition across all four models.

---

<sup>2</sup>I included in the survey Turow *et al*'s two survey questions measuring resignation to online advertising. Following their criteria, 77% of the sample in Study One qualified as resigned, providing another possible explanation for a lack of agency.[112]

Table 6.3: Fairness Measure

	Model 1		Model 2		Model 3		Model 4	
Condition	0.24***	(0.17)	0.25***	(0.22)	0.16***	(0.14)	0.17***	(0.15)
ITKS Scale	0.17**	(0.09)					0.05	(0.08)
Trust Scale	-0.01	(0.09)					-0.04	(0.08)
Caution Scale	0.14*	(0.11)					0.11*	(0.09)
Gen. Privacy Scale	-0.56***	(0.12)					-0.23***	(0.12)
Age			-0.05	(0.10)			-0.01	(0.07)
Eth. (0=White, 1=Non-White)			0.03	(0.25)			0.03	(0.18)
Gender (0=Male, 1=Female)			-0.05	(0.23)			0.02	(0.16)
Income (log)			0.20**	(0.19)			-0.01	(0.14)
Education			0.06	(0.12)			0.01	(0.08)
Real Prod. (0 = Unlikely/Neut)					0.21***	(0.18)	0.17**	(0.18)
Comfort w/Disclosure					0.58***	(0.06)	0.48***	(0.07)
Observations	250		233		250		233	
Adjusted $R^2$	0.38		0.09		0.57		0.59	

Standardized beta coefficients; Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 6.2.3.1 Model One

In Model One the experimental condition was significant ( $coef = .24, p \leq .001$ ), along with the technical competency scale ( $coef = .17, p \leq .01$ ), the caution scale ( $coef = .14, p \leq .05$ ), and general privacy ( $coef = -.56, p \leq .001$ ). Ultimately, higher levels of perceiving the relationship as fair were associated with the optional disclosure condition, higher technical competency, greater caution, and less concern with privacy. The adjusted  $R^2$  value for the full model was .38, a moderate fit.

### 6.2.3.2 Model Two

In Model Two, once again the experimental condition ( $coef = .25, p \leq .001$ ) and income ( $coef = .20, p \leq .01$ ) were the only factors significant in the model. The adjusted  $R^2$  value for this model was weak, .09.

### 6.2.3.3 Model Three

In Model Three fairness was significant with the experimental condition ( $coef = .16, p \leq .001$ ), the real product measure ( $coef = .21, p \leq .001$ ) and comfort with disclosure ( $coef = .58, p \leq .001$ ). Higher levels of fairness were associated with the optional disclosure condition, with respondents' reports that they would use the WearMe if it were a real product, and with increased disclosure comfort of personal information to the company. The adjusted  $R^2$  value was .57, demonstrating a strong fit.

### 6.2.3.4 Model Four

In Model Four the experimental condition was significant ( $coef = .17, p \leq .001$ ) with the caution scale ( $coef = .11, p \leq .05$ ), general privacy scale ( $coef = -.23, p \leq .001$ ), real product measure ( $coef = .17, p \leq .01$ ), and comfort with disclosure ( $coef = .48, p \leq .001$ ), with an adjusted  $R^2$  value of .59, indicating a strong fit. Respondents who reported the relationship as fair to them were associated with the optional disclosure condition, increased caution, lower concern with privacy, willingness to use the WearMe device, and increased comfort with disclosing to the company. Interestingly, the the ITKS scale was not significant in this model.

### 6.2.3.5 Discussion

All four models were significant with the experimental condition. Looking across them, it appears that the factors contributing the most influence to respondents' perceptions of fairness were the experimental condition, their general privacy orientation, and their comfort with disclosing to the company. Increased caution as well as willingness to use the device also played a role. Respondents who were comfortable with the premise of the vignette and the personal disclosure it required also had fewer general privacy concerns, but those who also had the power to negotiate their level of disclosure were more likely to perceive this relationship as fair.

## 6.2.4 Privacy Measures

Again, I measured privacy as one's perceived control over disclosure, as well as by using a combined measure comprised of participants' ratings of the company's privacy practices and whether the practices met their expectations.

### 6.2.4.1 Model One

For the disclosure measure, the experimental condition was significant ( $coef = .42, p \leq .001$ ), as was the caution scale ( $coef = .16, p \leq .01$ ) and the general privacy measure ( $coef = -.41, p \leq .001$ ). Higher ratings of control over personal disclosure were associated with the disclosure optional condition, higher levels of caution, and lower levels of general privacy concern; the adjusted  $R^2$  value for this model was .35, a moderate fit. The experimental condition coefficient was sizeable in this model (.42), suggesting that the assignment into the optional versus mandatory condition exerted sizable influence on one's propensity to disclose.

Higher ratings on the combined privacy measure were also associated with the disclosure optional condition ( $coef = .23, p \leq .001$ ), higher levels of both trust ( $coef = .15, p \leq .05$ ) and caution ( $coef = .16, p \leq .01$ ), as well as lower levels of general privacy concern ( $coef = -.55, p \leq .001$ ). The adjusted  $R^2$  value for this model was .40, also a moderate fit.

### 6.2.4.2 Model Two

In Model Two, once again the experimental condition and income were the only significant factors. Higher ratings of control over personal disclosure was associated with the optional disclosure condition ( $coef = .42, p \leq .001$ ) and with higher levels of income ( $coef = .19, p \leq .01$ ); the adjusted  $R^2$  value for this model was .20, a relatively poor fit. Higher ratings on the combined privacy measure were also associated with the disclosure optional condition ( $coef = .21, p \leq .001$ ) and higher levels of income ( $coef = .30, p \leq .001$ ); the adjusted  $R^2$  value for this model was .12, indicating a very poor fit.

### 6.2.4.3 Model Three

Both privacy measures were significant with the vignette-specific covariates independently and in the full model. Higher ratings of control over personal disclosure was associated with the disclosure optional condition ( $coef = .36, p \leq .001$ ), willingness to use the WearMe device ( $coef = .24, p \leq .001$ ), and increased comfort with disclosure to the company ( $coef = .31, p \leq .001$ ); the adjusted  $R^2$  value for this model was .41, indicating a good fit. Higher ratings on the combined privacy measure were also associated with the disclosure optional condition ( $coef = .14, p \leq .001$ ), willingness to use the WearMe device ( $coef = .35, p \leq .001$ ), and increased comfort with disclosure to the company ( $coef = .39, p \leq .001$ ); the adjusted  $R^2$  value for this model was .47, also indicating a good fit.

Table 6.4: Privacy Measure – Disclosure

	Model 1		Model 2		Model 3	
Condition	0.42***	(0.24)	0.36***	(0.20)	0.37***	(0.21)
Age	0.04	(0.11)			0.07	(0.10)
Eth. (0=White, 1=Non-White)	0.02	(0.28)			0.01	(0.24)
Gender (0=Male, 1=Female)	-0.11	(0.25)			-0.06	(0.22)
Income (log)	0.19**	(0.21)			0.03	(0.19)
Education	-0.03	(0.13)			-0.05	(0.12)
Real Prod. (0 = Unlikely/Neut)			0.24***	(0.25)	0.20**	(0.25)
Comfort w/Disclosure			0.31***	(0.09)	0.23**	(0.10)
ITKS Scale					-0.01	(0.12)
Trust Scale					-0.00	(0.11)
Caution Scale					0.14*	(0.13)
Gen. Privacy Scale					-0.20**	(0.17)
Observations	233		250		233	
Adjusted $R^2$	0.20		0.41		0.42	

Standardized beta coefficients; Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



Table 6.5: Privacy Combined Measure

	Model 1		Model 2		Model 3		Model 4	
Condition	0.23***	(0.09)	0.21***	(0.11)	0.14**	(0.08)	0.15**	(0.08)
ITKS Scale	0.08	(0.05)					-0.00	(0.05)
Trust Scale	0.15*	(0.04)					0.08	(0.04)
Caution Scale	0.16**	(0.05)					0.09	(0.05)
Gen. Privacy Scale	-0.55***	(0.06)					-0.28***	(0.07)
Age			-0.01	(0.05)			0.01	(0.04)
Eth. (0=White, 1=Non-White)			0.02	(0.13)			0.02	(0.10)
Gender (0=Male, 1=Female)			-0.09	(0.11)			-0.03	(0.09)
Income (log)			0.30***	(0.10)			0.10	(0.08)
Education			-0.02	(0.06)			-0.07	(0.05)
Real Prod. (0 = Unlikely/Neut)					0.35***	(0.10)	0.27***	(0.10)
Comfort w/Disclosure					0.39***	(0.04)	0.26***	(0.04)
Observations	250		233		250		233	
Adjusted $R^2$	0.40		0.12		0.47		0.52	

Standardized beta coefficients; Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

#### 6.2.4.4 Model Four

The Disclosure measure remained significant with the experimental condition ( $coef = .37, p \leq .001$ ) and relatively high in Model Four, along with: the caution scale ( $coef = .14, p \leq .05$ ), general privacy scale ( $coef = .28, p \leq .001$ ), the real product measure ( $coef = .20, p \leq .01$ ), and comfort with disclosure ( $coef = .23, p \leq .01$ ). The adjusted  $R^2$  value for this model was .42, indicating a good fit. The variance in the Disclosure dependent variable appears to be best explained by this model, which had the highest model fit. Interestingly, the experimental condition had the largest coefficient in this model (.37), with the values of the other covariates all roughly in a lower range (.14 to .28). Respondents who felt they could control their disclosure to the company were significantly associated with the disclosure optional condition, were more cautious, less concerned with privacy, more comfortable with disclosing personal information to the company, and more likely to use the WearMe device. These results also support the key finding from Brandimarte *et al*: increased control over disclosure was significantly associated with increased comfort with disclosure.

The Privacy combined measure also remained significant with the experimental condition ( $coef = .15, p \leq .01$ ), along with the general privacy scale ( $coef = -.28, p \leq .001$ ), the real product measure ( $coef = .27, p \leq .001$ ), and comfort with disclosure ( $coef = .26, p \leq .001$ ). The adjusted  $R^2$  value for this model was .52, indicating a strong fit. Model Four also had the strongest r-squared value for the Privacy combined measure despite the disappearance of the significant effects of the Trust and Caution scales from Model One. Again, this measure was a composite of respondents' ratings of the company's privacy practices and how well the practices met their expectations. Higher ratings on this measure were associated with the disclosure optional condition, greater comfort with disclosing to the company, greater willingness to use the WearMe device, and lower general privacy concerns.

#### 6.2.4.5 Discussion

While the experimental condition was significant across all of the dependent variables, the values of standardized coefficients have been modest (from .25 to lower). The experimental condition's effect on the Disclosure measure, however, was substantial (.36 to .42), demonstrating that the Disclosure measure itself strongly captures the dynamic posed by the experimental condition: the effect of having negotiative power over personal disclosure, versus none at all. The contribution of the experimental condition to the Privacy combined measure, in comparison, is much more modest (.15 in Model Four).

Interestingly, the explanatory effect of the general privacy measure is similar for both Disclosure and Privacy combined as it is with the other dependent variables (and in fact has greater effect on the Power measure). Meaning, the general privacy measure did not have a greater contribution to these measures than the other dependent variables; as with the other variables, the real product measure and the comfort with disclosure factors both contributed to predicting the values of both privacy variables. The other mediating factors, with the exception of caution (modestly) in Model Four for the Disclosure variable, did not contribute to the model. The importance of the

real product and comfort with disclosure measures indicate the role Contextual Integrity plays in understanding participants' perceptions, which I will discuss in Chapter Seven??

### 6.2.5 Study One Discussion

This study manipulated the role of optional versus mandatory personal disclosure in a direct negotiated exchange, hypothesizing that the disclosure optional condition would result in higher ratings by participants of their perceptions of trust, fairness, power, and privacy. All of the nine hypotheses were supported, demonstrating that encouraging or constraining control over disclosure has direct effects on participants' perceptions of the relationship as well as their expectations of privacy and willingness to disclose. In sum, giving respondents negotiation power increased their: perceptions of trust toward the company (both generally and specifically with regards to their personal data); perceptions that the relationship was fair to them; ratings of individual power as compared to the company; their perception of having more control over the terms of the relationship; perceptions of control over personal disclosure; and ratings of conformance with respondents' privacy expectations. It weakly influenced their perception that they gained greater benefit from the exchange than the company, but in both groups, averaged responses were below the midpoint of the scale, indicating that overall the respondents felt that the company benefited more from the relationship.

Control factors demonstrated a consistent association with privacy attitudes across the dependent variables, as well as associations with other mediating factors, such as respondents' orientations towards trust, caution, their level of technical competency, their willingness to use the WearMe device, and their level of comfort with disclosing information to the company. While the experimental condition was significant across all the dependent variables, the values of standardized coefficients were modest (from .25 to lower). Its effect on the Disclosure dependent variable, however, was substantial (.36 to .42), demonstrating that the measure itself captures the dynamic posed by the experimental condition: the effect of having some versus no negotiative power over personal disclosure. Its contribution to the Privacy combined measure, in comparison, is much more modest (.15 in Model Four).

Interestingly, the explanatory effect of the general privacy measure is similar on both the Disclosure and Privacy combined dependent variables as it is on the other dependent variables (and in fact has greater effect on the Power measure). Meaning, the general privacy measure did not have a greater contribution to these measures than the other dependent variables; as with the other variables, the real product measure and the comfort with disclosure factors both contributed to predicting the values of both privacy variables. The other mediating factors, with the exception of caution (modestly) in Model Four for the Disclosure variable, did not contribute to the model.

I discuss the broader implications of these findings, and emergent themes, in Chapter 7.

## 6.3 Study 2A Findings

In Study 2A I used the mandatory disclosure condition as the control, and then I paired it randomly with each of the three assurances when presenting them to participants. The first step in the analysis

involved running a series of ANOVAs using all four conditions against each of the dependent variables.<sup>3</sup> None of the ANOVAs yielded a significant result. Thus, with the first round of analysis, all the hypotheses were rejected. Even more compelling, the results did not follow the pattern I hypothesized: that, in general, the assurance conditions would always have higher means than the control. Tables reviewing the results of these tests are available in Appendix B.

After those surprising results, I shifted from comparing all four conditions against one another to conducting a series of *t*-tests comparing the means of the control condition against each experimental condition independently. This series of *t*-tests yielded a single significant result: for the Disclosure dependent variable, the mean of the law assurance condition was significantly lower than that of the control. Unfortunately, this significant finding is the opposite of what I predicted in Hypothesis 2a; the mean of the assurance condition (law) is actually lower than the control ( $\mu = 2.5$  for law,  $\mu = 2.9$  for the control).

Finally, I ran a series of chi-squared association tests against all of the categorical dependent variables. There were only two instances in which the distribution of responses were significant: when comparing the control against the anonymity condition for the trust in company dependent variable ( $\chi^2(6) = 14.32, p=.03$ ), and when comparing the control against the indirect social reputation condition for the disclosure dependent variable ( $\chi^2(6) = 12.58, p=.05$ ). For the first case, the distribution of responses in the anonymity condition were skewed towards trusting the company as compared to the control, where the responses were more normally distributed. In the second, the responses in the control condition were more extreme (greater responses at either end of the scale, with more disagreeing that they could control disclosure), as compared to the ISR condition, where the responses skewed towards disagreeing that they could control disclosure. The first case provided weak support for H1a; the respondents in the anonymity condition were more likely to find the company trustworthy than the respondents in the control. The second did not support H2a: while participants in the control condition were more likely to strongly agree that they had more control over their disclosure than respondents in the ISR condition, the total count of those agreeing was equal between both conditions (additionally, the total count of those disagreeing that they could control their disclosure was nearly equal between conditions, and much higher than those who agreed).

### 6.3.1 Summary

As I ran the mandatory disclosure study first, my initial impression was that one explanation for the lack of significance in the experiment was respondent discontent with the mandatory disclosure condition—essentially, that the mandatory disclosure condition was inherently so disliked that even with an assurance attached respondents weren't assuaged. Another aspect was the lack of consistency in the results; it wasn't just that the differences were not significant, but the differences between the means did not follow the pattern I anticipated. I thought it was possible the differences might not be significant yet still demonstrate the control condition as the lowest performer relative to the assurances. But this wasn't the case, either. In fact, the legal assurance, which one might

---

<sup>3</sup>Note: I used a Bonferroni correction for all the ANOVAs in these analyses.

argue would demonstrate the clearest assurance given a presumed consequence of sanctioning power, often performed the worst, while the anonymity assurance performed the best. In order to try to ascertain more, I proceeded to run Study 2B, using optional disclosure as the control condition.

## 6.4 Study 2B Findings

Study 2B was a mirror of Study 2A, except I used the optional disclosure condition as the control, and then paired it randomly with each of the three assurances. Again, the first step in the analysis involved running a series of ANOVAs using all four conditions against each of the dependent variables.

