

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Understanding and Improving Privacy Communication

Permalink

<https://escholarship.org/uc/item/0hw7s8j8>

Author

Smart, Mary Anne

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Understanding and Improving Privacy Communication

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science

by

Mary Anne Smart

Committee in charge:

Professor Kristen Vaccaro, Chair
Professor Lilly Irani
Professor Sorin Lerner
Professor Imani Munyaka

2024

Copyright

Mary Anne Smart, 2024

All rights reserved.

The Dissertation of Mary Anne Smart is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To all of my teachers throughout the years, especially my parents, who were my first teachers. To my partner, Adrian, for his support through all the ups and downs of life as a graduate student.

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Table of Contents	v
List of Figures	ix
List of Tables	xi
Acknowledgements	xiii
Vita	xiv
Abstract of the Dissertation	xv
Introduction	1
Chapter 1 Background	6
1.1 Conceptualizing Privacy	6
1.2 Privacy-Enhancing Technologies	7
1.2.1 Politics of PETs	7
1.2.2 Differential Privacy	7
Chapter 2 Understanding Risks of Privacy Theater with Differential Privacy	9
2.1 Introduction	9
2.2 Related Work	11
2.2.1 Differential Privacy	11
2.2.2 Communicating Privacy	14
2.3 Experimental Design	15
2.3.1 Internet Browsing Histories	15
2.3.2 Experiment 1: Developing Good Explanations of Differential Privacy ..	16
2.3.3 Experiment 2: Measuring Privacy Theater	19
2.3.4 Measures	22
2.3.5 Recruitment	26
2.3.6 Experimental Protocol	26
2.4 Analysis	28
2.4.1 Testing Hypotheses	28
2.4.2 Comparing Comprehension	28
2.4.3 Qualitative Analysis	29
2.5 Results	29
2.5.1 Experiment 1: Developing Good Explanations of Differential Privacy ..	29
2.5.2 Experiment 2: Measuring Privacy Theater	32

2.5.3	Qualitative Results	35
2.6	Limitations	37
2.7	Discussion	38
2.7.1	False Choices	38
2.7.2	Digital Resignation	38
2.7.3	Making Sense of Privacy Promises	39
2.7.4	Other Approaches for Addressing Privacy Theater	41
2.8	Conclusion	41
2.9	Acknowledgements	42
Chapter 3	Models Matter: Setting Accurate Privacy Expectations for Local and Central Differential Privacy	43
3.1	Introduction	43
3.2	Background	46
3.3	Interview Study	50
3.3.1	Initial Prototypes	51
3.3.2	Protocol	52
3.3.3	Participant Recruitment	53
3.3.4	Analysis	54
3.3.5	Findings	54
3.4	Large-Scale Evaluation	61
3.4.1	Results	65
3.5	Limitations	71
3.6	Discussion	71
Chapter 4	Negotiating Privacy Interdependence on Facebook	76
4.1	Introduction	76
4.2	Background	78
4.2.1	Evolving Conceptions of Privacy	78
4.2.2	Negotiating Privacy	79
4.2.3	Reasoning about Inferences	80
4.3	Designing the Friend Inference Generator (FIG)	81
4.3.1	Scraping from Facebook	82
4.3.2	Making Inferences	82
4.3.3	FIG's Four Stages	83
4.3.4	Ethical Considerations	87
4.3.5	Implementation	89
4.4	Methods	90
4.4.1	Recruitment	90
4.4.2	Interview Protocol	91
4.4.3	Selecting Pseudonyms	95
4.5	Analysis	95
4.6	Results	96
4.6.1	Reasoning About Inferences	96

4.6.2	Negotiating Privacy	99
4.7	Limitations	104
4.8	Discussion	105
4.8.1	Challenges of Networked Information	105
4.8.2	Facilitating Privacy Negotiation	106
4.8.3	Avoiding Privacy Theater	108
4.8.4	Utility of Probe Tools	108
4.9	Conclusion	110
Chapter 5	Selling Privacy: Privacy Conceptualizations Promoted Through Advertisements	111
5.1	Introduction	111
5.2	Related Work	113
5.2.1	Politics of Privacy-Enhancing Technologies	113
5.2.2	Critical Discourse Analysis	113
5.3	Methods	114
5.3.1	Data Collection	114
5.3.2	Analysis	115
5.4	Results	116
5.4.1	Viewers, Users, and Victims	117
5.4.2	Adversaries	120
5.4.3	Companies	125
5.5	Discussion	130
5.5.1	Harmful Narratives	130
5.5.2	Useful Techniques	131
5.5.3	Beyond Advertisements	132
Conclusion	134
Appendix A	138
A.1	Demographics	138
A.2	Replication Study	139
A.2.1	Experimental Design	139
A.2.2	Analysis	141
A.2.3	Results	141
A.3	Codes	142
Appendix B	144
B.1	Codes	144
B.2	Survey Instrument	149
B.2.1	Instructions	149
B.2.2	Scenario Description	149
B.2.3	Privacy Description	150
B.2.4	Comprehension Check	150

B.2.5	Trust	150
B.2.6	Self-Efficacy	151
B.2.7	Share	151
B.2.8	Objective Comprehension	152
B.2.9	Thoroughness	153
B.2.10	Subjective Understanding	154
B.2.11	Feedback	154
B.2.12	PETs	154
B.2.13	Background	155
B.3	Demographics	158
B.4	Designs	160
B.5	Descriptive Statistics	165
Appendix C	166
Bibliography	167

LIST OF FIGURES

Figure 2.1.	Perceived and actual understanding for Experiment 1. Most participants felt that they understood the explanations well (left), yet answered no more than one question correctly (right). Performance was similar across all four conditions.	30
Figure 2.2.	Performance on paired questions.	31
Figure 2.3.	Willingness to share for the three explanations and three settings of the privacy parameter. These figures show the averages for all participants (left) and participants who were initially unwilling to share their information (right).	32
Figure 2.4.	Willingness to share before and after differential privacy explanation. Most participants' answers did not change (left) when asked the second time. ...	33
Figure 2.5.	Perceived and actual understanding for Experiment 2. As in Experiment 1, participants' perceived understanding (left) is greater than their actual understanding (right), though the added questions improved comprehension scores compared to Experiment 1.	34
Figure 3.1.	The final version of our privacy labels for the local (left) and central (right) models.	62
Figure 4.1.	FIG's four stages. Each of the four stages serves a particular purpose. Scraping continues in the background during the first two stages. In the final two stages, FIG displays the inferences made from the scraped data. ..	77
Figure 4.2.	Strategy for making inferences. To make an inference about a particular friend, FIG turns to that friend's mutual friends with the participant. A simple voting process determines which inferences FIG makes.	83
Figure 4.3.	Stage 1. FIG provides an external link to the user's actual live profile on Facebook.	84
Figure 4.4.	Stage 2. This page shows how Facebook might use information shared by "Abby."	84
Figure 4.5.	Stage 3. This page shows a list of friends who shared in their about sections that they lived in Indianapolis (left) and a list of friends that FIG inferred lived in Indianapolis even though it was not shared in their about sections (right). Names in this screenshot are fabricated.	85
Figure 4.6.	Stage 4. FIG displays information from a friend's about section (gray) and any inferences made about that friend (blue).	87

Figure 5.1.	Diverse genders and racial backgrounds in “A New Era of Personal Privacy with Default End-to-Encryption”.	117
Figure 5.2.	The three frames from NordVPN’s “I am a fighter, but you gotta help me here, guys! Online privacy is no joke #shorts”.	119
Figure 5.3.	Stereotypical hacker depicted in “7 Tips to Secure your Wi-Fi Router — NordVPN”.	121
Figure 5.4.	The auctioneer gestures toward a holograph of Ellie in “Privacy on iPhone — Data Auction — Apple”.	123
Figure 5.5.	“Sweet Nana” sits on the auction stage.	124
Figure 5.6.	The colors change from yellow to black-and-white when guest mode is activated in “Guest Mode on Google Assistant”.	126
Figure 5.7.	Varied aspect ratios in “DuckDuckGo: None of Our Business”.	129
Figure A.1.	Replication study results.	141
Figure B.1.	Example of the Miro board setup used for the follow-up interviews.	161
Figure B.2.	Top: Diagram for local model. Bottom: Diagram for central model.	162
Figure B.3.	Original Privacy Labels.	163
Figure B.4.	Evolution of designs over time.	164

LIST OF TABLES

Table 2.1.	Example Table	16
Table 2.2.	Settings of Privacy Parameter	16
Table 2.3.	Explanations for Experiment 1	18
Table 2.4.	Low, Medium, and High Transparency Explanations for Experiment 2. All explanations are for the high privacy setting. The low transparency explanation conveys no information about the privacy parameter.	21
Table 2.5.	Comprehension Questions. All questions are multiple choice. Correct answers are in parentheses. For questions #3 and #4, the correct answer depends on the experimental condition.	24
Table 3.1.	Five Information Disclosures.	48
Table 3.2.	Explanation texts for the local and central models.	63
Table 3.3.	<i>Left:</i> results from linear regression models for objective comprehension. <i>Right:</i> results from ordinal regression models for subjective understanding.	66
Table 3.4.	Results from regression models for trust, perceived thoroughness, and self-efficacy, with the Xiong et al. explanation as the reference level explanation. Again we report odds ratios with 95% CIs. An OR > 1 indicates an increase in odds.	67
Table 3.5.	Familiarity with PETs	69
Table 3.6.	Results from regression model for data-sharing decision. We report odds ratios (OR) and corresponding 95% CIs. An OR > 1 indicates an increase in odds.	70
Table 4.1.	Selected interview questions for each stage. This table provides examples of the kinds of questions asked during interviews, though it is not a complete list.	91
Table A.1.	Participant Demographics	138
Table B.1.	Participant Demographics: Initial Interviews	159
Table B.2.	Participant Demographics: Follow-up Interviews	159
Table B.3.	Respondent Demographics	160

Table B.4. Original Metaphor Descriptions 161

Table B.5. Accuracy of Privacy Expectations 165

Table C.1. Participant Demographics 166

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Kristen Vaccaro, for her support and mentorship over the past several years. Kristen believed in me and supported my ideas even when I was unsure of myself. In addition to her invaluable support of me as a researcher, she has also been supportive of my efforts to become a better teacher. I am thankful to have had such a supportive advisor.

I am grateful to my committee members—Lilly Irani, Sorin Lerner, and Imani Munyaka—for their feedback and guidance. I am also grateful to all my collaborators. I am especially thankful to have had the opportunity to collaborate with Priyanka Nanayakkara, Elissa Redmiles, Rachel Cummings, and Gabe Kaptchuk through my experience as a visiting student at Columbia University. I have learned so much from this group and always enjoy working with them.

I am also grateful to my family and friends for their support over the past six years. To Mom, Dad, and Bobby—thank you for always being there for me. To Casey Meehan, Robi Bhattacharjee, Sormeh Yazdi, Jacob Imola, and Priyanka Nanayakkara—getting to know you all was the best part of graduate school. To all my lab mates and fellow Ph.D. students—thank you for building a supportive community at UCSD. To Dasha Kopulsky—thank you for visiting me in San Diego, for pilot testing my surveys, and for always being someone I can talk to. To Adrian Wolanski—thank you for accompanying me on this journey.

Chapter 2 is a reprint of the material as it appears in *Proceedings of the ACM on Human-Computer Interaction*. (M. A. Smart, D. Sood, K. Vaccaro. Understanding risks of privacy theater with differential privacy. *Proc. ACM Hum.-Comput. Interact.*, 6 (CSCW2), 2022.) The material in chapter 3 has been submitted for publication. (Smart, M. A., Nanayakkara, P., Cummings, R., Kaptchuk, G., Redmiles, E. M. Models Matter: Setting Accurate Privacy Expectations for Local and Central Differential Privacy.) The material in chapter 4 has been submitted for publication. (Smart, M. A., Broukhim, A., Sekhon, B., Tan, S., Vaccaro, K. Negotiating Privacy Interdependence on Facebook.) The material in chapter 5 has been submitted for publication. (Smart, M. A., Li, S., Liu, T., Vaccaro, K. Selling Privacy: Privacy Conceptualizations Promoted Through Advertisements.) The dissertation author was the primary author of the above papers.

VITA

- 2017 Bachelor of Arts in Spanish, Indiana University
- 2017 Bachelor of Science in Mathematics, Indiana University
- 2017 Bachelor of Science in Computer Science, Indiana University
- 2024 Doctor of Philosophy in Computer Science, University of California San Diego

ABSTRACT OF THE DISSERTATION

Understanding and Improving Privacy Communication

by

Mary Anne Smart

Doctor of Philosophy in Computer Science

University of California San Diego, 2024

Professor Kristen Vaccaro, Chair

Privacy communication takes many forms, from blog posts to news articles to advertisements. Existing privacy communication often falls short, misleading and discouraging readers. For example, confusing descriptions of privacy-enhancing technologies can lead users to overestimate their protective capabilities. In this dissertation, I examine the shortcomings of existing privacy communication, study the effects of privacy communication on data-sharing behavior, and explore opportunities for improvement. I use a range of methods, such as online surveys and interview studies, to design better ways to communicate about privacy that support engagement with privacy issues.

Introduction

Tech executives have sometimes claimed that privacy belongs to the past and that people no longer care about privacy [156, 207]. The truth, however, is more complicated. Researchers have consistently observed apparent inconsistencies between people’s expressed privacy concerns and their actions [177, 26]. For example, individuals’ stated data-sharing intentions tend to underestimate their actual willingness to share personal information [240]. This mismatch between privacy attitudes and behavior has been dubbed the “*privacy paradox*” [41]. One explanation of this phenomenon, supported by evidence from prior work, is that while people do care about privacy, they feel powerless to protect it [76, 128, 365].

It is hardly surprising that many people feel resigned about privacy issues. Much of what happens to our personal information is indeed out of our control. For example, consider the case of social media. Suppose that a particular user wishes to keep some of her personal information private. Even if she chooses not to share this information directly, the social media platform may be able to infer it using information about her social network or other information they have collected about her [298, 61]. Even reasoning about these threats may be quite challenging for many users [296]. Furthermore, tech companies employ a number of strategies that actively discourage engagement with privacy issues; for example, lengthy privacy policies written in confusing legalese feed feelings of powerlessness [76]. These feelings of powerlessness are not unwarranted—there is indeed a significant imbalance of power between data collectors and data subjects. Changing these power dynamics is a key part of addressing problems of privacy and feelings of resignation.

Scalable Strategies

A number of solutions have been proposed to address the inherent power imbalances between data collectors and data subjects that underlie most privacy issues. Of these proposed solutions, a few stand out as particularly scalable strategies; privacy advocates have called for improved technologies [279, 300, 56], for collective action campaigns [59, 158], and for legislative solutions [297, 349]. Each approach has its benefits and its limitations, but these strategies need not be viewed as mutually exclusive; rather, they can complement each other and contribute to the struggle for strong privacy protections.

One approach for empowering data subjects is to offer them technological solutions. In particular, privacy-enhancing technologies, such as encryption, have been proposed as tools to help people gain more agency of their privacy. Rogaway has argued that “*cryptology rearranges power*” and is thus an “*inherently political tool*” [272]. Strong cryptography allows individuals to communicate and exchange information secretly, even against powerful adversaries such as nation-states or multinational corporations. Put another way, “*a little bit of math can accomplish what all the guns and barbed wire can’t: a little bit of math can keep a secret*” [288]. When privacy-enhancing technologies are deployed at scale—for example, when major messaging platforms adopt end-to-end encryption [120]—mass surveillance by state or corporate actors becomes much more difficult. Nevertheless, while privacy-enhancing technologies certainly have a role to play, they are generally not a satisfactory solution in and of themselves [316, 123]. Privacy protection should not require extensive technical expertise, yet issues with communication and usability limit the accessibility of many privacy-enhancing technologies [359, 361, 135]. Furthermore, in some circumstances, these technologies may actually normalize ubiquitous surveillance [380].