Only a single ANOVA was significant in this study: the disclosure dependent variable, where the legal assurance had the lowest mean, while the control condition had the highest. This result violated both Hypotheses 2a and 2b: not only did the control condition have the highest mean, the anonymity mean was the second lowest, not the highest as I hypothesized. All of the hypotheses were rejected in this round.

Again, I ran a series of *t*-test comparing the means of the control condition against each experimental condition independently. These *t*-tests yielded two significant results, both for the Disclosure dependent variable. The control condition had a significantly higher mean than both the legal and the anonymity assurances. Again, this violates both H2a and H2b, where I hypothesized that the control condition would have the lowest mean, and the anonymity assurance would have the highest mean of all.

I ran a series of chi-squared association tests against all of the categorical dependent variables. This time there were two significant associations: disclosure and law ( $\chi^2(6) = 14.43, p=.03$ ), and disclosure and ISR ( $\chi^2(6) = 12.55, p=.05$ ). For the disclosure dependent variable, respondents in the legal assurance condition were far more likely to disagree that they had control over their disclosure to the company than the control; respondents in the indirect social reputation assurance condition were slightly more likely to disagree more strongly than those in the control. However, only the disclosure/legal assurance comparison provided any use in evaluating the hypotheses, and it violated Hypothesis 2a.

### 6.4.1 Summary

Similar to Study 2A, Study 2B produced virtually no significant results. Also similar, the response means by category did not conform to the predictions; the control condition had the highest mean four of nine times; the legal assurance mean was the highest three times and the lowest three times. The anonymity assurance, which had the highest mean most often in Study 2A, had the highest mean once and the lowest mean three times. The single time the indirect social reputation assurance mean stood out, it was the highest, whereas in Study 2B it was the lowest mean twice. In short, there was no pattern to the results either within the study or between the two studies.

## 6.5 Study Two A+B Discussion

The results of Study One demonstrated that the form of the exchange had a significant effect on respondents' perceptions of trust, fairness, power, and privacy. Possessing negotiation power increased respondents' ratings across the dependent variables. In turn, a lack of negotiation power—the take-it-or-leave-it scenario—resulted in significantly lower measures of these dependent variables. This association remained robust even after controlling for other factors.

In contrast, adding assurances to both forms of negotiation neither yielded consistently significant results nor supported the hypotheses for both Studies 2A and 2B. This outcome is curious given that existing research has demonstrated that people do rely on assurances in negotiated exchange relationships.[63] I performed additional analyses of the data in Studies 2A and 2B but ultimately it was not possible to disentangle any effect of the assurances from the influence of the forms of exchange.

### 6.5.1 Assurances in Detail

In order to attempt to disentangle why Studies 2A and 2B did not yield significant results, I first examined participant responses across all three surveys to two questions about assurances: a general question, and a vignette-specific question.

Figure 6.1 presents responses to the general question “Which of the factors below would make you more or less likely to use an online service that collects your personal information?” This question presented eleven assurances of varying forms to all respondents, with responses recorded on a five-point Likert scale from lowest (1 = extremely unlikely) to highest (5 = extremely likely), and an option for unsure (with 0 = unsure). While eight of the factors have very similar means (ranging from  $\mu = 3.44$  to  $\mu = 3.74$ , or between neutral to somewhat likely), one response has both a lower mean and wider spread in responses (“The company states on their website that: ‘Your privacy is important to us.’”), and two have higher means than the others: “A law specifically protects the type of information the company collects” ( $\mu = 4.2$ ), and “The company anonymizes all the data it collects” ( $\mu = 4.0$ ). The higher-mean responses are two of the three assurances I tested in Studies 2A and 2B.

Additionally, I also asked all respondents in the context of the vignette, “Please indicate how much each the following factors would provide you with assurance (*i.e.*, a sense of trust or certainty) about using the WearMe Device.” These responses are illustrated in Figure 6.2. Responses were recorded on a seven-point reversed Likert scale from highest (7 = strongly unassured) to lowest (1 = strongly assured), and an option for unsure (0 = unsure). In this case the responses with the lowest means represent the strongest assurance, which were: the pro-negotiation experimental condition (“If I had the option of deciding which data types the device will track”,  $\mu = 2.50$ ), and anonymity (“If the company anonymized my data (*i.e.*, my body and location data were not linked to my name or other personal information)”,  $\mu = 2.36$ ). Interestingly, the third assurance tested in Studies 2A and 2B (indirect social reputation: “If the WearMe Device was used by over a million people”) provided the least assurance of the group ( $\mu = 3.62$ ). The legal reputation assurance (“If all the data the WearMe Device collects were protected by law,”  $\mu = 2.79$ ) was nearly tied with the

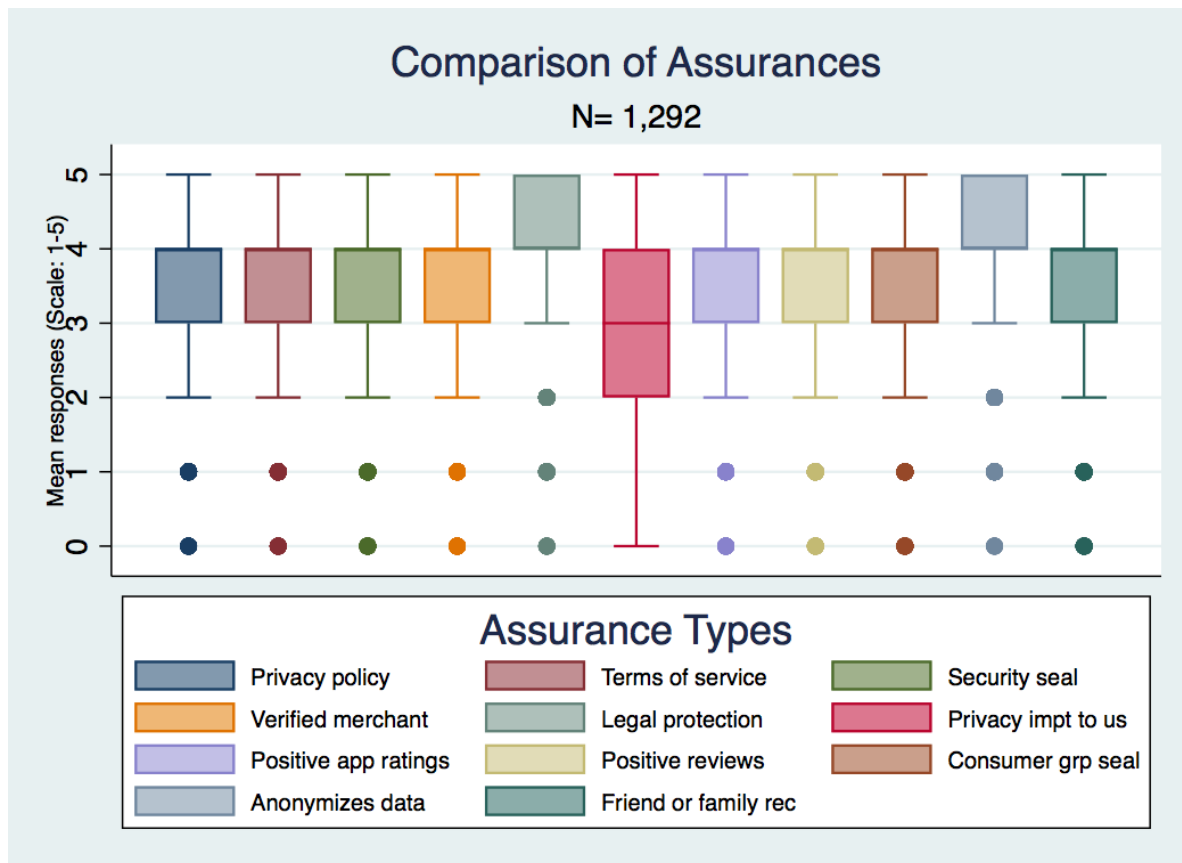


Figure 6.1: Boxplot of Assurance responses

direct social reputation assurance (“If the device manufacturer (the High-Tech Device Company) had a good reputation among its customers,”  $\mu = 2.83$ ).

The responses to these two questions indicate that at the least, all assurances were not considered equal; in responses to the general question, though eight of the eleven items were considered nearly neutral by respondents, two were clear leaders, which were also two of the three assurances included in Studies 2A and 2B. Similarly, in responses to the vignette specific assurances there were also two leaders, though the assurance with the lowest ratings was the third included in Studies 2A and 2B. But these ratings still did not translate into any pattern of responses in the two studies that demonstrated any clear effects.

Thus, even though respondents’ self-reports indicate that, at least in theory, some assurances matter more than others, why, then, did Studies 2A and 2B not produce significant results? I offer the following theories.

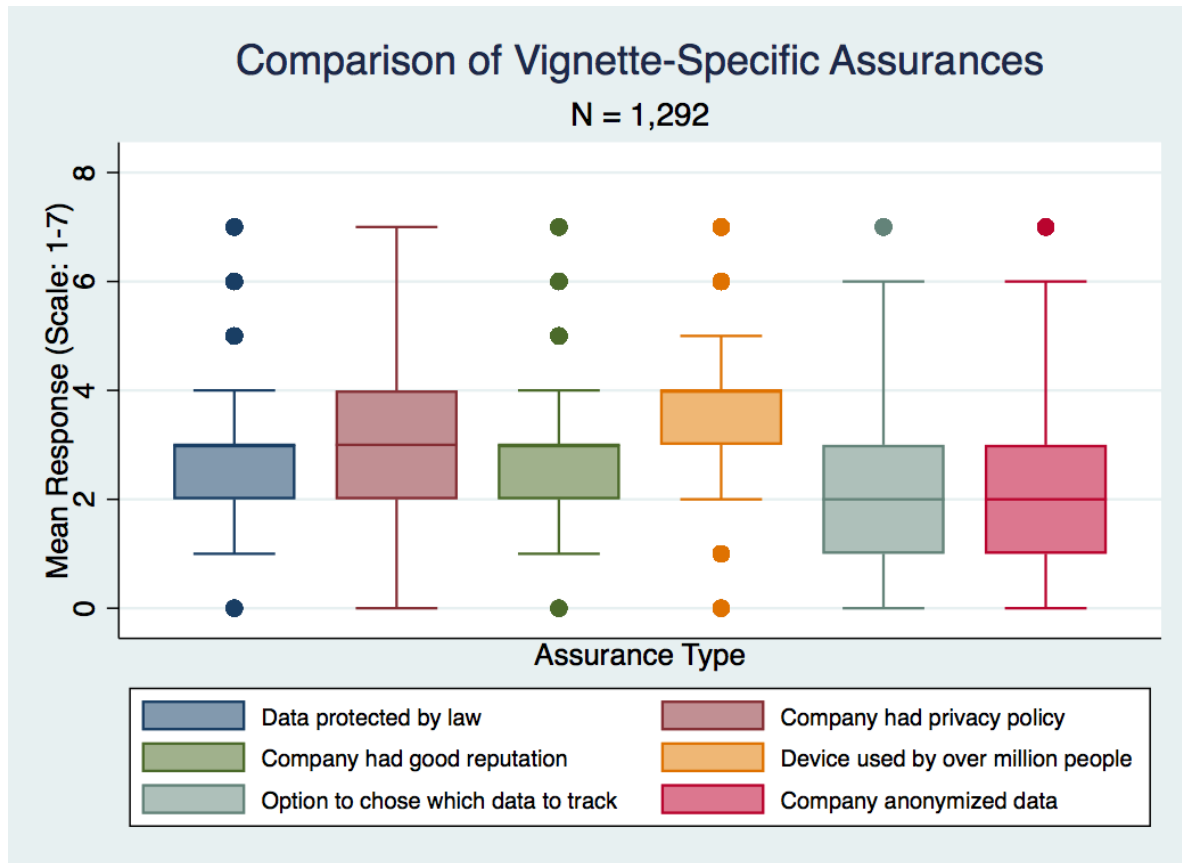


Figure 6.2: Boxplot of Vignette-Specific Assurance responses

### 6.5.1.1 Too Little Risk, or Risk Mismeasured

Assurances serve to mitigate risk and uncertainty, and it is possible that both of the studies simply did not pose enough risk to demonstrate any significant effects. In particular, I hypothesized that Study 2A, using the mandatory disclosure condition as the control, posed enough risk through mandatory disclosure that assurances would mitigate that risk. However, this assumption did not bear fruit. The entire scenario at its core may not have posed enough risk to respondents for the assurances to have any significant effects.

The vignette stated “We collect and store your data in accordance with standard industry practices” in order to constrain what respondents might infer about how the company protected their data while at the same time not introducing any confounding assurances or overly heightening their sense of risk. But in order to understand what participants did assume were included in ‘standard industry practices’ I asked the following question: “The WearMe device scenario states that “We collect and store your data in accordance with standard industry practices.” What do you believe are included in these “standard industry practices?” Please indicate how likely you think it is that each practice listed below is included.” I included seven options on a seven-point Likert scale from



### Suggested Standard Industry Practices

N = 1,042; scale ranges from 1 (extremely unlikely) to 7 (extremely likely)

Practice	$\mu(\sigma)$
The HTDC <u>does not share or sell any of its customer data</u> with other companies.	3.4 (1.9)
The HTDC will <u>share or sell your personal data only with your permission</u> .	3.9 (1.9)
The HTDC protects its customer data by using <u>strong encryption</u> .	4.6 (1.6)
The HTDC will <u>share or sell your data without your permission</u> .	4.6 (1.9)
The HTDC <u>shares anonymized data</u> with other companies that doesn't identify its customers.	5.0 (1.6)
The HTDC uses <u>basic security measures</u> to protect its customer data.	5.3 (1.3)
The HTDC has a privacy policy	5.5 (1.4)

Table 6.6: Means and Standard Deviations of Responses to Suggested Standard Industry Practices.

highest (7 = extremely likely) to lowest (1 = extremely unlikely), as well as an open text response. The questions were asked of all respondents in Studies 2A and 2B (n=1,042). A mix of high and low risk practices were presented to the respondents; responses indicate that, on the average, respondents assumed the standard practices did not include the most risk averse measures. The measures rated the least likely were that the company did not share or sell customer data, or would only do with the customer's permission.

At the same time, one item in the midrange—that the company uses strong encryption—suggests that respondents are assuming that the company by default incorporates more risk-averse practices than is arguably the norm. Based on these answers it is possible that many respondents assumed the company was taking measures to mitigate risk that it actually was not, and thus the baseline level of risk present wasn't high enough for the assurances to significantly mitigate.

Finally, it is also possible that the dependent variables I used to assess risk—trustworthiness, trust with data, and control over disclosure—were simply the wrong measures to use for assessing the effect of assurances. Future research would benefit from identifying more concise or relevant methods for assessing the direct effect of an assurance in a negotiated exchange.

### 6.5.1.2 Entrenched Cynicism Among Participants

In addition to the scaled responses for assessing standard industry practices, I also included an optional open response question: “Do you believe there was a ‘standard industry practice’ that wasn’t included above? If so, please describe here.” After removing basic responses (*e.g.*, “No,” “Unsure,” “Yes”) there were approximately 135 responses that covered a range of themes: requests for greater detail about company practices in the vignette, including a privacy policy; specific security practices, such as encryption; length of data retention; opting-out of all data collection; limiting targeting and sale of personal data; right of data removal and specificity of when data will be removed after account termination; accruing proceeds of the company sells your data for profit. In addition to a number of answers that expressed uncertainty about industry practices (from “I don’t know” to “what does standard industry practices mean?”), there were also many editorial comments about existing practices:

- “The term alone makes me believe that the company will behave similar to all the others, *i.e.* sharing too much personal data.”
- “The ‘standard industry practice’ is to collect as much information as possible and sell as much of it as possible. There might be a privacy policy, but there’s nothing private about it, it’s simply you giving them full permission to do whatever they want to do and agreeing that you have no legal recourse.”
- “That HTC will assume liability if any or all data is leaked. I don’t think they’ll do it, though, lol.”
- “Standard industry practice in the US is that they would sell your soul and make you pay extra for the service, so, no.”
- “Purposefully misleading the customer about how protected their personal information is.”
- “I’d assume unless explicitly stated otherwise that the industry practice is to maximize profit in any way regardless of the invasiveness to the consumer, which would include selling “anonymized” data to whoever is willing to pay, if the company believed the fallout would be less than the profit of doing so.”

While these did not comprise the majority of the comments, they indicate a limitation in this survey that the assurances possibly could not overcome: cynicism among the participants (and potentially the larger public) toward the intent of the companies who collect personal data. A baseline level of suspicion towards the company’s motives would be difficult to overcome, particularly if coupled with doubts about the effectiveness of assurances.

### 6.5.1.3 The Paradox of Disclosure

Another dynamic that might have affected the results is one that I call the paradox of disclosure—that the act of disclosing or highlighting a practice, even if the practice is positive—may

draw more attention to disclosure act and increase a user's concern about it. One minor finding in this survey hints at this effect. In the general assurances question, the inclusion of the statement "The company states on their website that: 'Your privacy is important to us.'" not only had the lowest mean (*e.g.*, it was the least assuring), it also had the widest spread in responses, indicating that it provoked a varied response in participants. Samat and Acquisti have found similarly that changing both the framing of a notice as well as its content under varying levels of risk can affect disclosure rates.[101] In short, for some participants the framing of the assurances may have had the opposite effect: they actually felt less assured. While the results of the studies show hints of this possibility, the differences between the controls and the assurances are inconsistent enough that this explanation can only remain as conjecture.

#### **6.5.1.4 Design Flaws**

Finally, it is possible that the display of the assurances in this particular format was simply ineffective for the purposes of this study, and that while this format worked well for Study One, adding the assurance statements confounded or disabled the effect of the manipulations.

## **6.6 Summary**

This chapter presented the results from the three experimental surveys. In Chapter 7 I will discuss in depth the implications of the results and how they fit in with the larger focus of this dissertation.

## Chapter 7

# Summary, Synthesis, and Conclusion

In this chapter, I first return to the primary research questions to review high-level findings. I then summarize the key findings from both studies and discuss related issues that emerged in the analysis. Finally, I conclude with a discussion of the broader implications for these findings.

### 7.1 Research Questions

Before diving into the specifics of the findings from each study, I first review the research questions motivating this study and how the findings address them.

#### 7.1.1 Qualitative Research Questions

In Chapter 2, I introduced the research questions for the qualitative portion of this project. To summarize the results: the individual to company disclosure relationships I explore in this study fit the definition of direct negotiated exchanges. They included multiple features of negotiated exchange: a transactional focus; benefits to the participants that were clearly understood; and a reliance by the participants on assurance structures when making disclosures. The participants in the qualitative study mentioned multiple forms of assurances that underlaid their decision to disclose, suggesting that these exchanges (from the perspective of the individual) are also binding, though the assurances they described encompass broader forms of assurance than what has been described in the SET literature.

Power differentials between the individual and company actors both existed and affected the disclosure relationship. The respondents described power in their relationships both as structural in nature (*e.g.*, Google's position in the information intermediary network provides it with far greater power in terms of controlling access to information and negotiative terms for exchange than individual actors) and based on dependency (the individuals had access to fewer comparable alternatives to obtain similar services). Finally, the length of the relationships affected them in multiple ways. One of the most obvious was that the longer the relationship, the more opportunities for disclosure from the individual to the company. In terms of power differentials, shorter relationships

had less information asymmetry between individuals and companies and fewer switching costs, while long term relationships had substantial information asymmetry and high switching costs.

### **7.1.2 Experimental Survey Research Questions**

The findings from the experimental portion of the study also provided answers to the three research questions I posed in Chapter 2. To summarize: the optional negotiation condition demonstrated significantly higher levels of trust, fairness, benefit, and power than the mandatory disclosure condition. Respondents in the optional condition reported that the exchange met their privacy expectations at significantly higher levels than those in the mandatory condition, and they also reported more control over their disclosure. Finally, introducing assurances as tested in this study did not produce statistically significant results between experimental conditions.

## **7.2 Summary of Qualitative Findings and Emergent Themes**

In this section I review the key findings from the qualitative studies and discuss the themes that emerged from this work.

### **7.2.1 The Nature of the Relationship**

First, the relationships fit the definition of direct negotiated exchanges. The relationships were transactionally focused and participants had clear expectations of what they were receiving for what they were exchanging. The extent to which a company had greater structural power (service lock-in; high switching costs; greater information asymmetry; fewer alternatives to the service; value or utility of the service) in the relationship resulted in stronger concerns articulated by participants about the relationship. The length of the relationship contributed to these power inequities between participants and the companies, particularly with regard to switching costs and information asymmetries.

### **7.2.2 Benefits and Fairness**

Across the three contexts the participants expressed that the exchanges provided them with direct benefits. Again, structural power affected the participants' sense of fairness. In particular, the participants' assessments of their relationship with Google demonstrated that one can be in a relationship in which one's benefits outweigh concerns about fairness, but at the same time one can still characterize the relationship, or at least aspects of it, as unfair.

### **7.2.3 Assurances and Trust**

As SET predicts, the participants relied on assurances in these negotiated exchanges rather than on a relationship built on trust, specifically those that the participants believed provided them with

a form of protection or guarantee when disclosing to a company. In this sense, the presence of those assurances helped enable disclosures of personal information. They relied foremost on informal structural assurances lacking sanctioning power, such as a company's reputation, the visual design of their website, the quality of the information or content on the website, the participant's perception of anonymity while using the company's service, expectations that the companies were interested in aggregate rather than individual analysis, minimal disclosure requirements, and adherence of the company's practices (specifically advertising practices) to the participants' existing mental models. The participants also relied on institutional assurances, such as government regulation and the legal system, but these were often characterized as ineffectual, difficult to access, or slow to respond. These findings suggest that in daily life people rely on a broader set of less formal assurances than those typically identified in lab studies when making disclosure decisions.

#### **7.2.4 Personal Disclosure**

The participants generally opted to disclose as little personal information as possible when given the choice, and they characterized the relationships in which they held negotiative control over their disclosure decisions as fairer and of greater benefit to them. One of the key expectations that the participants held was that the companies were less interested in them as individuals and more interested in aggregating their data with others'. This assumption minimized the participants' sense of risk and acted as an assurance structure to support disclosure. Most were fine with contributing to a company's aggregate data collection as long as the information collected did not violate norms of contextual integrity, and their disclosures did not result in their being individually targeted for advertising purposes or subject to inferences designed to personally manipulate them. The participants also utilized various strategies to protect their privacy and minimize over-disclosure, such as: neglecting to opt-in to public information sharing, using browser-based strategies (private windows, logging out, and clearing cookies).

#### **7.2.5 Risks and Trade-Offs**

Participants expressed concern about specific risks to their privacy across all three contexts, which I classified as risks of: targeting and manipulation; exposure and embarrassment; and violations of contextual integrity. The assessment of risks were lowest among the pregnancy app users, and were the highest and most varied when assessing online search. At the same time, as elective users of most of these services the participants felt that the benefits and trade-offs offered by these services outweighed these risks. While the perceived privacy risks did not discourage use, several participants engaged in strategies to reduce their risk exposure as described above.

### **7.3 Emergent Themes**

In addition to findings that addressed my research questions, I observed several emergent themes during my analysis.

### 7.3.1 The Importance of Power in Disclosure Relationships

Power differentials between the participants and the companies manifested as dependency (the availability of comparable service providers for the participants), and as structural power based on the company's relative position in the network of information intermediaries. A participant's dependency on a company's service was a function of the availability of other companies (exchange partners) with similar offerings. This dependency was noticeably lacking in the pregnancy app space, where there were multiple companies offering similar products. As such, the participants experienced minimal lock-in, fewer trade-offs, and lower switching costs. These apps also made the most minimal demands for information disclosure. In contrast, the participants felt most dependent on Google, as there are fewer options for online search of comparable quality, and the lock-in and switching costs from their use of other Google products are substantially higher. Perhaps not coincidentally, Google's defaults promote maximal data collection and the eternal storage of one's search queries. While aspects of the company's privacy settings are negotiable, the burden is on the user to discover and alter them.

Companies hold greater structural power in the information services marketplace than consumers do. The non-negotiable terms companies present to customers is one example of this uneven structural power. Large companies have the resources to directly influence power structures to ensure marketplace-wide acceptance of take-it-or-leave-it terms as the default practice. Thus, smaller companies may benefit from may benefit from the personal information collection practices of more powerful companies, from their influence on legislation and policy, as well as their leverage for setting the standard for the acceptability of business practices. While regulators can and do step in to alter this arrangement, the existing process enabled by the legal system today puts individuals at a disadvantage. For example, while the FTC has taken enforcement action against companies who violate their own privacy policies or are likely to cause substantial consumer injury, they have not directly regulated companies to set substantive limits of the types of information collection practices that may harm consumers. However, an effect of their enforcement actions has been to establish certain types of information collection practices for companies to avoid, as well as processes they must follow (*i.e.*, comprehensive information privacy programs).[10]

Further, the reliance of the participants on structural and informal assurances suggests that individuals may perceive institutional assurances to favor or be co-opted by more powerful actors or be expensive or otherwise difficult to access. For example, the participants who responded to my questions regarding pursuing legal avenues to rectify harms generally dismissed these options as too difficult or expensive for an individual to navigate (assuming an individual even retains the right to pursue a lawsuit against a company instead of being forced into arbitration). In contrast, the one tool multiple participants mentioned was the power of public shaming through social media and its potential to ensure that a company actor would uphold their end of the exchange. Further, social media tools allow individuals both to directly connect with companies and potentially build coalitions with other customers to increase their structural power, as suggested by Cook and Rice[28]. The turn to mechanisms that are outside the traditional legal and institutional redress channels may also be a response to feelings of powerlessness produced by a marketplace and regulatory structure that pretends choice and negotiation are equally available across the marketplace when in fact they

are not.

### 7.3.2 The Length of the Relationship

This study demonstrates that repeated information exchanges over time can contribute to information asymmetries and an increase of structural power for one actor at the expense of another, further exacerbating existing inequities. As several participants' long-term exchange relationships with Google demonstrated, substantial information asymmetries accrue over time, in turn giving the company greater structural power. Of course, the challenge to understanding this phenomenon is in documenting and assessing these inequities and their consequences when the information is privately held and the disclosures are so incremental. The effect on disclosure decisions of the interaction between the length of an information exchange relationship and power differentials needs more research before we can identify with greater certainty how these aspects affect exchange relationships.

### 7.3.3 Mediating Effects of Indirect Generalized Exchange

The participants' enthusiasm for contributing to research was a fascinating aspect of the 23andMe interviews. Their perception that they could both gain a direct benefit from the service while also contributing to the public good through the company's scientific research was a strong motivating factor for most and neutralized the risks participants held about the service. While it was not the sole factor that encouraged participation in the company's research, among this set of participants it was the most influential.

The effectiveness of the indirect generalized exchange depended on a number of assumptions by the participants themselves. First, the participants generally assumed that the value of their contributions were not individual but as part of the aggregate. Thus, when they were contributing their personal information to these large information pools (*e.g.*, 23andMe, or Google's search algorithms) they assumed their identities were anonymized, and that even if their contributions were identifiable, the company would not need to identify them. Some also articulated their belief that no humans ever came in contact with their data, and as such their privacy concerns were lessened as they were not being singled out. These assumptions led the participants to discount the risk contributing posed to them.

Further, the public's awareness that large predictive systems require massive amounts of training data in order to build reliability and precision at scale appears to be growing widespread, as well as the belief that their participation is necessary for improving these systems. As one participant put it when describing Google, "[t]here's this sort of weird 'fair trade.' You continuously get better and more improved services, in exchange for your data." [P16]

However, a private company using proprietary data to conduct research is not the same as, say, public health researchers using personal health information to do the same. Private companies engaged in research are not obligated to constrain their work to benefit the public (or the individual), and are not subject to the same oversight to ensure ethical compliance as academics are. Nor would 23andMe's decision to, say, use their proprietary database to identify a genetic disposition



for sweets, and then collaborate with the pharmaceutical industry to target weight loss drugs, or to target advertising at those people. Essentially, 23andMe's strategy appears to be the distillation of a larger trend of encouraging disclosure through a promise of indirect benefit and contribution to a greater good. After all, don't all Google users benefit if Google uses our collective data to improve its search algorithms? As companies continue to build out artificial intelligence at scale, encouraging current and potential users to give up their data in the name of greater good—even if the “public” benefit is limited to improving a company's product for its customers—is critical. Eric Horvitz and Deirdre Mulligan tackle the tension between big data for the public good versus increased privacy challenged in an article in *Science*, arguing that “[w]e need to strike a new balance between controls on collecting information and controls on how it is used, as well as pursue auditable and accountable technologies and systems that facilitate greater use-based privacy protections.”[55] I add that better understanding of how well promises of indirect generalized exchange motivates us, and what those promises are contingent on, is also crucial for understanding the dynamics of disclosure at scale in the next ten years.

As a motivation to inspire participation using indirect generalized exchange is an ingenious strategy. The potential research benefit of the public's data provides a noble and altruistic gloss to company requests for personal information. At the same time, it implicitly suggests that failure to opt-in is free riding, as we stand to benefit from research on a data set to which we ourselves won't contribute. The appeal of supporting research appears to be premised on a mental model formed around traditional academic, and regulated research. Policy makers may need to consider whether consumers are being misled by corporations generic statements about research, particularly where the personal information involved in the exchanges includes sensitive information such as genetic data.

## 7.4 Summary of Survey Experiment Findings and Emergent Themes

In this section I review the key findings from the survey experiments and discuss the themes that emerged from this research.

### 7.4.1 Study One

Study One demonstrated that manipulating the terms of optional versus mandatory personal disclosure in a direct negotiated exchange resulted in significantly higher ratings by participants of their perceptions of trust, fairness, power, and privacy in the disclosure optional condition. The regression models demonstrated the robustness of the experimental conditions and yielded insight into the general and vignette-specific control factors that affected the predictor variables. With three exceptions (Model Three for Trust and Power, and Model Four for Trust), the experimental conditions remained significant across all models when controlling for other factors.

Consumer choice is often touted in the U.S. as an unalloyed good. The results of this research study could be interpreted within that same framing: that by allowing consumers more choice, or

greater control, over the terms of disclosure, they in turn report greater levels of trust, fairness, power, control over disclosure, and greater consonance with their privacy expectations. However, I believe they suggest a different conclusion: that when one clearly articulates a mandatory disclosure practice, particularly one that requires maximal disclosure, users perceive it negatively. This isn't simply a reaction to having fewer choices; the participants rated the mandatory disclosure condition significantly lower as compared to the optional disclosure condition across all of the dependent variables. I present additional data below that challenges interpreting these findings as evidence for increasing user choice, suggesting that even when consumers are given greater choice over disclosure they still ultimately find the terms of disclosure objectionable.

### 7.4.2 Privacy Attitudes

In my regression analysis, the factor that provided the greatest explanatory power across all of the dependent variables was the general privacy scale, a scaled measure of the respondents' privacy attitudes towards institutional information privacy. All of the dependent variables had a negative relationship with this covariate, indicating that increases in the dependent variables were associated with a lower general concern for privacy. However, it is important to contextualize this lower concern; the mean of the privacy scale across the entire respondent pool is 4.8 on a scale of 1 to 7 with a median of 4.85; higher values indicate increased privacy concern. These scores demonstrate that as a whole the respondents' responses skew towards favoring privacy. The difference between means for the composite privacy score by experimental condition was not significant (mandatory disclosure  $\mu = 4.80$ ; optional disclosure  $\mu = 4.82$ ); neither were *t*-tests for all of the individual questions. Another general privacy question (not included in the scale) confirms this finding; responses to "In general, how important is it to you that companies respect your privacy?" skewed highly in favor of privacy concern, with 98 percent of the respondent pool reporting somewhat or very important. Thus, the respondent pool overall evinced a baseline moderate concern with privacy.

### 7.4.3 Implications for Design

My findings, similar to Brandimarte *et al*, have implications for design that can be viewed as either positive or negative for privacy. For example, if one follows the paradox of control findings to their logical conclusion, in order to encourage over-disclosure, one should give users as many options for control as possible. Similarly here, increasing negotiative power increased perceptions of trust, fairness, and individual power, yet also led to increased disclosure. As Brandimarte *et al* note, "higher levels of control may not always serve the ultimate goal of enhancing privacy protection." [19] Thus, it is vital to place these findings in a larger context and acknowledge that solutions that seek to protect privacy must be multifaceted while acknowledging the cognitive biases that challenge our individual abilities to manage our disclosure. In addition to limiting the amount of information collected in the first place, defaults should skew towards minimal disclosure, all disclosure should be opt-in rather than opt-out, and designs in general should seek to balance our collective need for privacy as a social good while allowing for individual variation.

Like Brandimarte *et al*, a goal of this study is to document a phenomenon and the conditions that produce it. A side-effect of this work is to highlight the means by which these conditions are created. My findings suggest that the participants were generally dissatisfied with the exchange scenario presented in this study because across both experimental conditions the terms are disempowering and reflective of the current status quo. While consumers may appreciate having choices, having a choice between two unfair options does not address the underlying unfairness. The extent to which these findings can be generalized to suggest that the public in general is dissatisfied with similar exchange scenarios is debatable, but I will argue that they point to the need to give consumers greater power in their negotiations with large companies over their personal information.