While privacy-enhancing technologies are all too often individualistic in nature, offering protection only to the individuals with enough tech savvy to use them correctly, collective strategies may be better suited to addressing privacy risks that are collective in nature [321]. Machine

learning models trained on data from millions of individuals can infer personal information about other people; this kind of threat cannot be countered through the action of one individual deleting their data. Thus, recognizing the limitations of individualistic approaches, scholars have directed their attention to the question of how to support collective action for privacy issues [158, 377, 59, 343]. Collectives may push for companies to adopt privacy-enhancing technologies like encryption at scale—offering messaging that is encrypted by default and thus requires little technical expertise. Alternatively, collectives may advocate for legislative solutions. Unfortunately, a number of challenges exist that make effective collective action difficult. For example, “*the nature of the collective harms*” is often “*invisible to the average person*” [321]. Furthermore, even people who are aware of the invisible algorithms that threaten privacy may be reluctant to act if they have resigned themselves to the idea that privacy is dead.

Regulation is another frequently suggested avenue for reigning in the power of data collectors. Again, the appeal of this approach lies largely in its ability to scale. An individual acting alone has little power against invasive and irresponsible data practices. A government acting on behalf of millions of such individuals is in a much more powerful position to address such issues. Calls for more government oversight often follow widely-publicized privacy debacles—one such example is the infamous 2017 Equifax data breach [255]. Over 100 million social security numbers were leaked in this data breach, leaving millions of people vulnerable to fraud and identity theft. Taking effort to freeze one’s credit or set up credit monitoring requires time and knowledge for the millions of affected individuals. The costs—both in terms of time and money—are particularly high for people who experience identity theft as a result of the data breach. Helping millions of individuals effectively cope with the aftermath of the Equifax data breach is much more expensive than it would have been for Equifax to invest in proper cybersecurity. Experts have suggested that better regulation might fix misaligned incentives for companies like Equifax and encourage them to follow best data practices, preventing future data breaches [378, 107]. This is merely one example of how regulation may address an issue at scale that is difficult for people to deal with individually.

Privacy Communication

Privacy communication takes many forms, including blog posts, privacy policies, advertisements, and news articles. Improved communication can support and complement a range of privacy strategies, including those described above. For example, understanding where existing communication falls short and designing more effective communication surrounding privacy-enhancing technologies can help people better understand both their advantages and their limitations. Improved communication can also help people better understand and navigate collective privacy threats by unveiling the hidden algorithms that are always operating in the background. Designing new ways to communicate about privacy may even go so far as to shift people's perceptions of what privacy is and why it matters.

Contributions

In this dissertation, I identify shortcomings in privacy communication and develop explanations and tools to help people deepen their understanding of privacy. Chapters 2 and 3 focus on explaining the privacy implications of differential privacy—a privacy-enhancing technology that is particularly interesting both for its political implications and for the challenges it poses for developing effective communication [38, 57, 233]. Chapter 4 discusses a tool designed to help Facebook users reason about algorithmic inferences and privacy interdependence. Finally, chapter 5 discusses conceptualizations of privacy in advertisements. I describe the contributions in more detail below.

In chapter 2, I design explanations of differential privacy that highlight the implications of a key implementation parameter. I conduct an online survey to study how the choice of explanation and the choice of privacy parameter affect data-sharing behavior. The results indicate that respondents prioritized other factors—such as their trust (or lack thereof) in the data collector—when making data-sharing decisions. Neither the choice of explanation nor the parameter setting had any significant effect on respondents' behavior.

In chapter 3, I design explanations that focus on a different aspect of implementations of differential privacy. In the local model of differential privacy, each individual's information is modified to preserve privacy before it is collected and stored. In the central model, on the other hand, a trusted data curator collects unmodified data. Privacy modifications are made later—for example, in the reporting of summary statistics. I explore a range of designs, evaluated through both an interview study and an online survey. The privacy labels, which list key privacy implications in a tabular format, help respondents understand which data flows are protected. I speculate as to how these designs might be extended to other privacy-enhancing technologies.

In chapter 4, I design an interface to help Facebook users reason about privacy interdependence. The tool shows examples of how Facebook might use the information someone shares to make inferences about their friends. An interview study conducted with the tool finds that algorithmic inferences complicate the way friends negotiate privacy concerns together. I discuss the implications of this study for the design of social media platforms.

In chapter 5, I collect and analyze YouTube videos advertising privacy-related features or products. The ads use a range of visual and rhetorical techniques to shape narratives about privacy that support the companies' best interests. For example, when Google claims to offer improved "*control over how your data is used and saved,*" the use of the passive voice disguises Google's own agency and positions the burden of privacy management on the shoulders of individual users. I discuss how the narratives that benefit tech companies may be harmful and how the techniques these ads employ may be borrowed in service of privacy education.

Chapter 1

Background

1.1 Conceptualizing Privacy

People often disagree about how to define privacy and about what privacy is for [241, 231, 272, 20, 130]. Scholars have proposed a number of different theories, definitions, and taxonomies to make sense of privacy as a concept [294, 239, 216]. In order to analyze how advertisements conceptualize privacy, it is useful to first offer an overview of some existing tensions in how privacy is understood.

Privacy has often been conceptualized as an individual right; Warren and Brandeis conceptualize privacy broadly as the “*right to be let alone*” [353], while other scholars have offered the more limited notion of privacy as control over one’s own personal information [228]. Lone individuals, however, have very little power to control what information is collected or inferred about them [296]. Since much of our data is interconnected with the data of others, and since privacy threats often operate at a group level, scholars have thus recognized the benefits of thinking about privacy collectively [209, 211, 321, 74, 377]. Solove’s influential taxonomy of privacy harms acknowledges that privacy violations can cause harm both at the individual and societal level [294]. Due to its relative comprehensiveness and its popularity in the literature, we use this taxonomy to characterize the privacy harms depicted in advertisements.

It is also common to view privacy as equivalent to *secrecy*; this view is particularly dominant in computer science [123]. Yet this understanding of privacy is also limited. It suggest

a false binary—that either information is secret, or it is public [293]. Nissenbaum offers an alternative and argues that “*a right to privacy is neither a right to secrecy nor a right to control but a right to appropriate flow of personal information*” [239]. This concept of appropriate information flows allows for a more nuanced understanding of privacy. Complications emerge, however, in determining which flows are appropriate [216].

1.2 Privacy-Enhancing Technologies

Privacy-enhancing technologies (PETs) are tools that help people protect the privacy of their information and/or communication. Below, we discuss the politics of PETs and provide an overview of a particular PET that is becoming increasingly popular—differential privacy.

1.2.1 Politics of PETs

PETs, such as encryption, are more than mere technical tools—they also have a political dimension [272]. PETs are often touted as the “solution” to mass surveillance, but this framing reduces larger political issues demanding solidarity to technological puzzles best solved by academic cryptographers [124, 272, 277]. In fact, PETs may have the counterintuitive effect of further normalizing surveillance—thus actively hampering anti-surveillance efforts [380]. Underlying the “*logic of [most] PETs*” is the implicit assumption that “*users are individually responsible for minimizing the collection and dissemination of their personal data*” [123]. This, however, is no small responsibility, especially since using PETs generally requires some technical expertise [316, 359]. Thus, instead of challenging surveillance in a way that builds collective power, overreliance on PETs means that “*challenging data collection becomes an individualized act based on perceived skill and ability to engage in privacy-enhancing digital practices*” [64].

1.2.2 Differential Privacy

Differential privacy has received a great deal of attention in academia, industry, and government, for its potential to empower data analytics with provable privacy guarantees. Dif-

ferential privacy is a PET of particular political importance, as it has recently been adopted by government agencies, most notably the US Census Bureau [4, 226]. The goal of differential privacy is to allow data analysts to learn from *aggregate* statistics while limiting the leakage of information that is specific to any *individual*. Differentially private mechanisms add noise to data to accomplish this. However, as more noise is added, the data becomes less useful for analysis. All data collection and analysis involves tradeoffs between privacy and accuracy. Differential privacy, however, makes this tradeoff explicit. A privacy parameter ϵ (which is related to the amount of noise injected into the data) controls the tradeoff between privacy for individuals and accuracy of aggregate data analysis. Therefore, the choice of this parameter has important practical and political consequences [5, 187].

Two common variants exist: central differential privacy and local differential privacy. In the central model, a central agent is responsible for storing the raw data and adding noise. In the local model, individuals add noise to their own data before sending it to the data collector. Thus, the local model offers stronger privacy protection, since it does not require trust in a central agent. Both models are popular in practice [69, 86, 4, 66]

Chapter 2

Understanding Risks of Privacy Theater with Differential Privacy

2.1 Introduction

In recent years, the HCI community has devoted increasing attention toward the issues of mass data collection and surveillance capitalism [380, 109, 344]. The role of privacy-enhancing technologies (PETs) within this landscape is complicated. On the one hand, PETs can offer protection from and resistance against harmful data collection practices [44]. On the other hand, some PETs may actually normalize surveillance, increase the power of data collectors, de-politicize surveillance issues by reframing them as technological puzzles, or otherwise distort political discourse around surveillance issues [380, 272, 316, 124]. In addition, these PETs may encourage *privacy theater*, where they provide the “*feeling of improved privacy while doing little or nothing to actually improve privacy*” [170]. We investigate this possibility using one popular PET: differential privacy.

Differential privacy has become one of the most widely used tools for privacy-conscious data analytics, with deployments across industry and government agencies (e.g., Google [86], Apple [69], US Census Bureau [4]). Differential privacy allows these organizations to collect data while protecting privacy by adding small amounts of statistical noise to the data that people share [80]. Importantly, the privacy protection provided by differential privacy is highly dependent on the setting of certain algorithm parameters, which control how much noise is added.

There is an inherent tradeoff, however, between the noise added (privacy protection provided) and the utility of the data for the organization. As a result, there is a risk that companies may misrepresent the actual privacy benefits afforded by differential privacy to persuade users to share more information. By understanding how users respond to different kinds of explanations of differential privacy, we can better understand whether users are likely to be harmed by explanations that obscure these tradeoffs.

Recent work has begun to probe users' understanding of differential privacy. For example, recent work studied how different explanations of differential privacy influenced users' (planned) willingness to share their data [367]. These explanations did not describe the role of algorithm parameters to users because the authors argued that users would struggle to understand how parameter settings would affect their privacy. A different study of differential privacy, however, provided transparency into the amount of noise added to participants' data. This transparency increased participants' comfort, understanding, and trust in the security of the differentially private mechanism [45]. In this work, we expand on these efforts, developing explanations of differential privacy that convey information about algorithm parameters and investigating whether incomplete explanations of differential privacy may mislead users.

Through multiple large-scale, online surveys, we study how different explanations of differential privacy influence participants' understanding and behavior when asked to share browser history data. Drawing and expanding on prior work, the first experiment aims to generate a "best possible" explanation for differential privacy, varying 1) whether the explanation includes a visual component, and 2) whether the explanation focuses on the process or outcome of the differentially private mechanism. The second experiment tests for potential *privacy theater*, pitting the best-performing explanation against standard industry explanations and observing whether industry explanations obscure risks of weak privacy settings. Unlike prior work [367], we do not assume that more willingness-to-share is always preferable. Instead, we hypothesize that good explanations will decrease willingness-to-share when privacy protection is poor.

Our results reveal a number of surprising findings. First, we find that participants

are overconfident in their understanding; although they feel that they have understood the explanations well, they perform poorly on comprehension questions. In particular, a surprisingly large number of participants do not understand that differential privacy offers protection from a wide range of adversaries—not just from hackers. This suggests that participants find the nature of the protection offered by differential privacy to be counterintuitive. The second surprising finding is that the explanations of differential privacy have little effect on individuals’ willingness to share browser history data. Only a small fraction of participants—less than 25%—actually change their minds after reading about the privacy protection; most participants have already decided whether or not to share their data. Our qualitative results help explain the reasons for this behavior. Many participants list other reasons that drive their decision making around disclosing browser data, such as trust (or distrust) in the research team. Many participants simply feel that since they have “nothing to hide,” they may as well share their information. For these participants, promises of privacy protection are largely irrelevant.

2.2 Related Work

Attempts to develop good explanations of differential privacy face a key challenge: it is difficult to explain how specific algorithm parameters affect privacy risk. In fact, understanding how to best set the privacy parameter remains a challenge even for experts [154, 4]. This section begins with some necessary background about differential privacy. It continues with an overview of the relevant prior work on privacy-related communication.

2.2.1 Differential Privacy

This paper focuses on *local* differential privacy, which has been deployed in a number of industry contexts [86, 69].

Formal Definition

The formal definition of local differential privacy reveals the importance of a parameter for how much privacy protection is provided. It states [80] that a randomized algorithm A is **ϵ -LDP** if and only if for any inputs u, v and any outputs $y \in \text{Range}(A)$:

$$\ln \left(\frac{P(A(u) = y)}{P(A(v) = y)} \right) < \epsilon$$

Since this definition requires the inequality to hold for all values of u, v , and y , it offers a kind of worst-case guarantee; replacing input u with any other input v will not change the probability distribution over algorithm outputs much, provided that the privacy parameter ϵ is small. In effect, when ϵ is small, anyone looking at an individual datapoint will be uncertain about its *true* value. Note that this privacy parameter¹, which we discuss in more detail below, is an essential factor in determining just how much privacy is offered.

Tradeoffs

To illustrate the tradeoff the privacy parameter makes between privacy and utility, we describe a popular survey technique that satisfies LDP²: the randomized response technique. Suppose a researcher wants to estimate the proportion of students at a particular school who use illegal drugs [117]. Since “*Have you ever used illegal drugs?*” is a sensitive question that many students may be reluctant to answer, the researcher gives students the following instructions. Privately flip a coin. If the coin toss is heads, write down your true answer. Otherwise, flip the coin again. If the second coin toss is heads, answer *yes*; if tails, answer *no*. This procedure offers a form of privacy to students; someone who saw a student’s response would not know whether it was the true response or not. In fact, this mechanism is $\ln(3)$ -LDP [86]. Nevertheless, these answers allow an estimate of the proportion of students that use illegal drugs. If out of n students

¹The privacy parameter ϵ is often referred to as the privacy *budget*. In other variants of differential privacy, there may be multiple privacy parameters, but in our setting, we can unambiguously refer to the privacy budget as *the* privacy parameter.

²Though RRT was invented before the formalization of differential privacy [352]

surveyed, y respond *yes*, then we can estimate that the true proportion of students who use illegal drugs is about $2(\frac{y}{n} - \frac{1}{4})$.

The privacy parameter ϵ controls the tradeoff between privacy and utility. Suppose that instead of flipping a coin, students roll a six-sided die. The student is instructed to report their true answer unless they roll a 1, in which case they flip a coin and report *yes* for heads and *no* for tails. This provides a more accurate estimate of how many students use illegal drugs, however, it reduces the privacy offered to students. Someone who sees a student's response to the question has more reason to believe that the response is true. This mechanism no longer satisfies $\ln(3)$ -LDP, but it does satisfy $\ln(11)$ -LDP. This example illustrates how increasing the privacy parameter ϵ increases the usefulness of the data for analysis, but weakens the privacy guarantee.