#### 7.4.4 Negative Perceptions of Company Relationships

Additional data from this study also demonstrates that while the participants' perceptions of the mandatory disclosure relationship were more negative than those in the optional disclosure condition, at the same time both groups provided relatively low ratings of their relationship with the company across multiple aspects. These findings highlight similar themes present in the qualitative portion of this dissertation related to the dependent variables in this study. While the questions here were vignette specific, the results suggest that examining these aspects in other disclosure relationships may be useful for establishing the extent to which these concerns hold true more broadly.

For example, in examining the two power-related dependent variables (power in relationship, control over terms), the means and medians for these measures demonstrate that the respondents overwhelmingly rated the company as having more power. The scale for each question runs from 1-10, with responses below the midpoint (5) favoring the company ("The company has more power;" "The company has greater control over the terms"). The responses for the relationship benefit measure, which was not significant between the two experimental groups, also favor the company over the respondents. And even respondents in the optional disclosure condition ranked their control over their data in the disclosure measure close to the midpoint of the scale (3.83 compared to a midpoint of 3.5), demonstrating that even when given power over disclosure, respondents still rated their ability to control their data as fairly low. Which is to say that while the optional disclosure respondents provided significantly higher ratings across all but one of the dependent variables, those ratings still illustrated a relationship where the company was perceived as having greater power, more control over the terms, and more control over the disclosed data. Table C.1 presents the means and median values for these questions in Appendix C.

A series of questions I asked the respondents about their relationship with the company yield additional insight on this point. Table C.2 presents the questions and summarizes their values in the Appendix C. I posed nine questions asking about different aspects of the relationship, measured on a seven-point Likert scale from strongly disagree (1) to strongly agree (7). I included an unsure option (0) for these questions as I was concerned that some respondents might not understand what the questions meant, and as I had no previously validated questions to draw from. I created these questions based on characteristics of negotiated and reciprocal exchange relationships with the

intention of using them to ascertain the extent to which the relationship had features that mapped to one of the two relationship forms.

Evaluating these question in the scope of their intended purpose indicate that the participants did agree more strongly with the statements representing characteristics of a negotiated relationship (statements F, G, H, and I) versus those more aligned with a reciprocal relationship (statements A, B, and C, and E). Statement D was included to assess the extent to which participants perceived the relationship as direct (one-to-one). But the responses can also be evaluated on their face as substantive evaluations of how participants perceived the relationship. Doing so demonstrates that the experimental condition affected some of the responses, but more broadly these responses also illustrate a predominantly negative interpretation of the relationship.

Three of the statements were significant in a *t*-test of means by experimental condition: “We are equal partners”; “Either one of us is free to end the relationship at any time”; and, “I have a positive feeling toward the relationship”. For each of these three statements, the optional disclosure condition had a significantly higher mean than the mandatory disclosure condition. The optional disclosure group agreed at higher levels that the partnership was equal ( $\mu = 3.4$  vs. 3.0), that they had positive feelings toward it ( $\mu = 4.1$  vs. 3.6), and that they could leave the relationship at any time ( $\mu = 4.9$  vs. 4.5) than the mandatory disclosure group. Only one of these means was above neutral (the midpoint of the scale, 4.5), and looking across the entire set of questions, the aggregate means for five of these statements were below the scale midpoint. As a whole, all of the respondents disagreed that they were equal partners with the company, did not have a positive feeling towards the relationship, and disagreed that the relationship was exclusively between themselves and the company. The respondents also disagreed with an assurance-focused statement (“The relationship relies on someone or something else to make sure we both hold up our ends of the agreement”), as well as with a statement intended to assess reciprocity (“Only one of us contributes to the relationship”). There were only four statements where response means were slightly above neutral, indicating weak agreement: that the relationship relied on an agreement; that it was based on trust; that there were clear expectation of what both parties contributed; and that both parties were free to end it at any time.

In all, these findings suggest at least two interpretations. The first is that the respondents had an overly negative view of this specific context. While this is certainly possible, participant responses to other general questions suggest otherwise. In response to the question “If the cost were not a concern, how likely would you be to use any kind of self-tracking wearable device?”, 31 percent were slightly to extremely unlikely, 10 percent were neutral, and 59 percent were slightly to extremely likely to use a wearable device. A comparison of means between experimental groups on this question was not significant. While this question doesn’t provide a check on participants’ attitudes towards companies generally, it does suggest that the majority of the participants were not inherently biased towards the type of product used in the study.

The second interpretation is that this pool of respondents might be biased towards online companies that engage in information collection. However, additional data appears to challenge that interpretation. Answers to the question “I appreciate that online services are more efficient because of the increased access they have to my personal data” demonstrate that 41 percent of the respondents somewhat to strongly agreed, 25 percent were neutral, and 34 percent somewhat to

strongly disagreed with this statement. A comparison of means between experimental groups on this question was also not significant. Again, these results suggest that these respondents were not biased against information companies in principle. Responses to another question do suggest concerns with risk: “I think there is little risk with sharing my personal information with online companies.” Answers to it, also not significant by experimental condition, ranged from 23 percent of the respondents somewhat to strongly agreeing, 11 percent neutral, and 66 percent somewhat to strongly disagreeing, indicating that respondents did have general concerns with online personal disclosure. In sum, this data suggests perceptions of negativity towards this form of exchange relationship independent of the experimental condition and also supported by the qualitative portion of this research project. Additional research would help determine whether these perceptions are widespread among the public or limited to this pool of respondents.

#### 7.4.5 Privacy Paradox

Study One also supports Brandimarte *et al*’s key finding that as individual control increased, disclosure also increased. Here, as respondents’ control over disclosure increased, so did their comfort with disclosure. Respondents in the mandatory disclosure condition reported less control over disclosure, which was also associated with less comfort with disclosing to the company.

A contrast between this study and Brandimarte *et al* revolves around the question of risk. In Brandimarte *et al*, the researchers varied the level of risk across experiments and found that the “privacy paradox” effect held even when objective levels of risk were higher. This study supports those findings: lower assessments of risk were associated with greater control over disclosure. However, it is important to note that I did not use the same methodology for assessing the effect of risk. In these experiments I held risk constant (I did not vary it across the two conditions) and attempted to assess its influence by posing a series of questions introducing specific risk scenarios to the vignette: “For this question, assume that the High Tech Device Company anonymizes all of the data it collects from you (meaning, your body data and location data are not linked to your name or other personal information). How much risk (the possibility of something bad or harmful happening to you) would each of the following scenarios pose to you?” I asked respondents to assess the level of risk posed to them if their anonymized data were sold by the company, shared with another company for advertising purposes, or if all of their collected data were stolen by hackers.

The second scenario posed the same three questions about identifiable data: “For this question, assume that the High Tech Device Company does not anonymize all of the data it collects from you (meaning, your body data and location data are linked to your name or other personal information).” Overall, respondents rated the anonymity scenario as less risky than the identifiable data scenario. The risk level ratings were significantly different between the two groups for two of the anonymity scenarios (data sharing and theft), and only the theft scenario for the identifiable data question. While these scenarios were not part of the vignette and thus did not directly affect respondents’ evaluation process, to the extent that the mandatory disclosure condition posed an inherently higher level of risk due to its requirement for mandatory disclosure, this manipulation did have an effect on the participants’ assessment of potential risk. The mandatory disclosure re-

### Anonymization Risk Scenarios

N = 250. Responses are scaled from 1 (Not at all risky) to 5 (Extremely risky)

<b>Anonymization Scenario</b>	$\mu(\sigma)$	<b>NY</b> $\mu(\sigma)$	<b>NN</b> $\mu(\sigma)$	<b>p-Value</b>
All of my WearMe data was sold by the HTDC to another company.	2.8 (1.1)	2.7 (1.2)	2.9 (.98)	—
All of my WearMe data was shared by the HTDC to another company for advertising purposes.	2.9 (1.1)	2.7 (1.1)	3.0 (1.1)	$p \leq .05$
All of my WearMe data was <u>stolen by hackers</u> .	3.4 (1.2)	3.2 (1.3)	3.5 (1.2)	$p \leq .05$

Table 7.1: Anonymization risk scenarios.

### Identifiable Data Risk Scenarios

N = 250. Responses are scaled from 1 (Not at all risky) to 5 (Extremely risky)

<b>Identifiable Data Scenario</b>	$\mu(\sigma)$	<b>NY</b> $\mu(\sigma)$	<b>NN</b> $\mu(\sigma)$	<b>p-Value</b>
All of my WearMe data was sold by the HTDC to another company.	3.9 (1.1)	3.8 (1.1)	4.0 (.95)	—
All of my WearMe data was shared by the HTDC to another company for advertising purposes.	3.8 (1.1)	3.8 (1.1)	3.9 (1.1)	—
All of my WearMe data was <u>stolen by hackers</u> .	4.4 (1.0)	4.2 (1.2)	4.5 (.85)	$p \leq .05$

Table 7.2: Identifiable data risk scenarios.

spondents reported that both the theft scenarios, as well as the anonymized data sharing scenario, posed significantly more risk to them than the optional disclosure respondents.

## 7.5 Synthesis and Discussion

This dissertation demonstrates the utility of social exchange theory for further understanding individual decisions to disclose personal information. In particular, it provides a framework for understanding the social context of a disclosure relationship and the impact of social structure on it. This research demonstrates that the social context does, in fact, have an effect—at least on the contexts I examined—on individual decisions to disclose personal information.

Based on this research, I believe that SET can add a valuable perspective to privacy research. Specifically, this perspective fills a gap present between the two research traditions I discussed in Chapter One: privacy as control, and contextual integrity. Privacy as control, with its individualistic approach to understanding disclosure, places the locus of control entirely in the hands of the individual without acknowledging that individuals are subject to social structural forces beyond their direct control. While both companies and individuals may be considered persons before the law in the U.S., they do not possess equal standing in the marketplace. As long as we continue to evaluate privacy and disclosure as if individuals were equal actors in these relationships, the distribution of power that informs the conditions under which decisions are made constrains the options individuals have to control disclosure in important ways.

Contextual integrity is useful for identifying normatively inappropriate flows of personal information and understanding why they constitute privacy violations. However, I think SET helps to identify the dynamics contributing to the violation in a way that CI may not. For example, consider the difference between analyzing personal disclosure using contextual integrity as the framework as compared to SET, with search engine use as the context. The goal of CI is to uncover normatively inappropriate flows of personal information that violate the context in which it was collected. Within the search context specifically, Nissenbaum writes that “[c]onsulting a search engine. . . is akin to conducting research, seeking information and association, searching a library catalog, and pursuing intellectual enlightenment. . . [i]f I am right about how search engines are used and for what purposes, then the governing norms would be strict confidentiality with regard to Web search histories and perhaps, as practiced by many public libraries, the prompt expunction of such records to minimize risks of leakage or mandated handovers as well as the temptation of future sharing for financial gain.”[85] CI locates the potential privacy harm in the violation of norms governing the activities or purposes of that sphere of social life—here, the use and reuse of the collected information outside of the context in which it was divulged—based on norms derived from corresponding activity in the non-digital realm.

Nissenbaum locates the contextually inappropriate flow of information based on past models of information seeking that she considers analogous to web search. My SET-based analysis did not unearth this line of thinking from my participants, but instead registered concerns about information asymmetry, individual targeting based on search queries, and inappropriate or inaccurate inferences by Google. These concerns are focused on the relationship between the participants

and the company and the power imbalance between them rather than the primary action itself (the collection and storage of search queries). I would argue that these approaches are not mutually exclusive but complementary; if SET informs us more broadly about the way in which people reason about disclosure this information will help us understand more about the composition of context. Identifying privacy risks is not limited to laws, ethics, and norms. Structural factors also create risk. From the perspective of this study, a participant's direct negotiated exchange with Google is marked by structural inequality: Google occupies a far more powerful position in the network of information intermediaries and consumers than both the individuals and other company actors. As such, the company's ability to dictate the terms of the exchange far exceeds the individual's. Depending on the length of the exchange relationship, it may further be influenced by aggregated power on the part of Google (manifest as information asymmetry) and power dependency by the individual on Google if her ability to obtain comparable resources from other actors is limited. The individual may rely on an assurance to bind the terms of the relationship, but as this study demonstrates, individuals more often than not lack the power to dictate the terms of those assurances. These conditions can place an individual in the position of disclosing against her best interests.

In Chapter 1, I wrote that Nissenbaum calls for the norms themselves to be judged in terms of "how they impinge on societal values, such as equality, justice, fairness and political liberties." CI doesn't provide an obvious means for doing this.[16] In contrast, SET provides a framework for evaluating these values at the micro-level of social interaction, and to potentially generalize upward to macro-level scale. Structural inequality isn't a problem limited to a single individual, but rather impacts a society at large.

### 7.5.1 Contributions to Privacy Literature

This research demonstrates that the power differentials that exist between individuals and companies can impact the individual's personal disclosure decisions. While power differentials may not be present in all disclosure relationships, or even every individual to company relationship, for researchers who are attempting to understand real world disclosure dynamics between individuals and companies these findings make an important contribution towards .

Relatedly, these findings also provide an additional perspective for interpreting the privacy paradox. They demonstrate that power dependency is a factor in disclosure relationships, and that an individual's limited access to alternatives may tie them to a disclosure relationship that they would prefer to exit if they could obtain a comparable service that addressed their privacy concerns. This undercuts one of the assumptions of the paradox: that people are willingly electing to use these services despite their information collection practices. Further, these findings show that in these relationships people engage in strategies to deliberately mitigate their exposure, demonstrating both an awareness of information collection practices and strategic planning to thwart them.

There appear to be several assumptions underlying the privacy paradox: that there is a competitive marketplace resulting in multiple options for consumers for information services or the ability to port their data between competitors; there is competition between actors on information collection practices or for privacy enhancing technologies that result in higher privacy standards; and that consumers have leverage to negotiate the terms of their exchange. My findings challenge



these assumptions, and suggest a bleaker interpretation—that consumers continue to engage with companies whose practices they dislike because they don't have better options, they do not have the power to negotiate beneficial terms, and because structural factors benefiting the most powerful companies (such as a lack of privacy legislation) have allowed companies to collect personal information with few consequences. Robust privacy protection requires addressing the overall conditions of the marketplace in which decisions take place, not just mechanics of decision-making.

### 7.5.2 Contributions to SET Literature

This dissertation contributes to studies of social exchange theory in several ways. First, I expand on the work of Foa and Foa[44] and Cheshire *et al* [22] in using an exchange of information goods, specifically personal information, as an object of social exchange. The integration of the privacy of information as an exchange good extends the application of social exchange theories that have focused tightly on experimental work in areas such as commitment, trust, affect, and power, rather than in real world studies of emergent valued resources such as digital information.

My research also suggests that there is more complexity and nuance to assurance structures than both the experimental and theoretical literature has addressed to date. As Cheshire *et al* note, “[i]n natural settings, agreements are most often guaranteed by law or a third-party mediator, while in experimental settings they are ensured by the design and structure of the environment created by the experimenter.”[24] The qualitative portion of my study suggests that assurances in natural settings are more broadly construed, as the participants relied on structural rather than formal assurances to protect their privacy. In particular, the participants believed that anonymity or relative obscurity provided by aggregate data protected their privacy. While lack of knowledge about legal protections may explain some participant reliance on structural assurances, my research suggests that even those who are aware of formal assurances still rely on structural ones to protect privacy due to perceived limitations of the legal system and other formal mechanisms.

These findings indicate that researchers should be cautious in applying tightly formulated experimental findings to real world situations involving assurances. For example, to assume that an negotiated exchange is binding because a law exists to prevent a breach of agreement might be overreaching given that a lay interpretation of assurance might diverge considerably from a formal conceptualization. Structural assurances, transaction costs, and social assurances, such as public shaming, should be modeled as assurances given that in the real world they inform people's decision-making.

### 7.5.3 Reconciling Assurances Between Studies

In the qualitative portion of my study, assurances played a key role in participants' decisions to disclose. However, in the experimental portion of this dissertation the assurances as manipulated in Studies 2A and 2B were not significant. What do these results imply?

I would argue that, foremost, the experimental results should not lead us to dismiss the importance of assurances. The qualitative results were clear that assurances mattered to the participants and influenced their decision-making. However, it might be that when signing up for a service,

only certain forms of assurance are salient to one's decision-making process (such as the structural assurance of visual design). Once one is already a customer and has had time to reflect on the experience other forms of assurance may be more salient. Returning to Daniel Kahneman's theory of System One and System Two cognitive processing, it is possible that when we are engaged in the process of evaluating a new online service, our cognitive attention relies on the intuitive and heuristically driven System One to assess risk rather than the contemplative processes of System Two.[58] Thus, formal assurances such as privacy policies and legal protections may not resonate with most of us when engaged in a sign-up process where we are focused on evaluating the benefit of a service. In contrast, once a relationship is already established we are able to consider assurances that rely on System Two thinking and assess potential risks with more care.