Industry Deployments

This privacy parameter setting has become an area of contention for industry deployments. One early use of LDP was Google's deployment of RAPPOR (which builds upon randomized response [86]), to collect information about certain Chrome users' browser settings. Apple followed suit, deploying its own versions of LDP to collect data for a variety of applications including discovery of new words, discovery of popular emojis, and analysis of memory usage in the Safari browser [69]. But Apple faced criticism for its initial decision not to publish the privacy parameter ϵ . When researchers reverse-engineered Apple's implementation of LDP on MacOS 10.2, they found the privacy loss to be "*significantly higher than what is commonly considered reasonable in academic literature*" [314]. In fact, this lack of transparency has been recognized as such a significant issue that one of the inventors of differential privacy called for the creation of an "Epsilon Registry," for firms to report implementation details such as the choice of privacy parameter ϵ . The underlying concern is that "*when ϵ is large it can [...] allow for a form of privacy theatre*" while offering little privacy protection [82].

2.2.2 Communicating Privacy

Researchers have extensively studied the questions of how privacy-related communication affects user behavior [313, 15] and how to communicate privacy risks effectively [167, 168, 96, 110, 24, 115, 83]. Previous studies have pointed out the shortcomings of traditional privacy policies [242, 327]. In theory, privacy policies could help people make informed decisions about their privacy. In practice, however, this is rarely the case [215]. In fact, privacy policies can actually be misleading. For example, people sometimes incorrectly interpret the mere presence of a privacy policy as an assurance that the website does not share users' data with third-parties [325]. Similar to this prior work, we are concerned that ineffective communication could give users a false sense of security.

Recently, these efforts have extended to how to communicate with users about differential privacy [45, 367, 57]. The study most closely related to ours is that of Xiong et al., which examines how different explanations of differential privacy shape users' willingness to share their data [367]. However, this work evaluates the quality of explanations in terms of users' resulting willingness to share. They state, for example, that *"when definitions of DP and LDP were communicated . . . participants increased their data disclosure for high-sensitive information, suggesting a positive effect of communicating differential privacy to laypeople."* In this work, we explore the potential risks involved in this framing and investigate whether misleading explanations can lead to increased data sharing when little actual privacy protection is provided.

Several previous studies have investigated factors that can shift users' willingness to share their data in different contexts. For example, iPhone users who are asked to grant certain permissions to an app are more likely to grant the permissions when the app provides an explanation of why the requested permissions are necessary [313]. A more nefarious example can be found in the study of phishing. Phishing emails seek to trick individuals into disclosing valuable information, such as banking credentials. Phishing emails that are information-rich tend

to be more successful [98]. In this work, we seek to understand the malleability of participants' existing data sharing preferences and to what extent participants' behavior can be shaped by explanations of differential privacy.

2.3 Experimental Design

Differential privacy is an approach for providing privacy by adding noise to the data users provide, resulting in a tradeoff between privacy protection (which increases as noise is added) and the usefulness of the data (which decreases as noise is added). This study develops explanations of differential privacy that make the necessary tradeoffs between privacy and utility explicit. The first experiment identifies the explanation that is most effective at communicating this information to users. This experiment tests four explanations' effectiveness at improving users' perceived and actual understanding of differential privacy. The second experiment measures how well explanations can convey the risks of poor settings of algorithm parameters and whether better understanding can induce more appropriate behavior (specifically, lower willingness to share data).

2.3.1 Internet Browsing Histories

The setting used for these experiments was collecting internet browsing history data. Prior experiments on explaining differential privacy have assessed *hypothetical* willingness to share data [367]. Unlike this prior work, our experiment seeks to have participants *actually decide* whether or not to share their private data. Browsing history data is an appropriate choice since many people consider browsing history data sensitive [259, 281]. Furthermore, many large-scale industry deployments of differential privacy have involved the collection of browser-related data [86, 69]. It is also plausible that the experiment could access participants' actual data since the experiment is conducted through an online survey.

Participants were told that we were interested in collecting internet browsing histories, focused on a list of 200 websites of interest. Using participants' browsers, an automated script

Table 2.1. Example Table

#	Website	Visited?
1	example.com	YES
2	example.org	NO
...
200	example.net	YES

Table 2.2. Settings of Privacy Parameter

Flip Probability	ϵ
1%	920
25%	220
49%	9

would collect all websites visited over the past 12 months. From this list, we could construct a table showing which of the 200 websites had been visited (Table 2.1). To satisfy differential privacy, the data collection would randomly change some answers. Participants were told that only this modified data would be shared with the researchers; the original data would never leave the participants’ browser. The explanation provided to participants is included in Table 2.3.

The percentage of changed answers is directly related to the privacy parameter ϵ (Table 2.2). By varying the percentage of answers that are changed, we vary the privacy parameter. As a result, the privacy parameter can be changed as an experimental condition.

While participants were told that they were deciding whether to share data, no browsing data was actually collected. Participants were debriefed on this deception at the end of the experiment. We did not deceive participants about our identity as researchers. Participants were explicitly informed that any data they shared would be shared with an academic research team. Our Institutional Review Board determined that the protocol was exempt from full IRB review.

2.3.2 Experiment 1: Developing Good Explanations of Differential Privacy

The first experiment evaluated which explanations of differential privacy would be easy for participants to understand. Prior work found that explanations that covered the implications of local differential privacy—rather than the definition or process of adding noise to data—improved participants’ comprehension [367]. And work on descriptions of encryption found that result-oriented descriptions led to greater feelings of security than process-oriented descriptions [72]. Nevertheless, explanations that fail to explain the underlying processes may confuse users, and

other authors have argued that “*visibility of system behavior*” is essential for encouraging good decision-making around privacy and security [75]. To test which is most useful to increase understanding of differential privacy, we developed two explanation texts, one of which focuses on the *outcome* of the privacy-preserving process and the other of which focuses on the *process* itself. This led to our first hypothesis³:

H1 *Process vs. Outcome*: Explanations focused on outcomes will increase understanding more than explanations focused on process.

Previous work has also found visual aids to be helpful in aiding understanding of differential privacy [45] and in understanding other privacy-related issues [358, 338, 214]. Visualizations have also played an important role in attempts to increase the explainability and interpretability of machine learning models [282, 193]. However, other work has found that people with limited statistics background often struggle to interpret visualizations related to probability distributions [148]. This presents a challenge, since it is important that our explanations convey information about the randomness inherent to differential privacy. These findings from previous work led to our second hypothesis:

H2 *Visual explanations*: Explanations providing a visual component will improve understanding compared to a text-only version of the explanation.

Experimental Conditions

To test these hypotheses, we use a 2^2 factorial design. The two factors are content and medium. Each has two levels; the content provides process- and outcome-focused content, and the medium provides either a text-only or text+visual explanation.

The four explanations are roughly matched for length (160-163 words) and reading level (6-8 on the Flesch-Kincaid grade level scale [174]). All also convey the same key information:

³All four hypotheses were preregistered with the Open Science Framework.

Table 2.3. Explanations for Experiment 1

Process	Outcome
<p>We will collect information about your internet browsing history using a method designed to protect privacy. The method provides local differential privacy. We have a list of 200 websites that interest us. First, the method will look at your browsing history from the past year. It will ignore any websites that are not on our list of interest. Then it will make a table. The table will show which of the 200 websites you visited. The table would look something like this: (see Table 2.1).</p>	
<p>The method will randomly select some rows in the table to change. Each row has a 49% chance of being selected to change. For these rows, the method will change the YES answers to NO and the NO answers to YES. We will not know which rows were selected. These changes will happen before you send us your information.</p>	<p>The method will have changed your information before you send it. For most people, about 98 of the 200 YES or NO answers will have changed. The changes will be different for each person. We will never have access to the original table. We also will not know what parts of your browsing history were changed.</p>
<p>Differential privacy changes each individual's information. But since we collect information from many people, we can still see overall patterns that interest us.</p>	

1) the kind of changes that would be made to protect privacy, i.e., that the record of whether particular websites had been visited would be changed, and that these changes would be made at random, 2) information related to the privacy parameter, i.e., how many changes would be made, and 3) the fact that the researchers would only see the altered data, and that as a result the researchers would be uncertain about which websites were actually visited. Four pilot testers—including one differential privacy expert—were recruited to evaluate the explanations on the basis of correctness, interpretability, and clarity. Discussions with these pilot testers revealed a desire to understand how the collected data could be used after undergoing the randomized privacy-preserving process; a final sentence clarifies this. Table 2.3 provides the explanations. One challenge in developing visual explanations of differential privacy is how to convey the

randomness inherent to all differentially private mechanisms. Our visualizations belong to the broader class of hypothetical outcome plots, which use animation to simulate the outcomes of multiple random draws [246].

2.3.3 Experiment 2: Measuring Privacy Theater

The second experiment examines how different explanations influence participant behavior — particularly willingness to share data. Of particular concern is how users decide whether to share data when presented with explanations that provide little to no transparency into the setting of the privacy parameter. We refer to these explanations that omit any discussion of the privacy parameter and its implications as *low-transparency* explanations; similarly, we refer to explanations that *do* discuss the implications of the privacy parameter as *high-transparency* explanations. If users are more willing to share data when given a low-transparency explanation than when given a high-transparency explanation, then data collectors may be incentivized to use low-transparency explanations. Furthermore, data collectors may be tempted to set inappropriately large values for the privacy parameter and thus abuse differential privacy as privacy theater.

Prior work suggests that explaining differential privacy’s protections can make people more willing to share sensitive data [367, 99]. Highlighting the particular benefits of *local* differential privacy has also been found to increase willingness to share [367]. So when the privacy protection offered is strong— i.e., the privacy parameter is small—highlighting this fact will likely increase willingness to share. We hypothesize that in the strong privacy setting (i.e. small ϵ), participants would be more willing to share their data when presented with a more transparent explanation that better highlights the strength of the privacy guarantee.

H3 *Strong Privacy Setting:* When the privacy protection is strong, users provided with a high-transparency explanation will be more willing to share data than those provided with a low-transparency explanation.

Prior work demonstrates that even explanations that omit any discussion of the privacy parameter can increase participants' willingness to share data [367]. In the weak privacy setting (i.e. large ϵ), an explanation of the privacy parameter might reveal the weakness of the privacy protection being offered, such that the increased willingness to share observed in [367] would disappear. Indeed, previous work on the randomized response technique (RRT)—the simplest variant of differential privacy—suggests that this is likely to happen, since it has been shown that participants' trust, comfort, and level of perceived protection fall as the privacy parameter increases [45, 290]. We hypothesize that in the weak privacy setting (i.e. large ϵ), participants would be less likely to share their information when given more transparent explanations.

H4 *Weak Privacy Setting:* When the privacy protection is poor, users provided with a low-transparency explanation will be more willing to share data than those provided with a high-transparency explanation.

Experimental Conditions

To test these hypotheses, we use a 3^2 factorial design. The two factors are transparency and privacy parameter setting. Each has three levels; the privacy parameter setting levels provide high, medium, and low privacy protection, and the explanations provide high, medium, and low transparency into the privacy parameter.

The first factor is the privacy parameter setting. We have a low privacy setting ($\epsilon = 920$), a medium privacy setting ($\epsilon = 220$), and a high privacy setting ($\epsilon = 9$). In general, the low privacy setting changes very few of users' true responses, while the high privacy setting changes many of the true responses. Therefore, as the level of privacy increases, the data becomes less useful for the data collector. Table 2.2 indicates exactly how these ϵ values translate into privacy protection.

The value of the privacy parameter in the high privacy setting is quite close to what has been used for certain industry deployments [69]. Nevertheless, future work could consider

Table 2.4. Low, Medium, and High Transparency Explanations for Experiment 2. All explanations are for the high privacy setting. The low transparency explanation conveys no information about the privacy parameter.

<p>Low (Uber)</p>	<p>Differential privacy is a formal definition of privacy and is widely recognized by industry experts as providing strong and robust privacy assurances for individuals. In short, differential privacy allows general statistical analysis without revealing information about a particular individual in the data. Differential privacy provides an extra layer of protection against re-identification attacks as well as attacks using auxiliary data.</p>
<p>Medium (Apple)</p>	<p>Differential privacy transforms the information shared with us before it ever leaves the user’s device such that we can never reproduce the true data. The differential privacy technology used is rooted in the idea that statistical noise that is slightly biased can mask a user’s individual data before it is shared with us. If many people are submitting the same data, the noise that has been added can average out over large numbers of data points, and we can see meaningful information emerge. Our differential privacy implementation incorporates the concept of a privacy budget (quantified by the parameter epsilon). We use a privacy budget with epsilon of 9.</p>
<p>High (Ours)</p>	<p>The method will have changed your information before you send it. For most people, about 98 of the 200 YES or NO answers will have changed. The changes will be different for each person. We will never have access to the original table. We also will not know what parts of your browsing history were changed. Differential privacy changes each individual’s information. But since we collect information from many people, we can still see overall patterns that interest us.</p>

even smaller values, such as $\epsilon < 1$, for even stronger privacy guarantees. The privacy offered in the low privacy setting, however, is much worse than what would be considered reasonable in practice—in fact, even the parameter value for the medium privacy setting offers relatively weak privacy guarantees [314]. This extreme value for the low privacy setting makes it clear that differential privacy in this case offers only privacy theater rather than any meaningful privacy protection.

The second factor is transparency. The explanations provide high, medium, and low transparency into the privacy parameter (Table 2.4). The high transparency explanation was the outcome-focused explanation from Experiment 1, as it directly conveyed the implications of the privacy parameter for this data. The medium and low transparency explanations are drawn from

real industry explanations of differential privacy. The low-transparency explanation is an excerpt from an Uber blog post⁴. This explanation does not mention the privacy parameter. The medium transparency explanation is one that provides information on the privacy parameter, but in a way that reveals little useful information to users. This explanation is an adapted excerpt from a larger document in which Apple explains its implementation of local differential privacy [69]. Interestingly, this document gives the exact ϵ values for Apple’s implementations of differential privacy; however, the document does not provide much context to help readers interpret the numbers. Both explanations were found to be effective in persuading users to share sensitive information in prior work [367].

The explanations were roughly matched for word count (60-108 words). But rather than controlling for reading level or content, we left the industry language intact. Both use technical language and jargon, which may be an intentional feature. Some scholars have noted that jargon-laden privacy policies actively discourage users from engaging with the information they contain [76, 253]; it is possible that the use of jargon in explanations of differential privacy could play a similar role. A text that leans heavily on technical jargon may give the sense that there exists “*some sort of crypto-magic to protect people from data misuse*” without actually providing an explanation that is accessible for the average user [272]. Similarly, the Uber explanation makes an appeal to authority—referencing “industry experts”—which may be designed to exploit authority bias [160].

2.3.4 Measures

Both experiments used the same measures for dependent variables: perceived understanding, actual understanding, and willingness to share. Similarly, both included measures for covariates including: demographics, general privacy concerns, sensitivity and internet browser usage.

⁴Uber actually uses central differential privacy rather than local differential privacy, but the selected excerpt applies equally well to either model.

Understanding

Drawing on prior work [367], we distinguish between a subjective component of understanding (perceived understanding) and an objective component (actual understanding).

For perceived understanding, we measure how well participants felt that they understood an explanation of differential privacy. We adapt three questions from [225] for measuring perceived comprehension of a text and a fourth adapted from [97]. All four questions use 6-pt semantic scales.

For actual understanding, we measure how well participants actually understood an explanation by asking them concrete questions about differential privacy. Two questions (#1 and #2 in Table 2.5) are drawn from [367]. Two additional questions (#3 and #4 in Table 2.5) are related to the privacy parameter, since our explanations were developed with the specific goal of conveying information about the privacy parameter.

In the second survey, some participants only see low-transparency explanations that do not convey the necessary information to answer these questions. Therefore, one of the answer options was “I do not have enough information to answer this question.” Prior work has shown that survey takers gravitate toward these choices [181], rather than thinking carefully about whether the relevant information is present [180]. To deal with this issue, in the second experiment, participants who select the “not enough information” option were prompted to make their best guess; prior work has found this strategy to be useful [345]. Participants who view low-transparency explanations get the question correct if they select the “not enough information” option; their best guess does not factor into their comprehension score. Participants who view high-transparency explanations get the question correct if they select the correct answer on the first try or if they select “not enough information” but subsequently select the correct answer when prompted to make their best guess.

After the launch of the first survey, it became clear that participants found our comprehension questions relatively difficult, so three additional, easier questions were added for the

second survey (#5-7 in Table 2.5).

Table 2.5. Comprehension Questions. All questions are multiple choice. Correct answers are in parentheses. For questions #3 and #4, the correct answer depends on the experimental condition.

1. If someone shares their information, will the researchers know with certainty which websites that person visited? (*No.*)

2. If someone shares their information with us, and an attacker steals the information, will the attacker know with certainty which websites that person visited? (*No.*)

3. If someone shares their information, the method will change how many of their true answers for the 200 websites of interest? (*One of the following: A few—for about 1 or 2 websites, their true answer will be changed. / Some—for about 50 websites, their true answer will be changed. / Many—for about 100 websites, their true answer will be changed. / I do not have enough information to answer this question.*)

4. If someone shares their information, how accurately will the researchers understand that person's individual internet usage from the information they share? (*One of the following: Very accurately [a pretty good guess] / Somewhat accurately [a bit better than a random guess] / Inaccurately [not much better than a random guess] / I do not have enough information to answer this question.*)

5. For those who choose to share their information, what kind of privacy protection will we use to protect that information? (*Local differential privacy*)

6. How does the method protect an individual's privacy? (*Changing some of their answers*)

7. How does the method help the researchers collecting the data? (*They can see meaningful information emerge from large datasets.*)

Willingness To Share

The survey asks participants whether they would be willing to share their browsing history data before the privacy protection is described. This provides a baseline for participants' willingness to share and helps pinpoint the actual effect of the explanation of privacy protection; we would not want to infer that a particular explanation has persuaded a participant to share information when in fact this participant would have been perfectly happy to share the information without any privacy protection. Therefore, it is useful to ask about willingness to share twice, both before and after participants read the explanation of privacy protections.

Prior work has frequently asked about hypothetical willingness to share information

rather than observing willingness to share directly [45, 367]. However, hypothetical willingness to share is an imperfect proxy for actual disclosure behavior; previous work has found both that individuals tend to underestimate their actual willingness to share sensitive information and that risk perceptions are more strongly related to behavioral intention (i.e. hypothetical willingness to share) than to actual disclosure behavior [240]. In other words, it is often the case that people do not actually act in accordance with their risk perceptions. As a result, it can be difficult to shift user behavior. Since we are most interested in actual behavior, we measure willingness to share by actually asking participants to share their browser histories with us. If participants agree to share this data, the study shows an animation of a “progress bar” that suggests the browser history data is being uploaded. However, no browser history information is collected, and participants are debriefed about this deception at the end of the study.

Of course, it does not make sense to ask participants to share their data twice. Therefore, the first ask is hypothetical and only in the second ask are participants actually asked to share their information.

Privacy Concerns

An individual’s decision about whether or not to share data will be informed by the concerns that this individual holds regarding information privacy. We use the Internet Users’ Information Privacy Concerns scale to measure privacy concern [205]. In addition to the ten questions for this scale, we add two additional related questions—adapted from previous work—that are relevant to our setting; we ask how frequently participants falsify personal information online [251, 205] and whether or not they ever try to hide their online activities from others [259]. The ordering of these two added questions is randomized. All questions use a 5-pt scale.

Sensitivity

Since information sensitivity can affect users’ willingness to share data [367], participants are asked to rate the sensitivity of the browsing history information we request from them. With

wording adapted from [281], participants rate the perceived sensitivity of the requested browsing history information. Participants also rate the harm that could result from this information being leaked. Both questions use a 5-pt scale and are averaged to produce the measure of sensitivity.

Browser Usage

When asked to share internet browsing data, people who regularly browse the Internet may behave differently than people who use Internet browsers less frequently. Therefore, participants are asked about their frequency of browser use.

2.3.5 Recruitment

For both surveys, English-speaking participants over the age of 18 were recruited through Qualtrics' paid recruitment service to be approximately representative of the United States population in terms of educational attainment, age, race/ethnicity, and gender. The surveys were also hosted on Qualtrics. Participants who failed the attention check questions—described in Section 2.3.6—were screened out of the survey. Some participants' responses were deleted for speeding (first survey: < 220 seconds, second survey: < 240 seconds), straightlining, or entering gibberish in the free response textbox. Finally, to achieve this representative sample, some participants were screened out after entering their demographic information. This resulted in 365 total participants for the first experiment and 308 for the second experiment. For these participants, the median completion time was approximately 7 minutes (7m 24s) for the first survey and 9 minutes (9m 7.5s) for the second survey. Participants who completed either survey were paid by Qualtrics. The first survey was conducted in March 2021 and the second in April 2021. Appendix A contains the full participant demographics.

2.3.6 Experimental Protocol

Both experiments use a between-subjects design. Participants begin the survey by answering demographic questions. Next, the survey explains the general concept of internet

browsing histories, providing a brief description for participants with low technology literacy. Then participants are prompted to reflect for a moment on their browser usage and answer the sensitivity questions. It is by design that participants reflect on the sensitivity of this information before they are asked to share it; only at the end of the sensitivity questions do we ask for the first time whether the participant would be willing to share their information.

We then explain the differentially private mechanism, with participants assigned at random to one of the four experimental conditions for the first experiment and to one of the nine conditions for the second experiment. For the second experiment, a timing mechanism prevents participants from advancing to the next section until they have spent at least one minute reading through the explanation. Participants are also required to click several times in order to continue reading; this was intended to slow participants down and encourage them to read carefully.

Next, four questions serve as an attention check in order to screen out participants who have not actually read the explanation. The questions are designed to be easy to answer, and the participants are able to reread the explanation in order to answer the questions; these choices aim to prevent screening out participants who have read the explanation carefully but may struggle with reading or have limited working memory capacity. Participants who fail to answer the first two questions correctly are given a second chance and presented with the remaining two questions. If these questions are not answered correctly, participants are screened out of the study.

Participants who pass this check proceed by completing items for the dependent measures: perceived understanding, actual understanding, and willingness to share. Participants can re-read the explanation as necessary while answering these questions. If participants report being willing to share their browsing history, an animation suggests this information is being uploaded automatically. Participants also answer an open-ended question about why they were or were not willing to share their data. Finally, participants complete questions related to the remaining covariates: general privacy attitudes and internet usage. Before they submit the survey, participants are given access to a debriefing document and informed that no browsing history

data was actually collected. At the completion of the survey, resources are provided so that participants can learn about best privacy practices, such as clearing browsing histories.

2.4 Analysis

2.4.1 Testing Hypotheses

To test our hypotheses for the first experiment, we used ANCOVA to compare the objective comprehension scores for the four treatment groups, controlling for education. Since only two participants declined to provide their educational background, these participants were dropped from this analysis. To test our hypotheses for the second experiment, we fit a logistic regression model that models willingness to share as a function of 1) the choice of explanation, 2) the choice of privacy parameter ϵ , 3) the interaction between these two variables, 4) perceived sensitivity, and 5) the privacy concerns measure. The privacy parameter was treated as an ordinal variable. Three participants who declined to answer questions for the covariates were dropped from this analysis. We fit one regression model on the full set of responses for the second survey and one regression model on a subset of responses, excluding participants who had been willing to share their data even before reading the description of differential privacy. The goal of this second regression was to test if our hypotheses would hold for these more privacy-conscious participants who were not willing to share their data without privacy protection.

2.4.2 Comparing Comprehension

We compared participants' performance on two closely related comprehension questions (#1 and #2 in Table 2.5)—one asks whether the researchers will be uncertain about participants' true answers while the other asks the same question about an external attacker. The question about the external attacker is more aligned with how discussions about security and privacy are traditionally framed. The question about whether the researchers will know the true responses is trickier—this idea of protecting privacy from the same people with whom they are sharing data

will be less familiar to participants. We compare performance on the two questions using the two proportion z-test; the responses from the two surveys are analyzed together.

2.4.3 Qualitative Analysis

To analyze the qualitative data from participants about why they chose (not) to share their information, two researchers used an iterative open coding approach [35]. The researchers performed two rounds of coding. First, both researchers familiarized themselves with the responses from the first survey. Next, the researchers agreed upon a set of codes. When the responses came in from the second survey, the researchers again began by familiarizing themselves with the data, and they decided that some additional codes should be added. Finally, all responses from both surveys were analyzed together and assigned codes; multiple codes could be assigned to a single response. Disagreements were remedied through discussion until agreement was reached. The list of codes can be found in Appendix A.

2.5 Results

The analysis of our results revealed no statistically significant effects. Interestingly, we found that most participants had made up their minds about whether or not to share their data before they even read about the privacy protection; therefore, these explanations had little effect on participants' behavior. Below we discuss the results in detail for both experiments and for the qualitative analysis.

2.5.1 Experiment 1: Developing Good Explanations of Differential Privacy

Participants were overconfident in their understanding of differential privacy; the scores for perceived understanding were substantially higher than their actual understanding (Figure 2.1). Participants may not have spent enough time reading the explanations carefully; many people are used to scrolling through privacy policies and clicking accept without actually reading

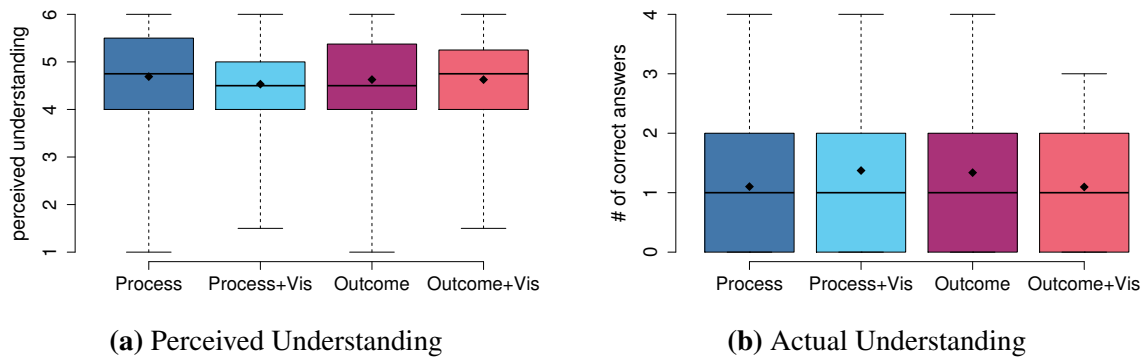


Figure 2.1. Perceived and actual understanding for Experiment 1. Most participants felt that they understood the explanations well (left), yet answered no more than one question correctly (right). Performance was similar across all four conditions.

them [242, 115, 215]. Participants who took more than 7 minutes to complete the survey (roughly 50% of participants) answered 0.47 more questions correctly — over a 10 percentage point improvement — compared to those who finished more quickly. To address this issue, experiment 2 set a minimum time for reading the explanation.

As hypothesized, the outcome-focused explanation outperformed the process-focused explanation; the average percentage of comprehension questions answered correctly was greater for outcome (33%) than process (28%) focused explanations. However, the difference was not significant. Figure 2.1b shows the distribution of the number of questions answered correctly for each condition.

Adding a visual component did not uniformly improve participants' understanding. The process-focused visual may have been helpful for participants. It slightly improved the average percentage of comprehension questions answered correctly (35%), but the outcome-focused visual decreased this performance (27%). Again, the differences were not statistically significant.

As none of the differences were statistically significant (ANCOVA comparing all four explanations gave $p = 0.1$), we selected the outcome explanation for the second experiment. Of all four explanations, process+vis resulted in the highest average comprehension score; however, the differences were small, and a text-based explanation would serve the fairest comparison for

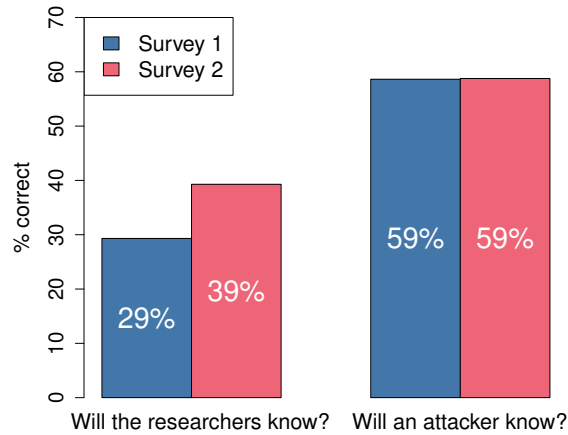


Figure 2.2. Performance on paired questions.

the text-only industry explanations.

Two paired comprehension questions identified one challenge for explaining differential privacy (Figure 2.2). Most participants correctly identified that attackers who stole survey data from the researchers would be unable to determine any participant’s true browsing history with certainty. However, most participants incorrectly believed that the researchers would have access to participants’ true browsing histories. Participants are likely more familiar with privacy tools to protect from hackers than with tools that allow for sharing information while preserving privacy. The difference in performance on these two questions is statistically significant ($p < 10^{-19}$). It may be counterintuitive that the information they have agreed to share will be obfuscated even for the researchers they agreed to share it with.

There are several factors that may have caused the poor performance on the comprehension questions. The fact that perceived understanding was high suggests that the readability of the explanations was not the problem, although participants may have found some of the comprehension questions confusing. As discussed previously, a lack of attentive reading is likely one reason for the low comprehension. Most people are habituated to skimming through privacy policies or clicking “accept” without reading [242]. Therefore, even though our explanations were designed to be much more accessible than the typical privacy policy, people may have

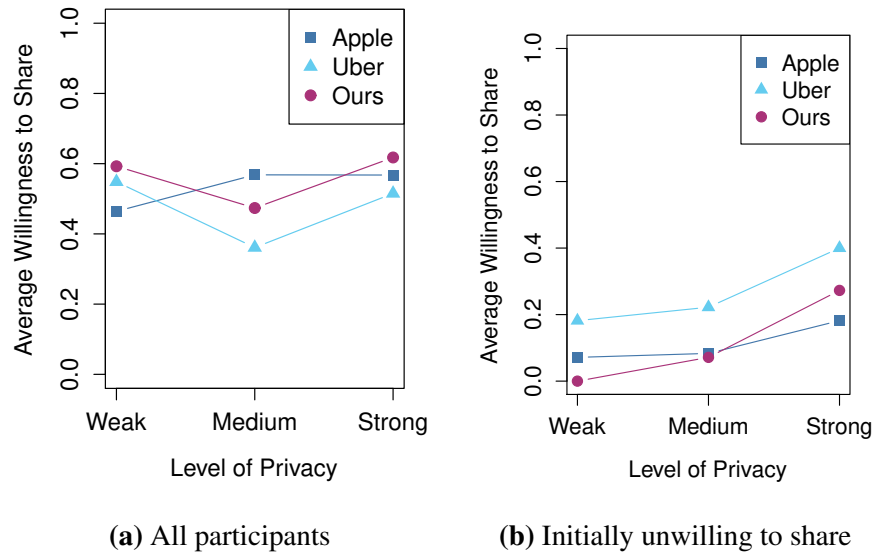


Figure 2.3. Willingness to share for the three explanations and three settings of the privacy parameter. These figures show the averages for all participants (left) and participants who were initially unwilling to share their information (right).

skimmed our explanations out of habit. Our results also indicate that most participants simply were not very concerned about their data privacy in this context—yet another reason they may not have paid close attention to the explanations. Another factor may be that certain aspects of differential privacy are simply counterintuitive. For example, participants may find it strange to think about hiding information from the same people with whom they are sharing data—this would explain the poor performance on the question about whether the researchers would know participants’ true information (Figure 2.2). Finally, two of the questions (#3 and #4 in Table 2.5) were expected to be more difficult, since they required participants to reason about probabilities.