Phelan *et al's* research into the privacy paradox using Kahneman's systemic thinking as the framework demonstrated that their participants only engaged in considered concern (System Two) of a privacy threat when their intuitive concern (System One) registered a privacy risk. Otherwise, the participants generally made decisions based on their intuitive judgments: "because the impressions generated by System One processes usually cannot be articulated, individuals evaluating considered concern may not understand the influence of intuitive concern. Therefore, if the benefits of disclosure are sufficiently large or they can achieve low considered concern, they may disregard their lingering intuitive concern." [91] Thus, from this perspective, the assurances I presented in the experimental studies may simply have not been salient to the participants' decision-making processes, and/or the experimental manipulations did not spark the participants' intuitive privacy risk, thus rendering them inconsequential to their decision-making.

#### 7.5.4 Suggested Interventions

Having identified the structural power differentials and power dependencies in these exchange relationships, the question remains: what can we do about them? I do not have a comprehensive or detailed solution to propose, but I will outline a few starting points here. To the extent that decisions to disclose data to companies are clearly impacted by these factors, we must first recognize this and understand the impact. There are structural inequalities that many companies benefit from, such as take-it-or-leave-it contracting terms, making the problem pervasive. Today individuals cannot negotiate meaningful terms and agreements with companies. The formal assurances intended to support these agreements favor company actors over individuals. Redress is difficult to obtain, particularly in an era where consumers are forced to agree to terms, such as mandatory arbitration, that limit their coalition-building power, such as class-action lawsuits. Schemes that could amplify the ability of individual consumers to build coalitions to specifically counter the power of larger companies to control personal data might hold promise.[46]

There is increased recognition in policy circles that a lack of competition among large technology firms leads to information collection practices that harm consumers, and that conventional competitive analysis may underestimate the power of platforms when a source of their power is the customer data they hold and their ability to offer their services for free. Law professor Howard Shelanski argues that "holding price, service quality, and everything else constant, digital platform customers would rather reveal less information about themselves, and would prefer that those

platforms maintain strong, rather than weak, privacy policies regarding the data that customers do disclose. . . [t]o the extent that competition promotes improved services and privacy policies, anti-competitive conduct diminishes both of these consumer benefits. In conventional antitrust terms, anti-competitive conduct can enable a platform to extract more information from customers without offering the level of quality a consumer could barter for in a more competitive market.”[104] In the case of market monopolization by large digital platforms, one could start with the most obvious approach in the U.S. of reducing their power through anti-trust mechanisms. In the specific case of digital platforms, however, Shelanski argues that any competitive analysis scrutinizing a platform should focus on the role of customer data in a firm’s conduct, as the often “free” cost of digital services may obscure the effect of data on a competitive effects analysis where the focus is often on the price of services. “Recognition of the role of consumer data as an input in digital platform products could therefore show competitive effects that are unrelated to prices or other terms on which the platform provides services.”[104]

What other solutions might provide consumers with greater power to negotiate their terms of disclosure? The European Union’s attempt to rebalance power in the hands of consumers materialized in the creation of a right to data portability, anchored to the concept of individual control, which will go into effect in May 2018. Companies with EU-based customers must provide them with the means to share their data with another provider by allowing them to request their personal data in a machine readable format.[49] Both the legal implications and implementation considerations of this approach are beyond the scope of this dissertation to consider, though Michelle De Mooy of the Center for Democracy and Technology has authored a detailed report that weighs the pluses and minuses of this approach. She notes that “[p]roponents of this concept believe it offers a way to give people a power of self-determination with regard to their information in big-data systems, leveling the playing field between individuals and the commercial and noncommercial entities that capture and share their information.”[31] At the same time, the challenges it presents for individuals to sort through the complexities are enormous, and the dominance of large technology companies means that “they would be less affected by users’ ability to move their information from place to place than would small business operators.”[31]

The GDPR data portability right assumes that individuals will derive power from direct control over their data, and the ability to decouple it from one platform and potentially move it to another. This may be the case. However, before we move to reproduce this model in the U.S., we should consider other possibilities as well, such as: prohibiting the collection of some forms of data entirely; regulating or severely curtailing the data brokerage industry; requiring an affirmative opt-in for any and all data collection; and, completely rethinking the notice and consent framework, which could include standardizing the consent interface and terminology, contemplating measures such as visceral notice[21], or proposing both a new legal and interface design strategy that radically departs from existing paradigms. Emergent technologies, such as blockchain, could also make it possible for people to store and manage personal data, granting or denying access, in ways that previously were considered too inefficient and cumbersome to implement.[69] Ultimately, any solution is likely to have multiple features, as this presents a problem that I argue cannot be solved either solely through legal or policy measures or by changes to interface design.

## 7.6 Conclusion

This dissertation forges new ground in the analysis of information privacy and personal disclosure. Namely, it demonstrates the utility of the relational analytic approach for identifying social structural factors that affect personal disclosure. This approach yields a different set of insights into the dynamic of personal disclosure and information privacy. It reveals the impact of power differentials on personal disclosure outside of the surveillance context, demonstrating that imbalances in power between individuals and companies can affect individual decisions to disclose.

The mixed-methods approach I used provides unique insights that either study alone might have missed, with the most substantive example being assurances. Assurances were clearly an important aspect of the disclosure relationship as the qualitative interviews demonstrated, but they did not have a significant effect when tested in the experimental portion of the study. The qualitative interviews also yielded additional nuance into assurances that would be difficult to discover through surveys. This experience demonstrates the value of a mixed-methods approach when researching a novel area.

But beyond merely revealing these aspects, this work demonstrates the value of examining disclosure more broadly than as a problem restricted to individual cognition. The structural factors affecting disclosure that I discuss in this dissertation are not limited to affecting individuals—they affect both society at large and implicate societal institutions in their complicity. To be sure, this approach goes much further than simply identifying flaws in an interface that contribute to disclosure, and from that perspective seem overwhelming. However, a decade-plus of research examining the effects of user interfaces on personal disclosure and privacy have left many of us in the field with a similar conclusion: there is only so much human-computer interaction scholars and others can do to preserve privacy when the underlying social and legal structures are a source of the conflict. Acquisti, Brandimarte, and Loewenstein presented a similar conclusion in their interdisciplinary review in *Science*: “[i]f the goal of policy is to adequately protect privacy (as we believe it should be), then we need policies that protect individuals with minimal requirement of informed and rational decisionmaking—policies that include a baseline framework of protection.”[3] If we as a society want to earnestly work to preserve information privacy as a value, we need to move beyond blaming the public for making poor choices, acknowledge that the factors affecting disclosure to companies exist beyond the user interface, and critically examine the power differentials between individuals and companies.

### 7.6.1 Final Words

In 1995 in her book *Legislating Privacy*, Professor Priscilla Regan wrote:

“As one European commentator pointed out, an “enormous imbalance of power between the isolated individual and the great data collection organizations” exists, and “under these conditions, it is a pure illusion to speak of ‘control.’ Indeed, the fact of insisting exclusively on means of individual control can in fact be an alibi on the part of a public power wishing to avoid the new problems brought about by the development

of enormous personal data files, seeking refuge in an illusory exaltation of the powers of the individual, who will thus find himself alone to run a game in which he can only be the loser.” A definition of privacy as the right of the individual to control access to himself or herself, in effect, rests upon an “exaltation of the powers of the individual.” It also explains the failure to examine the interests of the organizations collecting and using personal information; instead, the individual is given the means to mediate his or her relationship with the organization. By placing the burden on the individual, there is less need to evaluate whether organizational interests are indeed social interests or whether individual privacy interests could be conceived as social interests.”[96]

Like Professor Regan, I support the perspective that privacy is not merely an individual need but a social need. She recommends that we redefine privacy, in part, as “the right of a society to require institutions using personal information to do so in a manner that respects the shared interests in that information.[96] I agree; in working to respect and preserve privacy as a common value we can all enjoy, we need to move towards solutions that reside less in the individual and more towards addressing the structural imbalances that can make personal disclosure less a matter of personal choice and more of a mandate.

# Appendix A

## Experimental Survey Scales

### Scales

This section provides detailed information about the scales I used in the experimental surveys.

#### Trust and Caution Scale

The ten-item trust and caution scale consists of five items related to trust, and five items related to caution (which are negatively correlated with the trust items). I created two separate scale variables by computing the mean per respondent for each item group. Cronbach's  $\alpha$  for the five trust items was .90.  $\alpha$  for the five caution items was .74. The composite trust and caution scores have a negative correlation of -.22. Both scales run from low agreement (1) to high agreement (7); a higher score on the trust scale means the respondent's answers indicated a more trusting attitude, while a higher score on the caution scale means that the respondent's answers indicated a more cautious outlook.

#### Information Technology Knowledge Scale

The ITKS I used was a revised version of the scale used by Cheshire et al (Cheshire et al. 2010b). It consisted of four questions, answered on a 7 point agree/disagree Likert scale: I fully understand most of the technology I use on a daily basis I am comfortable working with computers and computer systems I easily learn how to use new information technologies When I encounter a problem with computers I can solve the problem myself

$\alpha$  for these four items was .83. I created a scale variable for this measure by computing the mean per respondent of the four items.

#### General privacy scale

I created the privacy scale by first including thirteen items that originally appeared in other surveys. [73][112][70] The full list had an  $\alpha$  value of .84. I then performed a principal component

factor analysis (PCF) in STATA in order to develop a privacy scale that focused on one underlying concept. Following Adcock's instruction for performing principal component factor analysis[1], after performing PCF on the full thirteen items, I identified three relevant factors with eigenvalues of over 1.0. In reviewing the factor loadings for each variable, twelve items had loadings of at least .40 on Factor 1, so these twelve were included in the composite scale. Factors 2 and 3 had far fewer loadings at the .40 cutoff value. The twelve items had an alpha of .86. I created a general privacy scale variable by calculating the row mean of the twelve items.

### Real Product Measure

I asked three questions related to the use of wearable devices, and of these three, one had significant effects: "If the WearMe Device was a real product, how likely would you be to use it?" This question was included as an ecological validity check in order to identify whether and to what extent respondents found the vignette premise to be inapplicable to them individually.

**Likelihood of using the Wear Me Device if it were a real product**

Response	Frequency	Percent	Cumulative
Extremely Unlikely	51%	20.4	20.4%
Moderately unlikely	44%	18%	38%
Slightly unlikely	2%7	11%	49%
Neither likely nor unlikely	25%	10%	59%
Slightly likely	40%	16%	75%
Moderately Likely	45%	18%	93%
Extremely likely	18%	7%	100%
Total	250	100%	—

Table A.1: Responses to the real product question

I recoded the variable into a dummy variable ('real\_prod\_dum') with 0 = extremely unlikely to neither/nor, and 1 = slightly likely to extremely likely. This allowed me to control for the influence of this variable in the regression analyses.

**Real Product dummy variable**

Real Product dummy	Frequency	Percent
0 (Unlikely)	147	59
1 (likely)	103	41%
Total	250	100%

Table A.2: Real product responses summarized as a dummy variable.



## Appendix B

### Mean Comparisons for Studies 2A and 2B

#### Mean Comparisons for Study Two

**Study 2A: Mean comparisons for Study 2A**

Dependent Variable	Study 2A				
	Total Mean	Mandatory Disclosure	LAW	ANON	ISR
Trust in company	4.46 (1.48)	4.34 (1.65)	4.58 (1.40)	4.62 (1.39)	4.29 (1.49)
Trust with data	3.86 (1.74)	3.88 (1.88)	3.87 (1.67)	4.01 (1.75)	3.69 (1.67)
Fairness of Relationship	3.73 (1.78)	3.71 (1.98)	3.60 (1.70)	3.96 (1.74)	3.64 (1.70)
Benefit of relationship	4.12 (1.97)	4.27 (2.08)	4.05 (1.99)	3.96 (1.85)	4.18 (1.94)
Power in relationship	3.44 (2.08)	3.45 (2.20)	3.29 (2.00)	3.70 (2.18)	3.30 (1.93)
Control over terms	2.98 (2.14)	2.89 (2.23)	2.71 (1.81)	3.21 (2.33)	3.10 (2.14)
Disclosure - control over data	2.73 (1.92)	2.89 (2.15)	2.48 (1.77)	2.74 (1.86)	2.80 (1.86)
Privacy expectations	3.71 (1.78)	3.65 (1.85)	3.58 (1.68)	3.96 (1.81)	3.64 (1.77)
Privacy practices	5.18 (1.90)	5.09 (2.02)	5.13 (1.91)	5.30 (1.74)	5.20 (1.93)

Items in red are the highest means for each row; items in blue are the lowest. Values in parentheses are standard deviations.

Figure B.1: Mean comparisons for Study 2A

In the table above, I illustrate the highest (red) and lowest (blue) means for each dependent

variable for Study 2A. In three occasions the control condition actually had the highest mean, rather than the lowest, and only once did it have the lowest. The law condition had the lowest mean the most often, followed by indirect social reputation (ISR).

**Mean comparisons for Study 2B**

Dependent Variable	Study 2b				
	Total Mean	NY	LAW	ANON	ISR
Trust in company	4.86 (1.33)	4.81 (1.23)	5.00 (1.37)	4.71 (1.29)	4.92 (1.40)
Trust with data	4.51 (1.64)	4.6 (1.50)	4.36 (1.75)	4.51 (1.66)	4.57 (1.62)
Fairness of Relationship	4.68 (1.68)	4.83 (1.55)	4.60 (1.79)	4.53 (1.67)	4.77 (1.69)
Benefit of relationship	4.69 (2.11)	4.79 (2.00)	4.58 (2.29)	4.75 (2.12)	4.64 (2.05)
Power in relationship	4.01 (2.23)	3.91 (2.02)	4.00 (2.44)	4.17 (2.30)	3.97 (2.18)
Control over terms	3.81 (2.46)	3.63 (2.22)	3.98 (2.66)	3.77 (2.54)	3.85 (2.41)
Disclosure - control over data	4.61 (1.83)	4.88 (1.71)	4.30 (1.97)	4.51 (1.84)	4.76 (1.76)
Privacy expectations	4.54 (1.63)	4.43 (1.60)	4.59 (1.72)	4.54 (1.61)	4.61 (1.58)
Privacy practices	5.92 (1.77)	5.86 (1.68)	6.00 (1.89)	5.80 (1.91)	6.00 (1.61)

Items in red are the highest means for each row; items in blue are the lowest. Values in parentheses are standard deviations.

Figure B.2: Mean comparisons for Study 2B

Similar to the previous table, the table above illustrates the highest (red) and lowest (blue) means for each dependent variable for Study 2B. In four occasions the control condition actually had the highest mean (and it had the lowest mean most often), rather than the lowest, and in three instances it had the lowest. The law condition had the next lowest mean the most often.

# Appendix C

## Reference Tables for Chapter Seven

**Means with Standard Deviations, Median, and Scale for dependent variables**

<b>Dependent Variable</b>	<b>Total Mean</b>	<b>Optional disclosure</b>	<b>Mandatory disclosure</b>	<b>Median</b>	<b>Scale</b>
<b>Trust in company</b>	4.53 (1.38)	4.68 (1.35)	4.39 (1.39)	4	1-7
<b>Trust with data</b>	4.06 (1.75)	4.32 (1.66)	3.81 (1.80)	4	1-7
<b>Fairness of Relationship</b>	4.14 (1.72)	4.58 (1.64)	3.72 (1.70)	4	1-7
<b>Benefit of relationship</b>	4.64 (2.04)	4.84 (2.08)	4.44 (1.99)	4.3	1-10
<b>Power in relationship</b>	3.89 (2.18)	4.24 (2.26)	3.56 (2.06)	3.2	1-10
<b>Control over terms</b>	3.51 (2.33)	3.83 (2.49)	3.20 (2.13)	3.2	1-10
<b>Disclosure - control over data</b>	3.64 (2.02)	3.83 (2.49)	3.20 (2.13)	3	1-7
<b>Privacy expectations</b>	4.05 (1.77)	4.39 (1.73)	3.72 (1.76)	4	1-7
<b>Privacy practices</b>	5.44 (1.95)	5.89 (1.81)	5.01 (1.98)	5.6	1-10

Standard errors are in parentheses.

Figure C.1: Means with Standard Deviations, Median, and Scale for dependent variables

**Characterizations of the Company-respondent relationship**

<b>Based on what you have read, how would you describe <u>your relationship</u> (exchanging your data for access to the WearMe device) with the High-Tech Device Company?</b>	<b>Mean</b>	<b>Median</b>	<b>T-test of means between experimental conditions</b>	<b>Total N (excluding unsure)</b>
A. We are equal partners	3.2	3	$p < .05$	249
B. The relationship relies on someone or something else to make sure we both hold up our ends of the agreement	3.8	2	---	238
C. I have a positive feeling towards the relationship	3.9	3	$p < .05$	247
D. The relationship is exclusively between the two of us (one-to-one)	3.9	2	---	242
E. Only one of us contributes to the relationship	3.9	4	---	249
F. The relationship relies on an agreement that lays out what we can/cannot do	4.6	4	---	248
G. The relationship is based on trust	4.6	5	---	240
H. There is a clear expectation of what we each contribute	4.7	4	---	246
I. Either one of us is free to end the relationship at any time	4.7	4	$p < .05$	234

Total N=250. All responses were measured on a 1-7 Likert scale (strongly agree =7, strongly disagree = 1).

Figure C.2: Characterizations of the Company-respondent relationship

## **Appendix D**

# **Survey Instrument**

Note: Survey instrument begins on the following page.

## 1. Consent Process

### Debriefing/Consent Form for Use of Research Data

CPHS# 2015-12-8189

#### Introduction and Purpose

My name is Jennifer King. I am a graduate student at the University of California, Berkeley working with faculty advisor Coye Cheshire in the School of Information. I would like to invite you to take part in my research study, which explores the reasons why people choose to disclose information to companies. This form is designed to give you information about this study.

#### Procedures

If you agree to participate in our research, we will ask you to complete the following online survey. In addition to demographic questions, the survey will involve questions about a fictional company and its product, and should take approximately 20 minutes to complete.

#### Benefits

There is no direct benefit to you from taking part in this study. It is hoped that the research will help us understand better how people make decisions to disclose their personal information.

#### Risks/Discomforts

We will be asking you questions about a fictional company and product, as well as general questions about your experiences using the internet. We don't anticipate that these questions will give you reason for concern, but if you feel uncomfortable or upset at any time, you are free to stop participating.

As with all research, there is a chance that your confidentiality could be compromised; however, we are taking precautions to minimize this risk.

#### Confidentiality

Your study data will be handled as confidentially as possible. Since this is an online study, the confidentiality of your data will be kept to the degree permitted by the technology being used. No guarantees can be made regarding the interception of data sent via the Internet by any third parties.

To minimize the risks to confidentiality, we will not collect identifying information about you through Qualtrics. All study data will be encrypted and password-protected.

When the research is completed, we may save the study data for use in future research done by us or others. We will retain these records for up to three years after the study is over. The same measures described above will be taken to protect confidentiality of this study data.

#### Compensation

To thank you for participating in this study, you will receive a \$3.00 payment through Prolific Academic. Upon completion of the survey you will be provided with a unique code that you can use to collect payment through Prolific Academic's payment system.

#### Rights

**Participation in research is completely voluntary.** You are free to decline to take part in the project. You are free to stop taking part in the project at any time by closing your browser window. Most of the questions in this survey are required; however, when asked about your age, gender, educational levels, and income, you may select "prefer not to say" if you do not wish to answer these questions. Whether or

not you choose to participate, or continue participating in the project, there will be no penalty to you or loss of benefits to which you are otherwise entitled.

**Questions**

If you have any questions about this research, please feel free to contact the research team at [jenking@berkeley.edu](mailto:jenking@berkeley.edu).

If you have any questions about your rights or treatment as a research participant in this study, please contact the University of California at Berkeley's Committee for Protection of Human Subjects at 510-642-7461, or e-mail [subjects@berkeley.edu](mailto:subjects@berkeley.edu).

**Consent** If you consent to the use of your data, please select the "I consent" button. If you do not consent to the use of your survey data, please select the "I do not consent" button. Please click [here](#) to download a copy of this consent document.

- I consent, begin the study (1)
- I do not consent, I do not wish to participate (0)

**2. Vignette Text****Q3.2 The WearMe Wearable Tracker -- created by The High-Tech Device Company (HTDC)**

The High-Tech Device Company (HTDC) is a world leader in consumer technology and wearable devices, building products powered by research and passion.

Our wearable tracker, the "WearMe", helps people live better by providing personalized insights into how they sleep, move, and feel.

Similar in size and shape to a wristwatch, the WearMe features five colors in three sporty and fashionable styles.

Our approach to lifestyle tracking is unique, relying on multiple data points to customize the WearMe's recommendations.

The WearMe tracks your:

- body data
- heart rate
- number of steps you walk or run
- body temperature
- sleep cycles
- location data
- precise physical location based on where you go throughout the day.

The WearMe then provides you with personalized recommendations for improving your health and fitness based on the tracking data collected from your WearMe Device and additional information about you that provide in your WearMe profile.

Our patented approach provides you with a health and fitness program tailored to your goals and needs.

You can access your personalized recommendations using the WearMe mobile app or at the WearMe website.

This requires activating an account with your personal data:

- name
- age
- gender
- email address home address
- mobile phone number.

Accounts are subject to our Terms of Use.

We collect and store your data in accordance with standard industry practices.



**3. Experimental Conditions and Assurance Statements**

**Mandatory Disclosure**

**WearMe Price: \$99.99**

**You must allow the WearMe to track all the data it requests about your body and your physical location.**

**Optional Disclosure**

**WearMe Price: \$99.99**

**You can choose the types of data about your body and your physical location the WearMe device will track.**

**Assurance Statements**

**Over a million people use the WearMe to improve their health and fitness.**

**The data the WearMe collects from your body and about your location is protected by law.**

**Your data is anonymized when used for any purpose other than to provide you with WearMe recommendations.**

#### 4. Survey Questions

Q4.3

As you answer the following questions imagine you have purchased and are using the WearMe device.

**Feel free to refer back to the WearMe Terms of Use and the Company Overview (in the split screen above) as you need when answering.**

**Q4.4 Please indicate your level of agreement with this statement:**

I am comfortable with disclosing my personal data (name, gender, age, email address, home address, mobile phone number) to the High-Tech Device Company in order to use the WearMe device.

Strongly agree (7)

Agree (6)

Somewhat agree (5)

Neither agree nor disagree (4)

Somewhat disagree (3)

Disagree (2)

Strongly disagree (1)

**Q4.5 Please indicate your level of agreement with this statement:**

I am comfortable with disclosing my location data (my precise physical location of where I go throughout the day) to the High-Tech Device Company in order to use the WearMe device.

Strongly agree (7)

Agree (6)

Somewhat agree (5)

Neither agree nor disagree (4)

Somewhat disagree (3)

Disagree (2)

Strongly disagree (1)

**Q5.3 Please indicate your level of agreement with this statement:**

I am comfortable with disclosing my body data (heart rate, the number of steps you walk or run, your temperature, your sleep cycles) to the High-Tech Device Company in order to use the WearMe device.

Strongly agree (7)

Agree (6)

Somewhat agree (5)

Neither agree nor disagree (4)

Somewhat disagree (3)

Disagree (2)

Strongly disagree (1)

**Q5.4 Please indicate your level of agreement with this statement:**

Based on what I have read, I find the High-Tech Device Company to be trustworthy.

Strongly agree (7)

Agree (6)

Somewhat agree (5)

Neither agree nor disagree (4)

Somewhat disagree (3)

Disagree (2)

Strongly disagree (1)

**Q6.3 Please indicate your level of agreement with this statement:**

Based on the information given in the scenario above, I am confident I would be able to control which data I disclose to the High-Tech Device Company.

*(Data includes your body data, personal data, and location data.)*

Strongly agree (7)

Agree (6)

Somewhat agree (5)

Neither agree nor disagree (4)

Somewhat disagree (3)

Disagree (2)

Strongly disagree (1)

**Q6.4 Please indicate your level of agreement with this statement:**

The relationship (exchanging my data for access to the WearMe device) I have with the High-Tech Device Company is fair to me.

Strongly agree (7)

Agree (6)

Somewhat agree (5)

Neither agree nor disagree (4)

Somewhat disagree (3)

Disagree (2)

Strongly disagree (1)

**Q7.3 Please indicate your level of agreement with each statement:**

Based on what you have read, how would you describe your relationship (exchanging your data for access to the WearMe device) with the High-Tech Device Company?

**Note: all questions were randomized**

**Scale:** Strongly agree (7); Agree (6); Somewhat agree (5); Neither agree nor disagree (4); Somewhat disagree (3); Disagree (2); Strongly disagree (1); Unsure (0)

We are equal partners (Q2.14\_1)

The relationship is based on trust (Q2.14\_2)

Only one of us contributes to the relationship (Q2.14\_3)

There is a clear expectation of what we each contribute (Q2.14\_4)

The relationship relies on an agreement that lays out what we can/cannot do (Q2.14\_5)

The relationship relies on someone or something else to make sure we both hold up our ends of the agreement (Q2.14\_6)

The relationship is exclusively between the two of us (one-to-one) (Q2.14\_7)

Either one of us is free to end the relationship at any time (Q2.14\_8)

I have a positive feeling towards the relationship (Q2.14\_9)

**Q8.3 Please indicate how sensitive you consider the types of body data and location data the **WearMe device** collects.**

*Note: by 'sensitive,' we mean information you would be concerned with others knowing or sharing about you without your permission.*

Scale: Extremely sensitive (5); Very Sensitive (4); Somewhat sensitive (3); Slightly sensitive (2); Not at all sensitive (1); Unsure (0)

- My heart rate (Q2.26\_1 HR)
- The number of steps I walk or run (Q2.26\_2 STEPS)
- My body temperature (Q2.26\_3 TEMP)
- My sleep cycles (Q2.26\_4 SLEEP)
- My precise physical location over time (Q2.26\_5 LOC)

**Q8.4 Please indicate how sensitive you consider the types of personal data the **High-Tech Device Company** collects.**


*Note: by 'sensitive,' we mean information you would be concerned with others knowing or sharing about you without your permission.*

Scale: Extremely sensitive (5); Very Sensitive (4); Somewhat sensitive (3); Slightly sensitive (2); Not at all sensitive (1); Unsure (0)

- My email address (Q2.27\_1 EMAIL)
- My real name (Q2.27\_2 NAME)
- My gender (Q2.27\_3 GEN)
- My age (Q2.27\_4 AGE)
- My home address (Q2.27\_5 ADD)
- My mobile phone number (Q2.27\_6 PHONE)

**Q9.3 In your opinion, who has more power in this relationship (exchanging my data for use of the **WearMe device**) -- you, or the **High Tech Device Company**?**


Scale: 1-10

Please indicate who you think has <u>more power</u> by moving the slider. (1)	
---	--

**Q9.4 In your opinion, who has more control over the terms of use in this relationship (exchanging my data for use of the **WearMe device**) -- you, or the **High Tech Device Company**?**

*(Data includes your body data, personal data, and location data.)*

Scale: 1-10


Please indicate who you think has <u>more control over the terms</u> by moving the slider. (1)	
--	--

Q10.3 **Please indicate your level of agreement with this statement:** Based on the information given in the scenario above, I trust the High-Tech Device Company with my data.  
 (Data includes your body data, personal data, and location data.)

- Strongly agree (7)
- Agree (6)
- Somewhat agree (5)
- Neither agree nor disagree (4)
- Somewhat disagree (3)
- Disagree (2)
- Strongly disagree (1)

10.4 In your opinion, who benefits more from this relationship: you or the High-Tech Device Company?

Scale: 1-10

Please indicate who you think <u>benefits more from this relationship</u> by moving the slider. (1)	
---	--

Q11.3 Please indicate how much each the following factors would provide you with assurance (i.e., a sense of trust or certainty) about using the WearMe Device:

**Note:** all questions were randomized

**Scale:** Strongly assured (1); Assured (2); Somewhat assured (3); Neither assured nor unassured (4); Somewhat unassured (5); Unassured (6); Strongly Unassured (7); Unsure (0)

- If all the data the WearMe Device collects were protected by law (Q2.28\_1 LAW)
- If the device manufacturer (the High-Tech Device Company) had a privacy policy (Q2.28\_2 PP)
- If the device manufacturer (the High-Tech Device Company) had a good reputation among its customers (Q2.28\_3 REP)
- If the WearMe Device was used by over a million people (Q2.28\_4 MILLION)
- If I had the option of deciding which data types the device will track (Q2.28\_5 OPTION)
- If the company anonymized my data (i.e., my body and location data were not linked to my name or other personal information) (Q2.28\_7 ANON)

**Q11.4 Please indicate your level of agreement with this statement:**


Based on the information given in the scenario above, the way in which the HTDC collects and uses my data meets my privacy expectations.

*(Data includes your body data, personal data, and location data.)*

- Strongly agree (7)
- Agree (6)
- Somewhat agree (5)
- Neither agree nor disagree (4)
- Somewhat disagree (3)
- Disagree (2)
- Strongly disagree (1)

**Q12.3 Based on the scenario and your own personal experiences, how would you rate the High Tech Device Company's privacy practices?**

Scale: 1=10

Please indicate your rating by moving the slider (13)	
--	--

**Q12.4 - STUDY ONE ONLY**

**For this question, assume that the High Tech Device Company anonymizes all of the data it collects from you (meaning, your body data and location data are not linked to your name or other personal information).**

**How much risk (the possibility of something bad or harmful happening to you) would each of the following scenarios pose to you?**

**Scale:** Not at all risky (1); Slightly risky (2); Somewhat risky (3); Very risky (4); Extremely risky (5)

All of my WearMe data was sold by the HTDC to another company. (Q12.4\_sold)

All of my WearMe data was shared by the HTDC to another company for advertising purposes. (Q12.4\_shared)

All of my WearMe data was stolen by hackers. (Q12.4\_steal)

## Q12.5 - STUDY ONE ONLY

For this question, assume that the High Tech Device Company does not anonymize all of the data it collects from you (meaning, your body data and location data are linked to your name or other personal information). How much risk (the possibility of something bad or harmful happening to you) would each of the following scenarios pose to you?

**Scale:** Not at all risky (1); Slightly risky (2); Somewhat risky (3); Very risky (4); Extremely risky (5)

All of my WearMe data was sold by the HTDC to another company. (Q113\_sold)

All of my WearMe data was shared by the HTDC to another company for advertising purposes. (Q113\_share)

All of my WearMe data was stolen by hackers. (Q113\_steal)

## Q12.6 - STUDIES 2A and 2B ONLY

The WearMe device scenario states that "We collect and store your data in accordance with standard industry practices." What do you believe are included in these "standard industry practices?"

Please indicate how likely you think it is that each practice listed below is included.

**Note:** all questions were randomized

**Scale:** Extremely unlikely (1); Moderately unlikely (2); Slightly unlikely (3); Neither likely nor unlikely (4); Slightly likely (5); Moderately likely (6); Extremely likely (7)

The HTDC has a privacy policy. (Q12.6\_PP)

The HTDC does not share or sell any of its customer data with other companies. (Q12.6\_NO\_SHARE)

The HTDC shares anonymized data with other companies that doesn't identify its customers. (Q12.6\_ADS)

The HTDC uses basic security measures to protect its customer data. (Q12.6\_SEC)

The HTDC protects its customer data by using strong encryption. (Q12.6\_ENC)

The HTDC will share or sell your personal data only with your permission. (Q12.6\_PERM)

The HTDC will share or sell your data without your permission. (Q12.6\_NOPER)

## Q128 – STUDIES 2A and 2B ONLY

**OPTIONAL:** Do you believe there was a 'standard industry practice' that wasn't included above? If so, please describe here:

---



**Q13.3 Imagine the WearMe device and service were offered free to you if you allowed your health insurance company access to all of the data the device collects.**

How likely would you be to agree to this condition?