2.5.2 Experiment 2: Measuring Privacy Theater

Contrary to our expectations, the choice of privacy parameter ϵ and choice of explanation had no significant effect on willingness to share ($p = 0.1 - 0.99$). Most participants made up their minds about whether or not to share their data before reading about the privacy protection.

The survey asks participants about their willingness to share twice. Early in the survey, participants are asked about their willingness to share their browsing histories. Later, after

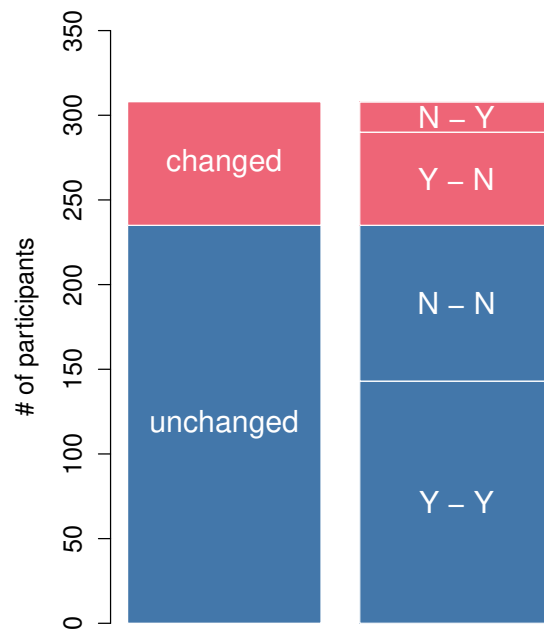


Figure 2.4. Willingness to share before and after differential privacy explanation. Most participants’ answers did not change (left) when asked the second time.

explaining the protection offered by differential privacy, participants are asked whether they agree to share this information via an automated upload process. Most participants (76%) do not change their minds after reading about the privacy protection (Figure 2.4).

The open-ended responses (see Section 2.5.3) provide some context for this finding. Many participants are willing to share their data because they trust researchers and want to help or because they do not consider browsing data particularly sensitive; for these participants, the promises of privacy protection are irrelevant. The set of participants for whom the privacy protection really mattered was relatively small—only 36% of participants were unwilling to share their data when first asked, even before any privacy protection was mentioned. Of those initially unwilling to share, many had strong negative reactions to being asked to share this information or found it too sensitive to share. The set of participants for whom the privacy protection made the biggest difference were those who changed their minds after reading the differential privacy explanation. This group was quite small (n=73).

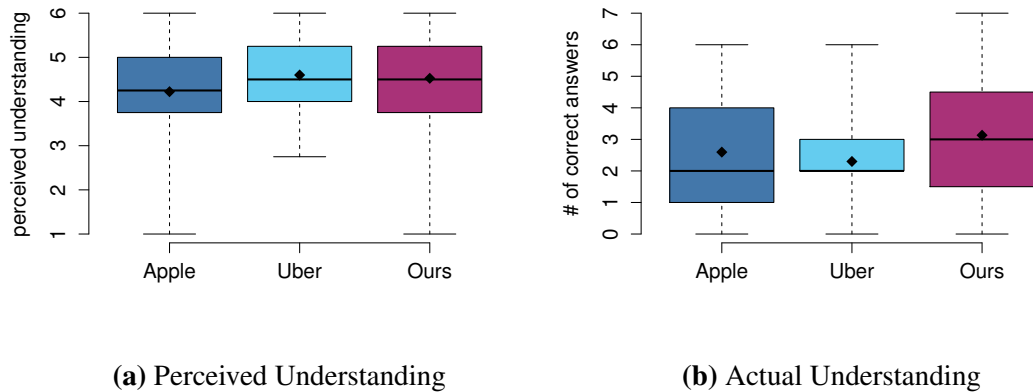


Figure 2.5. Perceived and actual understanding for Experiment 2. As in Experiment 1, participants’ perceived understanding (left) is greater than their actual understanding (right), though the added questions improved comprehension scores compared to Experiment 1.

Although we did not find a statistically significant relationship between the variables of interest, we noticed an intriguing result. In the high-transparency, weak-privacy setting, no one agreed to share, unless they had already agreed to share before they read any description of privacy protection. This suggested that for privacy-conscious participants, H4 may actually hold. In other words, our explanation may be effective in helping privacy-conscious users make more appropriate data-sharing decisions. However, we did not have enough data to discern whether this pattern was meaningful, since so few participants were initially unwilling to share data (only nine for this condition). To investigate this pattern further, we conducted a larger replication study. While the overall rates of (un)willingness to share remained consistent, this study failed to provide evidence to support our hypotheses. The details about this study are included in Appendix A. Although our failure to find a significant effect of explanation on sharing behavior seems to contradict some prior work [367], a more recent paper found that descriptions of differential privacy had a significant effect on privacy expectations yet not on willingness to share. The authors suggest that effective explanations may need to be tailored to the concerns of individual users [57]. Our results, however, show that it may be difficult to shift people’s preconceived feelings about what information should or should not be shared.

2.5.3 Qualitative Results

Both surveys asked participants to explain why they decided to share or not to share their information. This qualitative data helped clarify why some participants chose to share their information or not, as well as why some changed their minds after reading about the privacy protection offered by local differential privacy.

Reasons for Sharing and Withholding Information

Most participants did not change their willingness to share based on the explanation of privacy protections. This participant made it clear that the protection offered by local differential privacy was irrelevant to their decision: *No matter how it is shared, it does not seem like a good idea* [P588]. Other participants listed factors such as distrust of the researchers or the survey platform (n=37), fears of getting “scammed” or “hacked” (n=27), general discomfort (n=40), or sensitivity of the requested information (n=56). For example, two participants mentioned concerns about others seeing pornographic websites that they had visited.

Participants who were willing to share their information even before learning about the privacy protections and followed through in sharing listed a variety of reasons for doing so. Many participants—mainly in the first survey—mentioned concerns about being able to complete the survey and receive compensation (n=33), for example, writing that, *I need money badly so it's worth the weird risk* [P129] and *I really need the money from this survey to help take care of my family until pay day. I really hope its not a scam* [P192]. As a result, the second survey clarified that participants would be able to complete the survey and receive payment regardless of whether or not they agreed to share browsing data. These responses point to a larger issue; all too often, the choice of whether to share information is hardly a choice at all [203, 88].

Another common reason that participants agreed to share their information was that they trusted the researchers or the survey platform (n=46). In other words, these participants agreed to share their information because they trusted the data recipients. This finding is consistent with prior work that has shown that the reputation of the data recipient can be an important factor in

data sharing decisions [18].

Other participants who were willing to share their information listed reasons such as a desire to help the researchers (n=66), curiosity (n=31), and a sense that they had “nothing to hide” (n=82). For example, one participant wrote: *Research and study is important! I have mostly nothing to hide* [P125]. Another commonly expressed theme was indifference (n=27); when asked why they decided to share their data, many responded with another question—why not?

An interesting theme that arose in many of the responses was a concern about the scope of data collection more broadly (n=12). Some participants felt that so much data was being collected about them already that they might as well share more: *I decided to share because this information is already out of my hands given how my ISP has free access, might as well contribute to a study while i'm at it* [P30]. On the other hand, others expressed a desire to hang on to what they perceived as the last little bit of privacy that they had left:

I refuse to share this information because all of my information is already online, theres no need to go through my PRIVATE browsing history. My phone is my phone, and sorry but no one has a right to that. Privacy doesnt exist in this world anymore, the least I could do is keep my phone away from the world [P229].

It is interesting that both participants who shared their information and participants who declined to share cited these same kinds of concerns when asked to justify their choices. While some participants felt the need to exert extra effort to protect their privacy in a world of ubiquitous data collection, others expressed feelings of resignation and fatalism [76].

Perceptions of the privacy protection

Although most participants did not mention the protection provided by differential privacy in their responses, some participants did feel comforted by this privacy protection and trusted it to protect their information (n=40): *It seems like the process you're using will protect my information* [P294]. However, some participants felt suspicious and did not trust that the mechanism would protect their data (n=11): *The process didn't seem safe. I'm not sure what*

information would be changed. This seems like a hoax [P332]⁵. Other participants seemed concerned by the description of the “automatic upload” process. For example, one participant who initially expressed being willing to share information later declined to share, explaining that: *An upload of my information is a little daunting* [P199]. These participants were more concerned about a mysterious script running on their computer than about sharing browsing data. Finally, some participants remained confused after reading the explanation of the privacy protection (n=14). Nevertheless, most participants rated their own understanding of the explanations highly (Figure 2.1 and Figure 2.5).

2.6 Limitations

Our study has several limitations, some of which point toward directions for future work. The first limitation is that since most participants were perfectly willing to share their data, we can not draw firm conclusions about the behavior of individuals who were actually concerned about privacy. In order to better understand how people behave when concerned about their privacy, future work might try to explicitly target privacy-conscious individuals—for example, the topic of privacy could be emphasized in recruitment materials.

A second limitation concerns our qualitative results. Although the responses to the open-ended survey question provided valuable insight, most responses were quite short—a few sentences at most. A more thorough understanding of individuals’ thought processes might be obtained through interviews or laboratory studies. These kinds of studies might also shed light on any misconceptions that participants may have about differential privacy.

A third limitation of this work is that behavior in a research study may differ significantly from behavior in everyday life. For example, we asked participants to read the description of differential privacy carefully. In everyday life, when asked to share their data, many people will not read the fine print. Furthermore, the participants knew that they were sharing their data with researchers. Some participants specifically mentioned this fact as a reason that they trusted us

⁵This participant was assigned to the outcome+vis condition.

and were willing to share their data. People may behave differently when asked to share data with corporate (rather than academic) entities. Future work might study situations in which people encounter differential privacy “in the wild.”

2.7 Discussion

2.7.1 False Choices

Although we did not intend to force participants to choose between their privacy and their paycheck, some participants did feel compelled to make this choice due to a lack of clarity in the first survey. Several participants explained that despite their fears, they felt that they had no real choice, since they desperately needed the money. These responses offer a glimpse into the myriad of ways that marginalized people in particular are asked to surrender their privacy. Prior work has examined situations in which people living in poverty are forced to sacrifice privacy in order to access resources [88]. Studies that limit their focus to “*the often-studied White, American, middle-class subject*” will miss important aspects of the way that privacy and surveillance operate [212].

2.7.2 Digital Resignation

Several participants expressed concerns about the ubiquity of digital surveillance. Some of these participants opted to share their data, because they felt that trying to protect their privacy was a losing battle. Prior work has observed similar phenomena, sometimes referred to as “*privacy cynicism*”, “*surveillance realism*”, or “*digital resignation*” [76, 140, 65]. In fact, feelings of helplessness regarding surveillance appear to be widespread; a recent survey found that more than 60% of Americans agree that “it is not possible to go through daily life without being tracked” by companies or by the government [22].

Awareness of digital resignation should inform the design of privacy-enhancing technologies; for example, tools that increase awareness of online tracking without actually offering

improved control over one’s data may unintentionally lead to increased resignation rather than increased engagement with privacy-related issues [52, 326]. More work is needed to better understand the causes of digital resignation and to propose solutions for overcoming these feelings of futility, inevitability, and hopelessness. There may be particular strategies that help people regain a sense of agency. For example, in the context of environmental activism, Sarah Jaquette Ray argues that “*the perception that social change happens only on an individual scale creates defeatism*” and that “*if we see ourselves working collectively rather than individually, we can rest assured that we are contributing to a larger web of movement*” [263]. Perhaps collective strategies for resisting surveillance (e.g. obfuscation on Instagram [237]) and PETs that support such strategies (e.g. AdNauseam [143]) could be an effective way to combat digital resignation.

2.7.3 Making Sense of Privacy Promises

The results of our experiments suggest that for many people, factors such as the purpose of data collection or a (dis)trust of data collectors are more important than promises of privacy protection when deciding whether or not to share information. If people trust the data collector and if people expect to share in the benefits of data collection, they may feel that differential privacy is unnecessary. Perhaps discussions about optimal privacy-utility tradeoffs have at times overshadowed equally important discussions about the distribution of utility—in other words, discussions about who benefits from data analysis.

Even though explanations of differential privacy seem to only matter to a small set of participants, for companies with millions of users, even small effects matter. Future work could help clarify both 1) when users care about promises to protect privacy and 2) how users respond to such promises. As more tech companies begin to tout their privacy protection features in advertising campaigns, these questions will become increasingly important [250].

Our results revealed that some aspects of differential privacy are counterintuitive to users and may be difficult to convey even in well-crafted explanations. For example, users do not expect to need to protect themselves from the people they are sharing their data with.

In practice, industry explanations may not be well-crafted; in fact, they may be intentionally difficult to understand [76]. And even within a single industry (e.g., web browsers) there are a wide variety of technical approaches. As a result, people searching for useful information about ways to protect their privacy may grow frustrated with the “*widespread obfuscatory communication practices used by companies across the digital-media landscape*” [76] and struggle to understand which privacy promises are actually meaningful [153]. In trying to make sense of this landscape, people turn to journalists, to family, to friends, and to other trusted sources [257, 266]. Understanding how people engage within communities to learn and make sense of this information could help establish important avenues for intervention. And given the special opportunity that journalists have to help readers make sense of information about privacy that often lies buried in unintelligible privacy policies [89], the way that journalism shapes privacy practices is also worthy of further study.

More broadly, many questions remain about the rhetorical functions of privacy-enhancing technologies. Companies have a variety of avenues for communicating promises of privacy to their users—through privacy policies, app notifications, press releases, blog posts, and more. Understanding how this messaging shapes individual perceptions of technology companies is important. But it is worth asking not only how PETs and their associated messaging affect users’ interactions with particular apps or companies but also how they shape larger conversations around online privacy [124, 142, 143, 277]. The term “*privacy*” does not have a single, straightforward agreed-upon meaning [20, 295]. Thus, in addition to whatever technical affordances they provide, PETs may make implicit claims about the very meaning and value of privacy. For example, AdNauseam—a browser extension that hides ads while employing obfuscation mechanisms to sabotage advertising surveillance—serves both a practical, protective purpose and an expressive purpose; AdNauseam actively promotes the conception of privacy as a collective good [143].

2.7.4 Other Approaches for Addressing Privacy Theater

The problem of privacy theater is not unique to differential privacy [292, 92, 93, 137]. As a consequence, addressing the problem is likely to require an array of partial solutions. Carefully developed explanations can be one part, and prior work has discussed other approaches [82, 152]; still, more work is needed. While it is worthwhile to develop accessible explanations of privacy technologies, many people who are accustomed to the opaque, jargon-laden style of privacy policies may simply skim or skip such explanations out of habit. Furthermore, the burden of ensuring that the privacy protection offered by a particular company is adequate should not rest with individuals. Many researchers have noted that by offering control to users, technology firms can shift burdens from themselves to individual users [334, 316, 17]. Instead, given the power asymmetries that exist, policymakers and others should reflect on societal goals around privacy protection. In addition, there may be technical approaches like expert audits and other accountability mechanisms that can help drive organizational responses [367, 82].

2.8 Conclusion

In this work, we explore how people decided whether or not to share their data when offered the protection of differential privacy. We develop explanations of differential privacy for non-experts, but find that certain aspects of differential privacy remain challenging for people to understand. We also test whether explanations that offer little transparency into the role of the privacy parameter can be misused to trick people into sharing more data than they otherwise would. However, we find that most people are perfectly willing to share their browsing data, regardless of whether they are offered the protection of differential privacy; other factors seem to dominate decision-making around the sharing of browsing data. We point to the need for future work in order to better understand the effects of differential privacy and other privacy-enhancing technologies.