Extremely likely (7)

Moderately likely (6)

Slightly likely (5)

Neither likely nor unlikely (4)

Slightly unlikely (3)

Moderately unlikely (2)

Extremely unlikely (1)

**Q13.4 If the WearMe Device was a real product, how likely would you be to use it?**

Extremely likely (7)

Moderately Likely (6)

Slightly likely (5)

Neither likely nor unlikely (4)

Slightly unlikely (3)

Moderately unlikely (2)

Extremely unlikely (1)

**Q14.3 If the cost were not a concern, how likely would you be to use any kind of self-tracking wearable device?**

Extremely likely (7)

Moderately likely (6)

Slightly likely (5)

Neither likely nor unlikely (4)

Slightly unlikely (3)

Moderately unlikely (2)

Extremely unlikely (1)

**Q14.4 Have you ever used or are you currently using any kind of wearable device that tracks your physical or emotional activity?**

Yes (1)

No (2)

Uncertain (0)

*Display This Question:*

*If Q14.4 = 1*

**Q14.5 What is the name of device you use/used? (Your best guess is fine.)**

---

**Q15.2 Please answer the remainder of the questions based on your own personal experiences.**

*Note: transition from vignette-specific questions to general questions.*

**Q15.3 To what extent do you agree or disagree with the following statements?**

**Note: all questions were randomized**

**Scale:** Strongly agree (7); Agree (6); Somewhat agree (5); Neither agree nor disagree (4); Somewhat disagree (3); Disagree (2); Strongly disagree (1)

Most people are basically honest (Q18.1\_HONEST)

One can avoid falling into trouble by assuming that all people have a vicious streak (Q18.1\_VICIOUS)

If anything, I trust others (Q18.1\_TRUST)

Most people are basically good-natured and kind (Q18.1\_KIND)

You cannot be too cautious in dealing with others (Q18.1\_CAUTIOUS)

Most people trust others (Q18.1\_TRUSTOTH)

We do not always have to guard ourselves against being used by someone (Q18.1\_GUARD)

Most people are trustworthy (Q18.1\_TRUSTWORTHY)

If you are not careful enough, people will take advantage of you (Q18.1\_CAREFUL)

It is safer to believe that everyone has the capacity to be malicious (Q18.1\_MALICIOUS)

**Q15.4 To what extent do you agree or disagree with the following statements?**

**Note: all questions were randomized**

**Scale:** Strongly agree (7); Agree (6); Somewhat agree (5); Neither agree nor disagree (4); Somewhat disagree (3); Disagree (2); Strongly disagree (1)

I fully understand most of the technology I use on a daily basis (Q18.2\_UNDERST)

I am comfortable working with computers and computer systems (Q18.2\_COMFORT)

I easily learn how to use new information technologies (Q18.2\_LEARN)

When I encounter a problem with computers I can solve the problem myself (Q18.2\_SOLVE)

**Q16.2 When you hear the word “privacy,” what comes to mind for you? Tell us the first few words that pop into your head.**

---

**Q16.3 In general, how important is it to you that companies respect your privacy?**

Very important (5)

Somewhat important (4)

Neither important nor unimportant (3)

Somewhat unimportant (2)

Not important at all (1)

Unsure (0)

**Q16.4 Have you personally had an experience on the internet that you would consider to be an invasion of your privacy, or has this never happened to you?**

Yes, has happened (1)

No, never happened (2)

Unsure (0)

**Q17.2 For each of the following statements please indicate the degree to which you agree or disagree.**

**Note: all questions were randomized**

**Scale:** Strongly agree (7); Agree (6); Somewhat agree (5); Neither agree nor disagree (4); Somewhat disagree (3); Disagree (2); Strongly disagree (1)

I am willing to share some information about myself with companies in order to use online services for free. (Q3.5\_2 FREE)

Consumers have lost control over how personal information is collected and used by companies. (Q3.5\_3 NOCONT)

I've come to accept that I have little control over what advertisers can learn about me. (Q3.5\_4 RESG1)

I appreciate that online services are more efficient because of the increased access they have to my personal data. (Q3.5\_5 EFFI)

I want to have control over what advertisers can learn about me online. (Q3.5\_11 RESG2)

In general, I trust online companies with my personal information. (Q16.5\_TRUST)

In general, I think online companies act in my best interest. (Q16.5\_BESTINT)

Online companies are in general predictable and consistent regarding the usage of my personal information. (Q16.5\_PREDICT)

It is very important to me that I am aware and knowledgeable about how my personal information will be used. (Q16.5\_KNOW)

Online companies should never sell the personal information they collect about their customers to other companies. (Q16.5\_SELL)

Online companies should never share personal information with other companies unless it has been authorized by the individuals who provided it. (Q16.5\_SHARE)

I am unconcerned about threats to my personal privacy today. (Q16.5\_THREAT)

I think there is little risk to sharing my personal information with online companies (Q16.5\_RISK)

STUDIES 2A and 2B ONLY:

I feel as if I have nothing to hide from online companies. (Q17.2\_NOTH2HIDE)

Most online services have so many customers I don't feel as if me or my personal information would ever be singled out. (Q17.2\_OBSCURE)

**Q18.2 Which of the factors below would make you more or less likely to use ANY online service that collects your personal information?**

**Note: all questions were randomized**

**Scale:** Extremely likely (5); Somewhat likely (4); Neither likely nor unlikely (3); Somewhat unlikely (2); Extremely unlikely (1); Unsure (0)

The company has a privacy policy (Q3.10\_1 PP)

The company has a terms of service agreement (Q3.10\_2 TOS)

A security seal or other certification of the company's security measures (Q3.10\_3 SECS)

The company is a verified merchant by my credit card company (Q3.10\_4 VERF)

A law specifically protects the type of information the company collects (Q3.10\_5 LAW)

The company states on their website that: "Your privacy is important to us." (Q3.10\_6 PRIV\_IMPT)

Positive ratings of the company's app in an app store (Q3.10\_7 RATING)

Positive reviews or articles about the company's device or service (Q3.10\_8 REVIEW)

A seal on a company's website from a consumer group (e.g. the Better Business Bureau) certifying their business (Q3.10\_9 SEAL)

The company anonymizes all the data it collects (Q16.10\_10 ANON)

My friends and/or family use the service and recommend it (Q16.10\_11 FF)

**Q18.3 If a company acted in an unethical manner that made you angry with them – such as selling your personal data after promising not to do so – how might you react? Select all that apply:**

I would contact the company directly (1)

I would quit using the service or product (2)

I would complain about the company on social media (3)

I would rate the company poorly on consumer review websites like Yelp or the Better Business Bureau (4)

I would consider suing the company (5)

I would consider joining a class action lawsuit against the company (6)

I would complain to my state authorities (such as my state attorney general or bureau of consumer affairs) (7)

I would complain to federal government authorities (such as the Federal Trade Commission) (8)

I would do something else not listed here (please describe): (9)

---

⊗ I wouldn't do any of these things (10)

⊗ I would do nothing at all (0)

**Q19.2 What gender do you primarily identify with? Please select a single answer:**

Female (1)

Male (2)

Transgender (3)

Genderqueer (4)

I do not identify with any of these categories (5)

Prefer not to say (0)

**Q19.3 What is your current age?**

Under 18 (12)

18 - 24 (13)

25 - 34 (14)

35 - 44 (15)

45 - 54 (16)

55 - 64 (17)

65 - 74 (18)

75 - 84 (19)

85 or older (20)

Prefer not to say (21)

**Q19.4 What ethnicity do you primarily identify with? Please select a single answer:**

Asian or Pacific Islander (1)

African-American (2)

Hispanic or Latino (3)

Multi-racial (4)

White or Caucasian (5)

Prefer not to say (0)

**Q19.5 What is the highest level of education you have completed?**

- Some high school (1)
- High school graduate (2)
- Some college (3)
- Vocational degree (4)
- 2 year college degree (AA) (5)
- 4 year college degree (B.A., B.S.) (6)
- Master's/Professional degree (7)
- Ph.D or M.D. (8)
- Prefer not to say (0)

**Q19.6 What is your annual income?**

- Below \$25K (1)
- Between \$25K-\$50K (2)
- Between \$50K-\$75K (3)
- Between \$75K-\$100K (4)
- Between \$100K-\$150K (5)
- Over \$150K (6)
- Prefer not to say (0)

**Q19.7 Please let us know if you have any additional comments, feedback, or questions.  
Thank you for your time!**

Q20.1 Please enter your Prolific Academic ID:

*Note: Your Prolific Academic ID is required to verify your submission and process your payment.*

---

Q20.2 Thank you for taking part in our survey. Please click the Finish & Submit button below.

You will be automatically redirected to Prolific Academic after you click Submit.

## Bibliography

- [1] Alan C. Acock. *A Gentle Introduction to Stata*. 2nd Editio. College Station, TX: Stata Press, 2008, p. 357. ISBN: 9781597180436. URL: <http://books.google.co.id/books?id=YZIQ00acuWwC>.
- [2] Acquisti A. and J Grossklags. “Privacy and Rationality in Individual Decision Making”. In: *IEEE Security and Privacy Magazine* 3.1 (2005), pp. 26–33.
- [3] A Acquisti, L Brandimarte, and G Loewenstein. “Privacy and human behavior in the age of information”. In: *Science* 347.6221 (2015), pp. 509–514.
- [4] Alessandro Acquisti and Jens Grossklags. “What can behavioral economics teach us about privacy?” In: *Digital Privacy: Theory, Technologies, and Practices*. Ed. by A. Acquisti et al. Auerbach Publications, 2007, pp. 363–379.
- [5] Alessandro Acquisti, Leslie K John, and George Loewenstein. “What Is Privacy Worth”. In: 42.June (2013), pp. 249–274.
- [6] Alessandro Acquisti, Curtis R Taylor, and Liad Wagman. “The Economics of Privacy”. In: *SSRN Electronic Journal* (2015). URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C\\_%7Did=2580411](https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Did=2580411).
- [7] Idris Adjerid, Eyal Peer, and Alessandro Acquisti. *Beyond the Privacy Paradox: Objective versus Relative Risk in Privacy Decision Making*. 2016. URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C\\_%7Did=2765097](https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Did=2765097).
- [8] Idris Adjerid et al. “Sleights of privacy”. In: *Proceedings of the Ninth Symposium on Usable Privacy and Security - {SOUPS} ’13*. 2013.
- [9] Julia Angwin. *Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance*. Times Books, Henry Holt and Company, 2014. ISBN: 9780805098082.
- [10] Kenneth A Bamberger and Deirdre K Mulligan. *Privacy on the Ground: Driving Corporate Behavior in the United States and Europe*. MIT Press, Oct. 2015. ISBN: 9780262331357.
- [11] Michael Barbaro and Tom Zeller Jr. “A Face Is Exposed for AOL Searcher No. 4417749”. In: *The New York times* (Aug. 2006). ISSN: 0362-4331. URL: <https://www.nytimes.com/2006/08/09/technology/09aol.html>.
- [12] Susan B Barnes. “A privacy paradox: Social networking in the United States”. In: *First Monday* 11.9 (2006).

- [13] Solon Barocas and Helen Nissenbaum. “Big Data’s End Run around Anonymity and Consent”. In: *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. Ed. by Julia Lane et al. Cambridge, MA: Cambridge University Press, 2014, pp. 44–75.
- [14] A Barth et al. “Privacy and contextual integrity: framework and applications”. In: *2006 {IEEE} Symposium on Security and Privacy ({S&P’06})*. 2006.
- [15] Jerry Beilinson. *Glow Pregnancy App Exposed Women to Privacy Threats, Consumer Reports Finds*. July 2016. URL: <http://www.consumerreports.org/mobile-security-software/glow-pregnancy-app-exposed-women-to-privacy-threats/>.
- [16] Sebastian Benthall, Seda Gürses, and Helen Nissenbaum. “Contextual Integrity through the Lens of Computer Science”. In: *Foundations and Trends® in Privacy and Security 2.1* (Dec. 2017), pp. 1–69. ISSN: 2474-1558. DOI: 10.1561/33000000016. URL: <http://www.nowpublishers.com/article/Details/SEC-016>.
- [17] Peter Blau. *Exchange and Power in Social Life*. New York: Wiley, 1964.
- [18] Nellie Bowles. *Early Facebook and Google Employees Form Coalition to Fight What They Built*. San Francisco, Feb. 2018. URL: <https://www.nytimes.com/2018/02/04/technology/early-facebook-google-employees-fight-tech.html>.
- [19] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. “Misplaced Confidences”. In: *Soc. Psychol. Personal. Sci.* 4.3 (May 2012), pp. 340–347. DOI: 10.1177/1948550612455931. URL: <http://spp.sagepub.com/content/4/3/340.abstract>.
- [20] Alex Braunstein, Laura Granka, and Jessica Staddon. “Indirect content privacy surveys”. In: *Proceedings of the Seventh Symposium on Usable Privacy and Security - SOUPS ’11*. 2011. DOI: 10.1145/2078827.2078847.
- [21] M. Ryan Calo. “Against Notice Skepticism in Privacy (and Elsewhere)”. In: *Notre Dame Law Review* (2012), pp. 1–51. ISSN: 07453515. DOI: 10.1525/sp.2007.54.1.23.. arXiv: arXiv:1011.1669v3. URL: [http://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C\\_%7Ddid=1790144](http://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Ddid=1790144).
- [22] Coye Cheshire. “Selective incentives and generalized information exchange”. In: *Social Psychology Quarterly* 70.1 (2007), pp. 82–100. DOI: 10.1177/019027250707000109.
- [23] Coye Cheshire, Judd Antin, and Elizabeth Churchill. “Behaviors, adverse events, and dispositions: An empirical study of online discretion and information control”. In: *Journal of the American Society for Information Science and Technology* 61.7 (2010), pp. 1487–1501. DOI: 10.1002/asi.21346. URL: <http://dx.doi.org/10.1002/asi.21346>.
- [24] Coye Cheshire, Alexandra Gerbasi, and Karen S. Cook. “Trust and Transitions in Modes of Exchange”. In: *Social Psychology Quarterly* 73.2 (2010), pp. 176–195.



- [25] Federal Trade Commission. *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*. Tech. rep. Federal Trade Commission, 2012, p. 112.
- [26] K. S. Cook, C. Cheshire, and A. Gerbasi. “Power, Dependence, And Social Exchange”. In: *Contemporary Social Psychological Theories*. 2006, pp. 194–216. ISBN: 0804753474.
- [27] Karen S Cook and Richard M Emerson. “Power, Equity and Commitment in Exchange Networks”. In: *Am. Sociol. Rev.* 43.5 (1978), p. 721.
- [28] Karen S. Cook and Eric Rice. “Social Exchange Theory”. In: *Handbook of Social Psychology*. Ed. by John Delamater. New York: Kluwer Academic/Plenum Publishers, 2003. Chap. 3, pp. 53–76.
- [29] Karen S Cook et al. “Social Exchange Theory”. In: *Handbooks of Sociology and Social Research*. 2013, pp. 61–88.
- [30] Culnan Mary J. and Pamela K Armstrong. “Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation”. In: *Organization Science* 10.1 (1999), pp. 104–115.
- [31] Michelle De Mooy. *Rethinking Privacy Self-Management and Data Sovereignty in the Age of Big Data Considerations for Future Policy Regimes in the United States and the European Union*. Tech. rep. Center for Democracy and Technology, 2017. URL: <https://cdt.org/insight/rethinking-privacy-self-management-and-data-sovereignty-in-the-age-of-big-data/>.
- [32] Dienlin Tobias and Sabine Trepte. “Is the Privacy Paradox a Relic of the Past? An in-Depth Analysis of Privacy Attitudes and Privacy Behaviors”. In: *European Journal of Social Psychology* 45.3 (2014), pp. 285–297.
- [33] Tamara Dinev and Paul Hart. “Internet privacy concerns and their antecedents - measurement validity and a regression model”. In: *Behav. Inf. Technol.* 23.6 (2004), pp. 413–422.
- [34] Paul Dourish. “What we talk about when we talk about context”. In: *Pers. Ubiquit. Comput.* 8.1 (2004), pp. 19–30.
- [35] Charles Duhigg. “How Companies Learn Your Secrets”. In: *The New York times* (Feb. 2012). URL: <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.
- [36] Serge Egelman et al. “Timing is Everything?: The Effects of Timing and Placement of Online Privacy Indicators”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: ACM, 2009, pp. 319–328. DOI: 10.1145/1518701.1518752.
- [37] Jessica Vitak Charles Steinfield Rebecca Gray Ellison Nicole B. and Cliff Lampe. “Negotiating Privacy Concerns and Social Capital Needs in a Social Media Environment”. In: *Privacy Online* (2011), pp. 19–32.

- [38] Richard Emerson. "Exchange theory, part II: Exchange relations and networks". In: *Sociological Theories in Progress*. Ed. by Joseph Berger, Morris Zelditch Jr., and Bo Anderson. Boston: Houghton Mifflin Harcourt, 1972, pp. 58–87.
- [39] Richard M. Emerson. "Exchange theory, part I: A psychological basis for social exchange." In: *Sociological Theories in Progress*. Ed. by Joseph Berger, Morris Zelditch Jr., and Bo Anderson. Vol. 2. Boston: Houghton Mifflin Harcourt, pp. 38–57.
- [40] Richard M Emerson. "Power-Dependence Relations". In: *American Sociological Review* 27.1 (1962), pp. 31–41. ISSN: 00031224. DOI: 10.2307/2089716. URL: <http://www.jstor.org/stable/2089716>.
- [41] Richard M Emerson. "Social Exchange Theory". In: *Annual Review of Sociology* 2 (1976), pp. 335–362. ISSN: 03600572, 15452115. URL: <http://www.jstor.org/stable/2946096>.
- [42] Mary Flanagan and Helen Nissenbaum. *Values at Play in Digital Games*. MIT Press, July 2014.
- [43] Luciano Floridi. "The Ontological Interpretation of Informational Privacy". In: *Ethics and Information Technology* 7.4 (2005), pp. 185–200.
- [44] Edna B Foa and Uriel G Foa. "Resource Theory". In: *Social Exchange: Advances in Theory and Research*. 1980, pp. 77–94. ISBN: 978-1-4613-3087-5. DOI: 10.1007/978-1-4613-3087-5\_4. URL: [https://doi.org/10.1007/978-1-4613-3087-5%7B%5C\\_%7D4](https://doi.org/10.1007/978-1-4613-3087-5%7B%5C_%7D4).
- [45] U. G. Foa and E. B Foa. *Societal Structures of the Mind*. Springfield, IL: Charles Thomas, 1974.
- [46] Michael Geist. *Supreme Court Rules Facebook Can't Contract Out of B.C. Privacy Law - Michael Geist*. 2017. URL: <http://www.michaelgeist.ca/2017/06/supreme-court-rules-facebook-cant-contract-b-c-privacy-law/> (visited on 05/07/2018).
- [47] David Gelles. *Tech Backlash Grows as Investors Press Apple to Act on Children's Use*. New York, Jan. 2018.
- [48] Bob Gellman. *FAIR INFORMATION PRACTICES: A Basic History*. 2017. URL: <https://bobgellman.com/rg-docs/rg-FIPshistory.pdf>.
- [49] General Data Protection Regulation (GDPR). *Art. 20 GDPR – Right to data portability*. URL: <https://gdpr-info.eu/art-20-gdpr/> (visited on 05/07/2018).
- [50] Ralph Gross and Alessandro Acquisti. "Information revelation and privacy in online social networks". In: *Privacy in the Electronic Society 2005* (2005), p. 11. ISSN: 15206106. DOI: 10.1145/1102199.1102214. URL: <http://dl.acm.org/citation.cfm?id=1102214>.
- [51] Russell Hardin. *Trust and Trustworthiness*. Russell Sage Foundation, Mar. 2002. ISBN: 9781610442718.

- [52] Matthew Herper. *Surprise! With \$60 Million Genentech Deal, 23andMe Has A Business Plan*. Jan. 2015. URL: <https://www.forbes.com/sites/matthewherper/2015/01/06/surprise-with-60-million-genentech-deal-23andme-has-a-business-plan/>.
- [53] George Caspar Homans. *Social Behaviour: Its Elementary Forms*. Taylor & Francis, 1961.
- [54] Chris Hoofnagle. *Exit, Voice, and the Privacy Paradox*. 2014. URL: <https://medium.com/@hoofnagle/exit-voice-and-the-privacy-paradox-662a922ff7c8>.
- [55] Eric Horvitz and Deirdre Mulligan. "Policy forum. Data, privacy, and the greater good". In: *Science* 349.6245 (July 2015), pp. 253–255. ISSN: 0036-8075.
- [56] Tiffany Hsu. *For Many Facebook Users, a 'Last Straw' That Led Them to Quit*. Mar. 2018. URL: <https://www.nytimes.com/2018/03/21/technology/users-abandon-facebook.html>.
- [57] Carlos Jensen and Colin Potts. "Privacy Policies As Decision-making Tools: An Evaluation of Online Privacy Notices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '04. New York, NY, USA: ACM, 2004, pp. 471–478. DOI: 10.1145/985692.985752.
- [58] Daniel Kahneman. *Thinking , Fast and Slow*. New York: Farrar, Straus and Giroux, 2011, p. 499.
- [59] *How Come I'm Allowing Strangers To Go Through My Phone? Smartphones and Privacy Expectations*. 2012. URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C\\_%7Ddid=2493412](https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Ddid=2493412).
- [60] *Taken Out of Context: An Empirical Analysis of Westin's Privacy Scale*. July 2014. URL: <http://cups.cs.cmu.edu/soups/2014/workshops/privacy.html>.
- [61] Brendan I Koerner. *Your Relative's DNA Could Turn You Into a Suspect*. Oct. 2015. URL: <https://www.wired.com/2015/10/familial-dna-evidence-turns-innocent-people-into-crime-suspects/>.
- [62] Natasha F.; Krasnova Hanna; Veltri and Oliver Günther. "Self-disclosure and Privacy Calculus on Social Networking Sites: The Role of Culture - Intercultural Dynamics of Privacy Calculus". In: *Business and Information Systems Engineering* 4.3 (2012), pp. 127–135.
- [63] Airi Lampinen and Coye Cheshire. "Hosting via Airbnb". In: *Proceedings of the 2016 {CHI} Conference on Human Factors in Computing Systems - {CHI} '16*. 2016.
- [64] Sang M. Lee, Jeongil Choi, and Sang-Gun Lee. "The Impact of a Third-Party Assurance Seal in Customer Purchasing Intention". In: *Journal of Internet Commerce* 3.2 (2004), pp. 33–51. URL: [http://www.tandfonline.com/doi/abs/10.1300/J179v03n02%7B%5C\\_%7D03](http://www.tandfonline.com/doi/abs/10.1300/J179v03n02%7B%5C_%7D03).
- [65] H. Li, R. Sarathy, and H. Xu. "Understanding Situational Online Information Disclosure as a Privacy Calculus". In: *Journal of Computer Information Systems* 51.1 (2010), pp. 62–71.

- [66] Xueming Luo. "Trust production and privacy concerns on the Internet". In: *Industrial Marketing Management* 31.2 (2002), pp. 111–118.
- [67] Mary Madden. *Americans Consider Certain Kinds of Data to Be More Sensitive than Others*. Tech. rep. <http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/>: Pew Research Center: Internet, Science & Tech., 2014.
- [68] Mary Madden and Lee Rainie. *Americans' Attitudes About Privacy, Security and Surveillance*. Tech. rep. <http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/>: Pew Research Center: Internet, Science & Tech., 2015.
- [69] Michael Mainelli. "Blockchain Could Help Us Reclaim Control of Our Personal Data". In: *Harvard Business Review* (Oct. 2017). URL: <https://hbr.org/2017/10/smart-ledgers-can-help-us-reclaim-control-of-our-personal-data>.
- [70] Naresh K Malhotra, Sung S Kim, and James Agarwal. "Internet Users' Information Privacy Concerns ({UIIPC}): The Construct, the Scale, and a Causal Model". In: *Information Systems Research* 15.4 (2004), pp. 336–355.
- [71] Farhad Manjoo. *It's Time for Apple to Build a Less Addictive iPhone*. New York, Jan. 2018. URL: <https://www.nytimes.com/2018/01/17/technology/apple-addiction-iphone.html>.
- [72] Farhad Manjoo. *Tech's Frightful Five: They've Got Us*. May 2017.
- [73] Kirsten E Martin and Helen Nissenbaum. "Measuring Privacy: an Empirical Test Using Context to Expose Confounding Variables". In: *Colum. Sci. & Tech. L. Rev.* 18.176 (2017). URL: <http://dx.doi.org/10.2139/ssrn.2709584>.
- [74] Ginny Marvin. *Google AdWords Turns 15: A Look Back At The Origins Of A \$60 Billion Business*. Oct. 2015. URL: <https://searchengineland.com/google-adwords-turns-15-a-look-back-at-the-origins-of-a-60-billion-business-234579> (visited on 06/05/2018).
- [75] Marwick Alice E. and Danah boyd. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience". In: *New Media and Society* 13.1 (2010), pp. 114–133.
- [76] Claire Cain Miller. *Americans Say They Want Privacy, but Act as If They Don't*. 2014. URL: <https://www.nytimes.com/2014/11/13/upshot/americans-say-they-want-privacy-but-act-as-if-they-dont.html>.
- [77] Linda D Molm. "Risk and Power Use: Constraints on the Use of Coercion in Exchange". In: *Am. Sociol. Rev.* 62.1 (1997), p. 113.
- [78] Linda D Molm. "The Structure of Reciprocity". In: *Soc. Psychol. Q.* 73.2 (2010), pp. 119–131.
- [79] Linda D. Molm, Jessica L. Collett, and David R. Schaefer. "Conflict and Fairness in Social Exchange". In: *Social Forces* 84.4 (2006), pp. 2331–2352.

- [80] Linda D. Molm and Karen S. Cook. "Social Exchange and Exchange Networks". In: *Sociological Perspectives on Social Psychology*. Ed. by J. Fine, G. and House. Needham, MA.: Allyn and Bacon, 1995. Chap. 8, pp. 209–235.
- [81] Linda D Molm, Gretchen Peterson, and Nobuyuki Takahashi. "Power in Negotiated and Reciprocal Exchange". In: *American Sociological Review* 64.6 (1999), p. 876. URL: <http://dx.doi.org/10.2307/2657408>.
- [82] Linda D. Molm, Nobuyuki Takahashi, and Gretchen Peterson. "Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition". In: *American Journal of Sociology* 105.5 (2000), pp. 1396–1427. DOI: 10.1086/210434. URL: <http://www.journals.uchicago.edu/doi/10.1086/210434>.
- [83] Deirdre K Mulligan and Jennifer King. "Bridging the Gap Between Privacy and Design". In: *University of Pennsylvania Journal of Constitutional Law* 14.4 (2012).
- [84] Deirdre K Mulligan, Colin Koopman, and Nick Doty. "Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy". In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 374.2083 (Dec. 2016). DOI: 10.1098/rsta.2016.0118.
- [85] Helen Nissenbaum. "A Contextual Approach to Privacy Online". In: *Daedalus* 140.4 (2011), pp. 32–48.
- [86] Helen Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press, Nov. 2009.
- [87] Helen Nissenbaum. "Respecting Context to Protect Privacy: Why Meaning Matters". In: *Sci. Eng. Ethics* (July 2015).
- [88] Daniel R Horne Patricia A. and David A Horne. "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors". In: *The Journal of Consumer Affairs* 41.1 (2007), pp. 100–126.
- [89] Koray Özpölat and Wolfgang Jank. "Getting the most out of third party trust seals: An empirical analysis". In: *Decision Support Systems* 73 (2015), pp. 47–56. DOI: 10.1016/j.dss.2015.02.016.
- [90] Eyal Peer et al. "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research". In: *Journal of Experimental Social Psychology* 70 (2017), pp. 153–163. DOI: 10.1016/j.jesp.2017.01.006.
- [91] Chanda Phelan, Cliff Lampe, and Paul Resnick. "It's Creepy, But it Doesn't Bother Me". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. 2016. DOI: 10.1145/2858036.2858381.
- [92] Andelka M Phillips. *Take an online DNA test and you could be revealing far more than you realise*. 2016. URL: <http://theconversation.com/take-an-online-dna-test-and-you-could-be-revealing-far-more-than-you-realise-52734>.

- [93] Andrew Pollack. *23andMe Will Resume Giving Users Health Data - The New York Times*. Oct. 2015. URL: [https://www.nytimes.com/2015/10/21/business/23andme-will-resume-giving-users-health-data.html?%7B%5C\\_%7Dr=0](https://www.nytimes.com/2015/10/21/business/23andme-will-resume-giving-users-health-data.html?%7B%5C_%7Dr=0).
- [94] Lorrie Faith Cranor Ponnurangam Kumaraguru. *Privacy Indexes: A Survey of Westin's Studies*. Tech. rep. Carnegie Mellon University School of Computer Science, 2005.
- [95] Sören Preibusch. "Guide to measuring privacy concern: Review of survey and observational instruments". In: *Int. J. Hum. Comput. Stud.* 71.12 (2013), pp. 1133–1143.
- [96] Priscilla M Regan. *Legislating Privacy: Technology, Social Values, and Public Policy*. Univ of North Carolina Press, 2000.
- [97] Thomas Hughes Roberts. "A Cross-Disciplined Approach to Exploring the Privacy Paradox: Explaining Disclosure Behaviour Using the Theory of Planned Behaviour". In: *UK Academy for Information Systems Conference Proceedings 7* (2012).
- [98] Michelle Robertson. *After a 23andMe DNA test, Northern California woman discovers 3 siblings*. Sept. 2017. URL: <http://www.sfgate.com/bayarea/article/23andme-genetic-dna-test-siblings-sperm-donor-12209687.php>.
- [99] Dan Ryan. "Getting the Word Out: Notes on the Social Organization of Notification". In: *Sociological Theory* 24.3 (2006), pp. 228–254.
- [100] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. 2016, 368 pages.
- [101] Sonam Samat, Alessandro Acquisti, and Linda Babcock. "Raise the Curtains: The Effect of Awareness About Targeting on Consumer Attitudes and Purchase Intentions". In: *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. Santa Clara, CA: {USENIX} Association, 2017, pp. 299–319. ISBN: 978-1-931971-39-3. URL: <https://www.usenix.org/conference/soups2017/technical-sessions/presentation/samat-awareness>.
- [102] Charles Seife. *23andMe Is Terrifying, but Not for the Reasons the FDA Thinks*. Nov. 2013. URL: <https://www.scientificamerican.com/article/23andme-is-terrifying-but-not-for-the-reasons-the-fda-thinks/>.
- [103] Kim Bartel Sheehan and Mariea Grubbs Hoy. "Dimensions of Privacy Concern Among Online Consumers". In: *Journal of Public Policy & Marketing* 19.1 (2000), pp. 62–73.
- [104] Howard Shelanski. "Information, Innovation, and Competition Policy for the Internet". In: *University of Pennsylvania Law Review* 161.6 (May 2013). URL: [https://scholarship.law.upenn.edu/penn%7B%5C\\_%7Dlaw%7B%5C\\_%7Dreview/vol161/iss6/6](https://scholarship.law.upenn.edu/penn%7B%5C_%7Dlaw%7B%5C_%7Dreview/vol161/iss6/6).
- [105] Smith et al. "Information Privacy Research: An Interdisciplinary Review". In: *Miss. Q.* 35.4 (2011), p. 989.
- [106] Daniel J Solove. *Understanding Privacy*. Cambridge, MA: Harvard University Press, 2008.

- [107] Jeffrey M Stanton and Kathryn R Sham. “Information Technology, Privacy, and Power within Organizations: a view from Boundary Theory and Social Exchange Perspectives”. In: *Surveillance and Society* 1.2 (), pp. 152–190.
- [108] Richard H. Thaler and Cass S. Sunstein. “Libertarian Paternalism Is Not an Oxymoron”. 2003.
- [109] Richard H. Thaler and Cass S. Sunstein. *Nudge : improving decisions about health, wealth, and happiness*. New Haven: Yale University Press, 2008, p. 293.
- [110] Tobias Dienlin Trepte Sabine and Leonard Reinecke. “Risky behaviors: How online experiences influence privacy behaviors”. In: *Von Der Gutenberg-Galaxis Zur Google-Galaxis. From the Gutenberg Galaxy to the Google Galaxy* (2014).
- [111] Janice Y. Tsai et al. “The effect of online privacy information on purchasing behavior: An experimental study”. In: *Information Systems Research* 22.2 (2011), pp. 254–268. DOI: 10.1287/isre.1090.0260.
- [112] Joseph Turow and Michael Hennessy. “The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation”. In: *SSRN Electronic Journal* (2015). DOI: 10.2139/ssrn.2820060.
- [113] U.S. Food and Drug Administration. *FDA allows marketing of first direct-to-consumer tests that provide genetic risk information for certain conditions*. 2017. URL: <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm551185.htm> (visited on 05/05/2018).
- [114] Janet Vertesi. “How one woman’s experiment opting out of big data ended up making her look like a criminal”. In: *Time Magazine* (May 2014). URL: <http://time.com/83200/privacy-internet-big-data-opt-out/>.
- [115] Jessica Vitak. “The Impact of Context Collapse and Privacy on Social Network Site Disclosures”. In: *Journal of Broadcasting & Electronic Media* 56.4 (2012), pp. 451–470.
- [116] Jeff Warshaw, Nina Taft, and Allison Woodruff. “Intuitions, analytics, and killing ants: Inference literacy of high school-educated adults in the US”. In: (2016). URL: <https://research.google.com/pubs/pub45458.html>.
- [117] Alan Westin. *Privacy and Freedom*. New York: Atheneum, 1967.
- [118] Benjamin Winterhalter. *A genetic ‘Minority Report’: How corporate DNA testing could put us at risk*. Jan. 2014. URL: [https://www.salon.com/2014/01/26/a%7B%5C\\_%7Dgenetic%7B%5C\\_%7Dminority%7B%5C\\_%7Dreport%7B%5C\\_%7Dhow%7B%5C\\_%7Dcorporate%7B%5C\\_%7Ddna%7B%5C\\_%7Dtesting%7B%5C\\_%7Dcould%7B%5C\\_%7Dput%7B%5C\\_%7Dus%7B%5C\\_%7Dat%7B%5C\\_%7Drisk/](https://www.salon.com/2014/01/26/a%7B%5C_%7Dgenetic%7B%5C_%7Dminority%7B%5C_%7Dreport%7B%5C_%7Dhow%7B%5C_%7Dcorporate%7B%5C_%7Ddna%7B%5C_%7Dtesting%7B%5C_%7Dcould%7B%5C_%7Dput%7B%5C_%7Dus%7B%5C_%7Dat%7B%5C_%7Drisk/).
- [119] Rafael Wittek, Tom A B Snijders, and Victor Nee. *The Handbook of Rational Choice Social Research*. Stanford University Press, June 2013.

- [120] T. Yamagishi and M. Yamagishi. "Trust and commitment in the United States and Japan". In: *Motivation and Emotion* 18 (1994), pp. 129–166. DOI: 10.1007/BF02249397.