2.9 Acknowledgements

MAS was supported by a Qualcomm Fellowship. We also thank the anonymous reviewers for their helpful comments, suggestions, and insights. This chapter is a reprint of the material as it appears in *Proceedings of the ACM on Human-Computer Interaction*. M. A. Smart, D. Sood, and K. Vaccaro. Understanding risks of privacy theater with differential privacy. *Proc. ACM Hum.-Comput. Interact.*, 6 (CSCW2), 2022. The dissertation author was the primary author of this paper.

Chapter 3

Models Matter: Setting Accurate Privacy Expectations for Local and Central Differential Privacy

3.1 Introduction

Usable privacy research has long focused on making information about privacy protections transparent to end users in order to allow for informed decision-making about data sharing [195, 95, 84]. Prior work has sought to explain privacy enhancing technologies (PETs) such as end-to-end encryption in messaging [63] and HTTPS/TLS [94]. Existing explanation strategies vary in terms of how much they try to explain data protection mechanisms or processes versus implications of the particular PET. As PETs become increasingly complex and offer more nuanced notions of privacy, there is a significant need for increased research into best practice for transparent PET messaging that matches these new techniques and empowers end users.

Differential privacy (DP) [81] is a privacy-enhancing technology that has quickly been adopted by industry and government agencies [86, 221, 317, 4, 226, 66]. DP deployments provide provable privacy guarantees by adding statistical noise to data or computations; this noise obfuscates the information of each individual while preserving aggregate-level insights. In response to DP's rapid success, a growing body of work has started to document the inadequacies of existing messaging around DP [57] and design new messaging techniques for DP systems [367,

368, 163, 235, 45, 287, 101].

The technical and mathematical complexity of PETs like DP makes effective communication challenging [57, 233, 6, 63]. DP is an interesting case study for constructing transparent PET messaging because it is an instance of an emerging PET paradigm that has received relatively little attention: privacy-preserving, outsourced computation. This paradigm is increasingly important as more users rely on computationally-weak mobile devices [252]. In response to this growing need, new approaches to privacy-preserving computational outsourcing are being actively developed, both in industry [16, 224] and academia [186]. Although there has been some work on people’s mental models of other PETs in this category [161] and creating transparent messaging for functional encryption in particular [10], further work is needed. DP is particularly interesting because of its widespread deployment and use of statistical methods (as opposed to encryption).

In this work, we develop messaging for DP that highlights the threat models that are implicit in different approaches to deploying DP. These implicit threat models are critical for end users to understand before sharing their data, as the chosen threat model may not provide protections against the classes of attackers about which they are concerned. We do this by exploring three explanation formats drawn from the existing PETs messaging literature: nutrition labels [167], diagrams [368, 287, 306], and metaphors [372, 63, 260, 163, 373, 311]. Each of our evaluated explanations aims to communicate the consequences of DP in terms of which information flows the deployment protects against. Prior work has tended to focus on explaining individual PETs in isolation, and while we focus on DP, we discuss how our designs may be modified to account for the effects of multiple PETs deployed together.

Differential Privacy Deployment Models. There are multiple deployment *models* for DP, each of which is associated with a particular threat model. The two most widely deployed models are the *central model* [81] and the *local model* [164].¹

The central model assumes there exists a data curator who is *trusted* to see raw data from

¹Although there are many variations of DP [67], we focus on the *central* and *local* models due to their popularity.

individuals; the adversary can only access released, aggregate results. The data curator collects data from individuals, performs statistical analyses on the collected dataset, and then injects statistical noise into the results before release. This process limits the ability of the adversary to reconstruct individual records from summary statistics, at the cost of reduced accuracy. The potential danger of this model, however, is that the data curator might *not* be trustworthy; the database storing individuals' data could be vulnerable to hackers or misuse by insiders if other, complementary security practices are not adopted in tandem. Well-known deployments of the central model include the U.S. Census Bureau's data products for the 2020 Decennial Census [4].

In the local model, noise is added to each individual's data *before* collection, meaning the unmodified data are never stored together. As such, there is not the same need to trust the data curator (i.e., it is assumed that the data curator is honest but curious). This higher level of security comes at a cost: significantly more noise must be added to the data in order to ensure the same level of privacy protection, reducing the accuracy—and, thus, utility—of the collected data. Notably, Google and Apple have both used local DP to analyze browser data in Chrome and Safari, respectively [86, 69].

Helping Users Understand DP Models. Ensuring that descriptions of DP accurately convey information about the model is crucial to designing transparent messaging for DP deployments: the threat surface associated with the two main models differ significantly, even if they provide the same privacy guarantees for the *data releases*. Specifically, data collected under the central model can be hacked, leaked, or abused by an insider threat if sufficient complementary privacy measures are not taken, while data collected under the local model does not share these risks.

Data subjects cannot be expected to make informed data-sharing decisions if they believe that DP is “some sort of crypto-magic to protect people from data misuse” [272]. Prior work has demonstrated that existing DP description strategies do a poor job aligning users' privacy expectations with the privacy protections provided by DP models [57]. In other words, the kind of protection that users expect does not align with the actual nature of the protection offered

by DP. Misaligned expectations exacerbated by poor communication can lead to data subjects underestimating or—even more alarmingly—overestimating the privacy protection that DP offers.

The goal of our work is to find effective ways to incorporate information about the model into descriptions of DP deployments. More specifically, we use a mixed-methods approach to study the following research question: **What are effective design strategies for explanations that help people understand which information flows are protected by DP, given a deployment model?**

We first explore three kinds of explanations that build on prior work [163, 236, 368, 167, 260, 57] through an interview study: metaphors, diagrams, and privacy labels for DP. Based on the results of this study, we identify the most promising strategies—privacy labels and metaphors—and further refine these explanations based on participant feedback. We evaluate our refined explanations in an online survey ($n = 698$), measuring objective comprehension, subjective understanding, perceived thoroughness, and trust. We compare our explanations against existing state-of-the-art text-based explanations of DP [367]. Based on our results, we make suggestions for future research and design of explanations of PETs.

The process of investigating design strategies also allowed us to characterize mental models people form around DP. While studying mental models was not the primary goal of our study, we include particularly interesting insights that can guide future work. For example, we found that participants often tried to make sense of DP through comparisons to other PETs such as encryption. We conclude with a discussion of the potential design implications of our findings.

3.2 Background

A growing body of work provides guidance for effective security and privacy (S&P) communication [280, 115]. Awkward interfaces or ineffective communication can lead to dangerous misconceptions and risky behaviors [359, 179, 57, 125]. One reason that people may

misjudge privacy risks or misuse PETs is that they lack appropriate *mental models*. Prior work has argued that “efficacy of risk communication depends not only on the nature of the risk, but also on the alignment between the conceptual model embedded in the risk communication and the user’s mental model of the risk” [21]. Unfortunately, existing depictions of DP appear to be misaligned with people’s mental models, resulting in misaligned privacy expectations [57].

In this section, we outline the relevant prior work on the challenges of designing effective, transparent communication about PETs. First, we discuss the prior work on communicating with data subjects about DP. Next, we discuss three particularly popular privacy explanation strategies—metaphors, diagrams, and nutrition labels. Finally, we discuss prior work on mental models in S&P.

Implications vs Process. One of the most important findings from prior work is that explaining data protection processes is not enough for most readers to grasp the implications of the protection offered by PETs [287, 57, 367, 72]. Xiong et al. [367] studied explanations of both central and local DP, and found that when the implications of the local and central models were stated explicitly, participants were more willing to share information under the local model. Kührtreiber et al. [182] replicated this study with German participants. Differently from these studies, we explore a variety of best-practice methods from the usable S&P literature (i.e., metaphors, diagrams, and nutrition labels) to communicate which information flows are protected by DP under the central and local models.

Cummings et al. [57] explored the implications of DP through six information disclosures about which people care and DP may protect—depending on whether the local or central model is used. They found that existing descriptions of DP fail to appropriately set privacy expectations regarding these disclosures, in part because many descriptions are not specific to the model (e.g., central or local) being used. In contrast to this work, we build new explanations rather than evaluating existing ones. We draw on their framework to present the implications of DP (entities to whom data can potentially be disclosed) as part of our nutrition labels. For improved clarity,

Table 3.1. Five Information Disclosures. We combine the “organization” and “data analyst” categories from prior work, since a data analyst is simply an employee [57]. Although some implementations of the central model limit employees’ access to the data (e.g., Uber [157]), we consider the more common case where only published information is privacy-protected.

Information Disclosure	Local	Central
Hack: <i>A criminal or foreign government that hacks the non-profit could learn my medical history.</i>	False	True
Law: <i>A law enforcement organization could access my medical history with a court order requesting this data from the non-profit.</i>	False	True
Org: <i>An employee working for the non-profit, such as a data analyst, could be able to see my exact medical history.</i>	False	True
Graph: <i>Graphs or informational charts created using information given to the non-profit could reveal my medical history.</i>	False	False
Share: <i>Data that the non-profit shares with other organizations doing medical research could reveal my medical history.</i>	False	True

in our study, we combine two of the disclosures (organization and data analyst), resulting in five total (Table 3.1).

Finally, Frazen et al. [101] and Nanayakkara et al. [235] developed methods of explaining the implications of the privacy budget, drawing from the risk communication literature. Nanayakkara et al. [235] found that participants were more willing to share information as the privacy budget decreased (i.e., protections were strengthened). In our study, we assume a small privacy budget (i.e., strong privacy), so that we can focus on the implications of the deployment model. Future work could consider combining our explanations with explanations of the privacy budget.

Metaphors Metaphors are one approach for improving mental models, and have been studied extensively in the S&P domain [372, 63, 260, 163, 373, 311]. For example, physical security metaphors can improve users’ understanding of personal firewalls [260]. In other cases, however, metaphors have been less effective. For example, descriptions of end-to-end encryption using metaphors failed to improve understanding [63]. Prior work has also begun to explore the effectiveness of metaphors specifically for explaining DP [163]. They find that functional metaphors can be useful for explaining both that injected randomness protects privacy and

that there exists a tradeoff between privacy and accuracy. The metaphors we develop are also functional (i.e., focused on *what* DP offers), rather than structural (i.e. focused on *how* DP works) [11]. While the metaphors from prior work aim to cover a long list of facts about DP, they are not designed to emphasize the different kinds of disclosures against which DP may or may not protect—the focus of our work.

Diagrams Another strategy for explaining PETs is the use of visualizations. For example, hypothetical outcome plots [145] have been used to visualize the protection offered by DP [287, 249]; they have also been used to visualize DP’s accuracy implications for data curators [234]. In the case of randomized response [352]—a simple instantiation of local DP—the injected noise can be represented through a spinner [45, 62]. Recent work has also explored the use of diagrams and animations in the specific context of location privacy [368]. Diagrams have also been used to explain other PETs such as encryption [306]. We build on this prior work to develop diagrams for DP in both the local and central models.

Nutrition Labels One influential approach in privacy communication broadly has been the use of “nutrition labels” for privacy [167]. Drawing inspiration from standardized nutrition labels on food products, privacy labels have been proposed as an alternative or supplement to typical privacy policies with their notorious usability issues [242, 327]. Privacy labels have proven to be a useful way to present privacy-related information [167]. Organizing key information into carefully-designed labels helps users find information more quickly than they would by perusing a traditional privacy policy [168]. Although originally proposed for websites, similar labels have since been developed for datasets [141] and Internet of Things devices [84]. Nutrition labels have even been proposed for describing DP [366, 57]. Apple has recently integrated the nutrition labels approach into their iOS app ecosystem. Unfortunately, the utility of these labels has been hampered by the fact that labels are not always easy to find and can be misleading or inaccurate [55, 178]. Our work adapts the privacy label concept for the purpose of explaining DP—specifically for explaining how local and central DP may or may not protect against particular disclosure risks.

Mental Models.

Mental models refer to simplified versions of complex processes that people mentally hold and which may help them understand key pieces of information [54, 194]. Researchers have argued that mental models are important for effectively communicating security risks to end users [305]. Camp [47] argues that a medical or public health mental model is particularly useful for conveying the implications of malicious code—in particular, “that everyone is at risk,” “the importance and continued autonomy in the face of risk” and the “shared responsibility for community health.” In this way, mental models can rely on people’s existing knowledge to help them better grasp attributes of a new setting.

However, flawed mental models can lead to dangerous decisions [340, 354]. For example, Wash [354] proposes folk models of viruses and hackers and describes how these models help explain why people ignore security advice. A mental models approach can also clarify how people’s backgrounds may impact their understanding of risks [162, 241, 40]. For instance, people’s level of computer science background impacts the complexity of their internet mental models, and therefore the number of privacy threats they perceive [162]. Oates et al. [241] find that when asked to create illustrations of the meaning of privacy, experts’ illustrations tend to depict privacy as more “nuanced” than non-experts’ illustrations. Bravo-Lillo et al. [40] also find that novice and advanced users have different mental models and risk perceptions.

Finally, researchers have noted the value of studying privacy expectations [194, 261]. For example, Lin et al. [194] propose evaluating mobile app privacy by studying people’s privacy expectations of apps, while Rao et al. [261] suggest that understanding misalignment’s between people’s expectations and privacy policies can help reduce privacy risks.

3.3 Interview Study

We began designing explanations by developing a set of initial prototypes, drawing from prior work in S&P communication. Through an interview study,² we use these prototypes to

²All studies were approved by the Human Research Protection Office.

solicit feedback on what makes an effective explanation of DP.

Scenario

We situate our designs within the medical data collection scenario from [57]. In this scenario, a non-profit organization is collecting health data for medical research. Because medical information is considered highly sensitive [149, 281], data subjects are more likely to care about understanding the privacy implications of DP in this scenario.

3.3.1 Initial Prototypes

Metaphors can help non-experts develop more useful mental models. We develop four initial metaphors—two for the local model and two for the central model—designed to clarify the kinds of risk involved. All four can be found in Appendix B.4.

We also draw inspiration from prior work on visualizations of DP [45, 234, 249, 287, 236, 235] to design our own diagrams that highlight how DP protects or fails to protect against the disclosures listed in Table 3.1. We developed our diagrams through an iterative process. We discussed the accuracy and clarity of initial diagrams as a group, and based on the discussion, iterated on our designs. In the end, we developed four diagrams—two for the local model and two for the central model—with slight differences in iconography. After initial interviews, we added a third variation for both models that included a caption. All diagrams used a vertical line to depict the “privacy barrier,” as in [236], and used icons—most selected from the Noun Project³—to represent the different kinds of disclosures. Instead of using an illustration of a database, as in [236, 368], we use an icon of a filing cabinet to represent the collected data. Representative diagrams can be found in Appendix B.4.

Following guidance from prior work, we also developed privacy labels to clearly demonstrate which kinds of information disclosures DP can protect against. Each row corresponds to a specific information disclosure and clarifies whether protection is offered against said disclosure. We tested three different versions of the tables (six distinct tables in total, across the two models).

³<https://thenounproject.com>

One version of the table listed only the disclosures against which DP can protect. Thus for the local model, this table had five rows, whereas for the central model, this version had only one row. This table uses a red circle-backslash symbol to indicate that a particular disclosure is not permitted. The other two versions always included information about all five disclosures, but used different iconography to depict protection or lack thereof. Both of these versions incorporated lock icons to indicate when DP protected against a particular kind of disclosure. In one of these tables, we use a green lock icon—as recommended in prior work on connection security icons [95]—to indicate safety, whereas a red unlocked icon indicates disclosures against which DP does not protect. The other table is in black-and-white and uses the presence or absent of a lock icon to indicate (lack of) protection. Chrome previously used lock icons to indicate connection security, but has recently backed away from this choice due to concerns about overtrust; some Chrome users incorrectly assumed that a lock icon was a reflection on the safety of the website itself rather than the connection [51, 201]. Varying the use of icons allowed us to evaluate their appropriateness in a DP context. Appendix B.4 includes representative versions of our original privacy labels.

3.3.2 Protocol

We used a 3 x 2 study design: each participant evaluated either the metaphors, diagrams, or privacy labels for either the local or central model. Our goal was to solicit feedback to help us iterate on our designs of each type. All interviews began by describing the same hypothetical scenario:

A non-profit organization is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. The non-profit would like to explain to people how they will protect patients' privacy.

Next, participants are informed that the non-profit plans to “use an extra layer of privacy protection in order to protect patients’ medical information.” Then, they are shown the first explanation of this privacy protection. After reading the explanation, the participants are asked to explain how patient data will be protected in their own words, as in [116]. Next, they are

asked how they feel about the explanation, how well they feel that they understand the privacy protection after reading the explanation, what concerns they would have about sharing their data, and what else they would like to know about how patient data will be protected, adapting questions from [267]. If the design under discussion includes the use of color, they are also asked about these color choices. Finally, they are asked how the explanation could be improved.

Next, participants are shown an alternate version of the explanation of the same type, still describing the same model (i.e., local or central). They are asked if the new explanation has changed their understanding. Then, they are asked the same questions they were asked about the original explanation. Some participants were then shown a third version—since we had three versions of the privacy labels and added a third version of the diagrams—and the above questions were repeated. We vary the order of explanations shown between participants. After viewing all explanation versions, participants are asked which one would be most useful for patients deciding whether to share their data. Finally, participants are asked how they would explain to patients how their data would be protected.

Participants who viewed the privacy label explanations or metaphor explanations were then asked to draw a diagram that conveyed their understanding of how patient data would be protected. Participants who struggled to draw on their screens could choose to tell the interviewer what to draw. The purpose of these drawings is two-fold. The drawings serve both as a way to clarify participants' mental models and as a source of inspiration for iterating on our own designs. The participants who viewed the diagram explanations were not asked to do any drawing, since they would be heavily biased towards the diagrams they had already been shown. Finally, in concluding the interview, participants were asked to self-report gender, race, and ethnicity. Additional demographic information was provided through the recruitment platform.

3.3.3 Participant Recruitment

The first author interviewed 24 U.S. residents recruited through Prolific. We wanted our explanations to be broadly accessible, so we used filters to ensure that at least half of participants

had no college degree. A breakdown of participant demographics can be found in Appendix B.3. Participants whose interviews included drawing a diagram were paid \$15, whereas participants who evaluated the diagrams were paid \$12 since these interviews were shorter. Interviews lasted about 10-30 minutes and were conducted over Zoom.

3.3.4 Analysis

The interviewer first transcribed and summarized all the interviews. Next, an interview from each condition was selected at random, forming a set of six interviews. The first two authors reviewed these six transcripts to develop a set of codes, organized into four distinct themes (Appendix B.1). They then shared this codebook with the research team and modified it based on the group's feedback. Finally, the same two authors coded all 24 interviews together⁴ with the updated codebook.

3.3.5 Findings

Effectiveness of Initial Designs

While we found some strategies more effective than others, across all conditions, participants had additional questions that our explanations did not answer.

Metaphors.

Responses to the metaphors were mixed. Some participants appreciated the concision of the metaphors, while others wanted more details. For example, one participant criticized an explanation's brevity, saying it is "*a little bit simple and [...] doesn't go into too many details.*" (P8) In contrast, a different participant complimented this very quality by describing an explanation as "*reader-friendly, very concise*" (P5). This tension between accuracy and thoroughness of explanations on the one hand, and simplicity on the other has also been reported

⁴We do not calculate or report inter-rater reliability (IRR) for two reasons. One, while calculating IRR can be useful to establish agreement before researchers divide a corpus to code different subsets individually, in our case both researchers coded all of the data together. Two, we are not seeking to make quantitative claims about our codes [217].

in other domains, such as explainable machine learning [3] and privacy policies [115].

Explanations that make use of metaphor can help people develop useful mental models, and, conversely, people’s use of metaphor can reveal their own understanding. Participants across all conditions provided a range of metaphors conveying their understanding of DP, some of which could be adapted as explanations of DP. For example, a participant in the metaphor condition explained that after their data passed through the privacy barrier, they would be like a ghost, no longer identifiable. Another participant in the metaphor condition explained the obfuscation applied in the local model as follows:

I have long hair, but you don’t know what color it is. You don’t know that I have contacts and not glasses, so you wouldn’t be able to pick me out of a lineup, is what I would imagine it as. (P4)

These metaphors of ghosts and lineups both hint at the idea of DP as a form of anonymization.

This same participant provided another particularly creative metaphor:

It’s kind of like an egg. You know, you crack it open and you don’t know if it’s going to be rotten inside or not. But I don’t know what chicken it came from, so I can’t blame the chicken. (P4)

The phrase “can’t blame the chicken” seems to convey the protection offered by DP as a form of plausible deniability.

Design Changes: We replaced our original metaphors with a new metaphor inspired by those generated by participants. Synthesizing metaphors related to hiding or changing one’s appearance—like not being recognizable in a lineup or becoming a ghost—we developed a new metaphor: this metaphor compares protecting data with DP to wearing a “disguise.”

Diagrams.

Of all the explanation methods, the diagrams were the least successful. Of the eight participants assigned to this condition, five explicitly expressed that the diagram was confusing. Although the other three participants did not explicitly use the term “confusing,” they also struggled to understand various aspects of the diagrams. For example, when asked to explain

the privacy protection in their own words, one participant started to try to explain, then cut themselves off and responded: “*Well, I don’t really know*” (P23).

A number of participants expressed confusion or disagreement with the underlying threat model, particularly for the central model. The central model only prevents disclosure from published reports. Although responsible data collectors will employ other technologies such as encryption to protect against hackers or criminals, DP in itself does not protect against this kind of disclosure in the case of the central model. For some participants, this was counterintuitive. For example, after viewing a diagram explaining the central model, one participant expressed their confusion as follows:

I don’t really get it. [...] There’s supposed to be a barrier between my medical information and the people who read the published reports. It seems. And then people who want your data seems like that’s open and free, and it seems backwards to me. (P24)

Two other participants viewing diagrams for the central model incorrectly stated that the privacy barrier was protecting data from hackers, even though the diagrams showed hackers on the left side of the privacy barrier (i.e., the same side as the data collection). One of these participants realized their mistake later in the interview. First, they explained:

The privacy barrier [...] allows the people who utilize the information, say the law enforcement and medical professionals, [...] to share that information amongst themselves on a secure in a secure network without allowing the people who want to get that information to abuse that information, the hackers. (P22)

However, a bit later, they realized their mistake:

I’m looking at it again. It says well the people want that data, it’s just letting them take it, it looks like. So I guess that would kind of be a concern there [...] we’re letting the scientists and the policymakers, the scientists, the people who need to see maybe medical data not allowing them to see the data. But it has a backdoor that allows the people who want to steal that information. So it really has a flaw. (P22)

Despite the fact that the diagrams showed the hacker to the left of the privacy barrier, two of the four participants in this condition nevertheless explicitly stated that the privacy barrier would

protect their data from hackers. Many people may expect PETs to protect against hackers and criminals, making the protection offered by central DP alone somewhat unintuitive [287]. We dropped the diagram explanations due to the pervasive confusion expressed by participants.

Xiong et al. [368] previously investigated the use of diagrams for explaining location privacy and found less than ideal levels of comprehension, particularly for the local model, though they speculate that data quality issues with Amazon Mechanical Turk may be to blame. Alternatively, it is possible that data flow diagrams inherently overemphasize *processes* at the expense of clearly enumerating *implications*.

Design Changes: The diagram explanations were dropped, due to persistent confusion.

Privacy Labels.

Responses to the privacy labels were largely positive, though not universally so. Participants praised the privacy labels for their simplicity and clarity. In addition, several participants appreciated the use of color. For example, one participant explained that: “*Having the colored icons does make it a bit faster for a person to get the message*” (P16). However, participants did not always agree about the meaning of the colors green and red. On the one hand, green is often associated with safety while red is associated with danger. Given these associations, one might use green to indicate protection and red to indicate vulnerability. On the other hand, green is also used to mean “go” whereas red means “stop.” Given these associations, one might use red to indicate protection, since the flow of data is “stopped.” Some participants felt that our use of green and red should be switched, while others felt that our use was appropriate.

Design Changes: To ameliorate the confusion with red and green, we eliminated red and chose to highlight protection in green. The rest of the content was black.

Importance of Process.

All of our explanations were designed to communicate the *implications* of DP rather than the details of *how* DP works. Prior work has shown that explaining the process of adding

noise to data is not enough to help people understand the consequences data sharing [367]. Nevertheless, omitting any discussion of process seems to leave people unsatisfied and confused. Most participants had questions about how the data protection worked. Providing a detailed explanation of the mathematical and technical details of DP is likely to overwhelm most people, but people nevertheless do want some information about how DP works—finding the right balance may be challenging. This finding aligns with prior work on explaining encryption. While explanations of encryption focused on *outcome* lead to greater perceived security than explanations focused on *process*, hybrid explanations that incorporate information on both process and outcome lead to the greatest perceived security [72]. Prior work on metaphors for DP also found that some participants were interested in understanding how DP works [163].

Design Changes: Another text was added to provide context about how DP works; we adapted a state-of-the-art text explanation by Xiong et al. [367], while aiming for improved readability by eliminating terms like “database” and “aggregated.” We anticipated that this additional information on *process* could complement our other explanations that focus on *implications*.

Mental Models

Our interviews reveal a number of different mental models that participants constructed to understand DP, based on the explanations they were shown. In many cases, participants’ mental models were informed by their prior knowledge of and experience with other technologies.

Comparison to other PETs.

Some participants—especially in the diagram and privacy label conditions—reasoned about DP through comparisons to other PETs. For example, one participant understood the privacy barrier as “*some kind of firewall that keeps [their] privacy safe*” (P21). Encryption in particular was mentioned frequently, perhaps because it is a particularly familiar and ubiquitous PET or perhaps because participants associated our lock icons with encryption [95, 135]. One

participant, who assumed that encryption was the technology being described, wanted to know “*what type of encryption*” (P23) was used. Prior work has also found associations between DP and encryption, and found that associations with encryption correspond to higher trust [163]. DP is distinct from encryption, so while it may be possible to leverage people’s knowledge about encryption to construct better explanations of DP, it is also likely that associations with encryption may lead to misconceptions.

One particular source of confusion is that with encryption, the protection offered should be binary—information is either encrypted (i.e., protected) or not. This corresponds nicely with the physical metaphor of a lock that has exactly two states: locked and unlocked. In the case of DP, however, the goal is to allow some information “leakage” while still offering some protection—the amount of information leakage depends on the the privacy budget parameter. Although many participants liked the lock icons, other participants pointed out this issue. For example, one participant in the diagram condition said:

If you’re releasing some form of my information to these published reports, it’s not completely locked. (P20)

Thus, the use of lock icons and their association with encryption may in some cases prove problematic.

Design Changes: We designed an additional version of the privacy labels that uses arrows to indicate whether data flows are permitted or blocked instead of locks. This version uses red to denote flows that are blocked.

DP as anonymization.

Several participants understood DP as an anonymization technique—especially those who read the metaphor explanations. These participants often had an overly-simplistic view of DP. For example, one participant explained that in their understanding, the data “*would be protected by virtue of being anonymized and not including the patient’s name, social security number, or date of birth*” (P13). Of course, DP provides better guarantees than such a naive

anonymization strategy; nevertheless, this mental model may provide a useful approximation of practical DP guarantees.

DP as fake data.

A few participants understood DP as the injection of fake data. One participant explained it as follows:

You're collecting my name, but it's a fake one, so it's like a shield up in front of me. (P4)

Once again, while this model oversimplifies DP, it shares key elements with the truth and thus is likely useful overall. However, it is important for people to understand that the “fake” data nonetheless reveals useful information about the overall distribution; therefore, DP does not necessarily offer protection against inferential privacy risks [171, 172].

Validating Design Changes.

We recruited 10 additional participants through Prolific to pilot our updated explanations. These participants were shown explanations of various types—including the two privacy labels and various texts that evolved somewhat over the course of the interviews—and asked to build their own explanation by editing or combining existing explanations or creating their own from scratch (Figure B.1). Participants expressed more satisfaction and few substantive edits as compared to our initial evaluations, however they suggested a wide range of ways to combine the texts and privacy labels. No singular combination was preferred by several participants. Therefore, in our quantitative evaluation, we test not only the texts and privacy labels alone but also these explanations in combination with each other as further detailed in Section 3.4.

Further, prior to launching the quantitative evaluation of our designs, we compared our two privacy labels in a survey using the evaluative criteria outlined in Section 3.4. We found no significant differences between the two versions on any of the evaluation criteria. We chose to continue with the version with arrows instead of the version with locks for a few reasons. One participant expressed their preference for the version with arrows over the one with locks as

follows:

I felt better seeing the same people being blocked rather than the lock because you see those everywhere nowadays. (P28)

In other words, the lock symbol has become so ubiquitous that this participant found it meaningless. We also felt that the arrows more clearly showed that certain information disclosures were protected against while others were not, whereas lock icons might suggest that certain people are given a “key.” This is a fundamentally different kind of protection since keys can be leaked or shared. Finally, although the difference was not significant, comprehension scores were slightly better for the version with arrows. Thus, we dropped the version with locks. The evolution of our designs is visualized in Figure B.4.

Design Changes: We dropped the label with locks in favor of the version that emphasized information flows.

3.4 Large-Scale Evaluation

We conducted an online survey in March 2023 to evaluate the impact of our explanations on understanding and to assess their efficacy in setting appropriate privacy expectations.

Protocol.

Respondents are first asked to read the scenario description (the same medical scenario discussed in Section 3.3.2) and the description of how data will be protected. Next, respondents answer a simple, multiple-choice comprehension question to ensure that they have read the scenario description. They are given the option to re-read the description. If they do not answer correctly, they are given a second attempt, in accordance with Prolific’s policies. If after the second attempt, they again answer incorrectly, they are prevented from advancing further in the survey.

Respondents who pass the comprehension check are then asked whether they trust the non-profit to protect patient privacy [367]. Next they are asked two questions related to self-efficacy. Finally, they are asked whether they would be willing to share their information with

the non-profit. An open-text box asks them to explain their decision.

Respondents then answer five true/false questions on privacy expectations, followed by the Likert-scale questions about understanding and thoroughness. They are also invited to share feedback on the explanations in a free-response text box. When answering the above questions, respondents have the option to reread the descriptions of the scenario and privacy protection at any time. Next, respondents are asked about their familiarity with various PETs, including DP and a non-existent technology (“deliquescent security”). If they indicate familiarity with some of the listed technologies, they are asked which of the technologies (if any) was described in the survey. In a free-response text box, they are asked to explain their reasoning. Finally, respondents answer questions about themselves. In addition to standard demographic questions (i.e., age, income, race, ethnicity, gender, education, job field), the survey also includes measures of internet skill [127]. The full survey instrument is included in Appendix B.2.



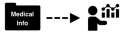



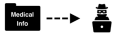
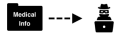












Who Can See Your Data	Without Privacy Protection	With Privacy Protection	Who Can See Your Data	Without Privacy Protection	With Privacy Protection
Viewers of graphs or informational charts created using information given to the non-profit...	 ...might be able to see your information.	 ...will not be able to see your information.	Viewers of graphs or informational charts created using information given to the non-profit...	 ...might be able to see your information.	 ...will not be able to see your information.
Hackers—like criminals or foreign governments— who successfully attack the non-profit...	 ...might be able to see your information.	 ...will not be able to see your information.	Hackers—like criminals or foreign governments— who successfully attack the non-profit...	 ...might be able to see your information.	 ...might be able to see your information.
Law enforcement with a court order requesting your information from the non-profit...	 ...might be able to see your information.	 ...will not be able to see your information.	Law enforcement with a court order requesting your information from the non-profit...	 ...might be able to see your information.	 ...might be able to see your information.
Employees of the non-profit, such as data analysts, who work with the non-profit's data...	 ...might be able to see your information.	 ...will not be able to see your information.	Employees of the non-profit, such as data analysts, who work with the non-profit's data...	 ...might be able to see your information.	 ...might be able to see your information.
Organizations collaborating with the non-profit that are given access to the non-profit's data...	 ...might be able to see your information.	 ...will not be able to see your information.	Organizations collaborating with the non-profit that are given access to the non-profit's data...	 ...might be able to see your information.	 ...might be able to see your information.

Figure 3.1. The final version of our privacy labels for the local (left) and central (right) models.

Experimental Conditions.

We use an 8 x 2 experimental design (all conditions listed in Appendix B.4), varying both the explanation type and the deployment model—local or central. We evaluate the privacy labels (Figure 3.1), process text, and metaphor text individually as well as in combination: metaphor + process, metaphor + process + label, metaphor + label, and process + label. We compare

Table 3.2. Explanation texts for the local and central models.

Type	Local	Central
Process	To protect your information, your data will be randomly modified before it is sent to the organization. Only the modified version will be stored, so that your exact data is never collected by the organization.	To protect your information, the organization will store your data but only publish reports, graphs, or charts that have been randomly modified. These modifications hide information that is unique to you as an individual.
Metaphor	The technology works something like this: Your data will be disguised before it is stored by the organization. Therefore, anyone who accesses the data collection will only see this disguised version of your data.	The technology works something like this: The collected data will be disguised when any graphs, charts, or reports are published. However, anyone who accesses the organization’s data collection will see the undisguised data.
Xiong et al.	To respect your personal information privacy and ensure best user experience, the data shared with the non-profit organization will be processed via an additional privacy technique. That is, your data will be randomly modified before it is sent to the organization. Since the organization stores only the modified version of your personal information, your privacy is protected even if the organization’s database is compromised.	To respect your personal information privacy and ensure best user experience, the data shared with the non-profit organization will be processed via an additional privacy technique. That is, the organization will store your data but only publish the aggregated statistics with modification so that your personal information cannot be learned. However, your personal information may be leaked if the organization’s database is compromised.

these seven conditions against the implication-focused explanations from [367]. Although [367] evaluated a number of different explanations, we chose to compare against the explanation that led to the highest comprehension of privacy protections. Table 3.2 provides all three text explanations.

Dependent Measures.

Our goal is to set privacy expectations appropriately. Thus, we use a series of true/false questions from prior work about whether certain types of disclosure are possible to measure *objective comprehension* (Table 3.1). We also ask respondents about their *subjective understanding*

of the explanations and how *thorough* they perceive the explanations to be [167]. Additionally, we ask whether respondents *trust* the non-profit organization to protect patient privacy [367], and we ask two questions related to *self-efficacy* in decision making [235]. All five questions use 5-pt Likert scales. Finally, although we do ask about *willingness to share* data with the non-profit, we caution against using willingness to share as a measure of explanation quality. The explanation that convinces the most people to share their data is not necessarily the best explanation. For example, we hope that a patient who is particularly concerned about disclosure to law enforcement would choose *not* to share data when it is protected using the central model.

Participant Recruitment.

698 total respondents were recruited through Prolific, using the “balanced sample” feature—in accordance with best practices—to recruit an approximately representative sample in terms of gender [315]. We conducted a power analysis to estimate an appropriate sample size; due to the large number of experimental conditions, we lack the statistical power to detect very small effects, but such small effects are unlikely to be meaningful in real-world contexts [268]. Respondents were paid \$2 for completing the survey, and the median completion time was just under six minutes. A detailed breakdown of respondent demographics can be found in Appendix B.3.

Analysis.

We analyze⁵ the effect of our explanations on our dependent measures. We construct a set of regression models studying the effect of our independent variables—explanation and model—on our dependent variables: objective comprehension, subjective understanding, perceived thoroughness, trust, self-efficacy, and data-sharing decision. We use logistic regression models to study data-sharing decisions, linear regression models to study comprehension, and ordinal regression models to study the remaining dependent variables. In all models, we control for internet skill [129]. We use the following convention in reporting statistically significant results:

⁵Analysis code: https://osf.io/3acvw/?view_only=f12174861ffd4cd0872a54a8e1326a26

* denotes $p < 0.05$; ** denotes $p < 0.01$; and *** denotes $p < 0.001$.

We also perform a qualitative analysis of the responses to two of the free-response questions. The first author reviewed all the reasons respondents gave for their data-sharing decisions and developed a set of codes. The first and second authors then reviewed the codebook together and coded 30 responses, resolving disagreements through discussion and refining the codebook as necessary. Then they separately coded 25 responses and evaluated inter-rater agreement by calculating Cohen's Kappa—the average across all codes appearing in this sample was 0.75, indicating substantial agreement. Remaining responses were divided between both authors for coding. After finding the privacy labels to be most effective, the first author additionally reviewed all feedback provided for the privacy labels and developed a second set of codes—despite some overlap in the themes discussed in the feedback responses and the data-sharing decision responses, the content was sufficiently distinct to merit separate codebooks. Again, the first two authors reviewed the codebook and coded 10 responses together, resolving disagreements through discussion. Then they separately coded 25 responses and calculated Cohen's Kappa, with an average of 0.98 across all codes appearing in the sample, indicating near perfect agreement. The remaining 393 responses from participants in any of the privacy label conditions were divided between both authors for coding. For both sets, multiple codes could be applied to a single response. Both sets of codes are available in Appendix B.1.

3.4.1 Results

Effectiveness of Designs

We find some explanations are more effective than others in terms of objective comprehension and trust.

Comprehension.

Across all explanations, we find a significant difference in objective comprehension (Table 3.3) between the local and central model—the local model is associated with fewer correct

Table 3.3. *Left:* results from linear regression models for objective comprehension. We report regression coefficients (β) and 95% CIs for these coefficients. $\beta > 0$ indicates an increase. *Right:* results from ordinal regression models for subjective understanding. We report odds ratios (OR) and corresponding 95% CIs. An OR > 1 indicates an increase in odds. For both columns, we use the Xiong et al. explanation as the reference level explanation.

Variable	<i>Objective Comprehension</i>	<i>Subjective Understanding</i>
	β	OR
Model: Local	-1.17*** [-1.42, -0.92]	0.79 [0.61, 1.03]
Expl: Metaphor	0.09 [-0.45, 0.62]	1.63 [0.91, 2.90]
Expl: Process	-0.33 [-0.86, 0.19]	1.39 [0.79, 2.45]
Expl: Process+Metaphor	0.47 [-0.06, 1]	1.79* [1.01, 3.18]
Expl: Arrow Label	1.15*** [0.62, 1.68]	1.35 [0.75, 2.42]
Expl: Label+Metaphor	1.26*** [0.73, 1.8]	1.51 [0.86, 2.67]
Expl: Label+Process	0.97*** [0.45, 1.5]	1.39 [0.79, 2.44]
Expl: Label+Process+Metaphor	0.95*** [0.43, 1.48]	1.48 [0.84, 2.6]
Internet Skill	0.20** [0.05, 0.35]	1.36*** [1.16, 1.6]

answers ($\beta = -1.17$; $p < 0.01$). This finding is consistent with prior work which suggests that privacy expectations are more closely aligned with the central model than with the local model [57, 368]. It may be difficult to realign a reader’s understanding if they come in with strong expectations that do not match the actual protection offered by DP. Compared to the Xiong et al. explanation, all of the explanations that include a privacy label are associated with more correct answers ($\beta = 0.95$ – 1.26 ; $p < 0.01$). The text-only explanations, on the other hand, showed no significant improvement over the Xiong et al. explanation. This improvement is expected since the privacy labels are designed explicitly to highlight the information flows that respondents are asked about in the comprehension questions. Interestingly, there

Table 3.4. Results from regression models for trust, perceived thoroughness, and self-efficacy, with the Xiong et al. explanation as the reference level explanation. Again we report odds ratios with 95% CIs. An OR > 1 indicates an increase in odds.

Variable	<i>Trust</i>	<i>Thoroughness</i>	<i>SE (Info)</i>	<i>SE (Confidence)</i>
	OR	OR	OR	OR
Model: Local	1.70*** [1.3, 2.23]	1.08 [0.83, 1.41]	0.87 [0.67, 1.13]	0.90 [0.69, 1.17]
Expl: Metaphor	1.46 [0.82, 2.6]	1.28 [0.72, 2.27]	1.21 [0.67, 2.19]	1.65 [0.91, 2.99]
Expl: Process	0.99 [0.56, 1.73]	0.81 [0.46, 1.43]	0.70 [0.39, 1.23]	0.95 [0.53, 1.68]
Expl: Process+Metaphor	1.45 [0.81, 2.58]	1.55 [0.88, 2.74]	0.93 [0.52, 1.64]	1.19 [0.67, 2.11]
Expl: Arrow Label	0.85 [0.48, 1.52]	0.68 [0.38, 1.21]	1.15 [0.64, 2.05]	1.44 [0.8, 2.61]
Expl: Label+Metaphor	1.18 [0.66, 2.12]	1.73 [0.97, 3.09]	0.96 [0.53, 1.72]	1.16 [0.65, 2.06]
Expl: Label+Process	1.97* [1.12, 3.47]	1.22 [0.7, 2.15]	1.08 [0.62, 1.87]	1.06 [0.61, 1.87]
Expl: Label+Process+Metaphor	0.94 [0.54, 1.64]	1.38 [0.79, 2.41]	0.98 [0.55, 1.73]	1.07 [0.6, 1.9]
Internet Skill	0.96 [0.82, 1.13]	1.06 [0.9, 1.25]	1.21* [1.03, 1.41]	1.28** [1.09, 1.5]

is a misalignment between objective comprehension and subjective understanding. The process+metaphor explanation is the only one that significantly improves subjective understanding compared to the Xiong et al. baseline (OR = 1.79; $p < 0.05$), even though it does not improve objective comprehension. Indeed, objective comprehension and subjective understanding are not strongly correlated ($\tau = 0.085$; $p < 0.01$), though there is stronger correlation for the local model ($\tau = 0.159$; $p < 0.001$) than for the central model ($\tau = -0.0389$; $p > 0.1$). Prior work has found similar misalignment between objective comprehension and subjective understanding [287, 101]. Unsurprisingly, internet skill is also associated with higher objective comprehension ($\beta = 0.20$; $p < 0.01$) and subjective understanding (OR = 1.36; $p < 0.001$).

Other Evaluation Criteria.

Table 3.4 summarizes how the explanations compare on our other evaluation criteria. Although comprehension is better for the central model, trust is higher for the local model (OR = 1.97; $p < 0.05$). This is promising, since the local model does offer stronger privacy. The label + process explanation is also associated with greater trust. This aligns with the qualitative feedback from our interviews. While information about process is not enough to help readers understand implications, it seems that explanations that focus only on implications leave readers feeling skeptical. This result is consistent with prior work on explaining encryption that finds benefits of combining information on process and outcome [72]. There were no significant effects of model or explanation on perceived thoroughness or self-efficacy, although higher internet skill is associated with higher self-efficacy (OR = 1.21–1.28; $p < 0.05$).

Feedback.

As in our interview study, one of the most common themes in our respondents' feedback was a desire for more information about how the privacy protection works (n=76). For example, one respondent wrote:

It doesn't explain at all how this supposed "privacy protection" works, so how do I know if it's credible? I have a lot of cybersecurity training: I want technical details!

Even respondents who read the process text sometimes requested more information about data protection processes. Respondents also requested other kinds of additional information (n=28), for example, about the organization and how it would use their data. A tension was again evident between respondents who requested additional information and those who praised our concision or requested further simplification. One respondent suggested "*more detailed explanations of the privacy protections that are available to view if needed.*" This adaptive approach was also suggested in interviews.

Table 3.5. Familiarity with PETs

PET	#
End-to-end encryption	439
Differential privacy	32
Secure multi-party computation	
Deliquescent security (distractor)	
None of the above	237

Prior Familiarity with PETs

In interviews, we found that some participants understood DP through comparisons with other PETs. Of the PETs we mention in our survey, end-to-end encryption was by far the most familiar, whereas only a minority had heard of DP (Table 3.5). Of respondents who answered the question asking which technology was described in the survey, most correctly selected DP, though several respondents explained in their free-text responses that they were simply guessing.

Data-Sharing Decision

Respondents are more willing to share data (Table 3.6) under the local model (OR = 1.46; $p < 0.05$). This replicates findings from prior work and is likely due to the fact that the local model offers stronger privacy guarantees [367]. None of the explanations had a significant effect on data-sharing decisions.

When people are deciding whether to share information, they consider many other factors in addition to privacy protections [287, 232, 105]. In fact, many respondents simply were not worried about privacy (n=75). For example, one respondent felt that they had nothing in their medical history that they would “*need to hide or be particularly private about.*” Other respondents were interested in sharing their information, because they value helping others, mentioning benefits of data sharing (n=151). In the words of one respondent: “*I do not have a problem with sharing my records if it will help someone.*” On the other hand, respondents who were less willing to share their data often indicated that they felt it would be too risky or

Table 3.6. Results from regression model for data-sharing decision. We report odds ratios (OR) and corresponding 95% CIs. An OR > 1 indicates an increase in odds.

Variable	<i>Share</i>
	OR
Model: Local	1.46* [1.07, 2]
Expl: Metaphor	1.09 [0.57, 2.12]
Expl: Process	0.75 [0.38, 1.45]
Expl: Process+Metaphor	0.81 [0.41, 1.58]
Expl: Arrow Label	0.52 [0.26, 1.04]
Expl: Label+Metaphor	0.80 [0.4, 1.56]
Expl: Label+Process	1.33 [0.7, 2.56]
Expl: Label+Process+Metaphor	0.65 [0.33, 1.28]
Internet Skill	0.95 [0.79, 1.15]

that their medical information was simply too private (n=242). For example, one respondent explained they were “*not comfortable sharing [their] medical records with anyone but [their] doctor.*” Other respondents wanted more information before they would be willing to share their data (n=155). However, the information they requested was not always related to DP. For example, some respondents wanted to know more about the non-profit organization. Finally, some participants distrusted either the non-profit or the privacy protection (n=88). In the words of one respondent:

Companies say that your information is secure all the time, but all the time there are security breaches. I do not trust my private information to be secure with anyone.

Other respondents also mentioned the frequency of data breaches as a cause for concern (n=35).

3.5 Limitations

Our designs are limited in their focus on a single scenario. Although medical applications are often cited as motivation for studies of DP [159, 30], DP has not been widely deployed in medical contexts [58]. Nevertheless, our privacy labels are transportable to other domains. Future work could transfer our designs to other scenarios and test whether our findings still hold. A limitation of our evaluation is that encountering explanations of DP in practice differs significantly from encountering explanations in an online survey. Future work could investigate comprehension when these explanations are encountered in more natural settings. A third limitation is our focus on a U.S. audience. Our privacy labels may be received differently in a different cultural context. Finally, we present the nature of DP’s protection as binary, when in fact the level of protection depends on the choice of privacy budget. This simplification may be appropriate for small privacy budgets, but the question of determining an acceptable range for the privacy budget is itself a nontrivial problem.

One concern may be that our privacy labels are “teaching to the test,” since we design them specifically to highlight the information disclosures that we ask about to measure comprehension. Thus, it is not surprising that comprehension is higher for our privacy labels than for explanations designed with a different emphasis. However, if the purpose of an explanation is to inform readers about which information flows are restricted—i.e., if we are using the “right” test—perhaps teaching to the test is not such a problem. Nevertheless, we incorporate additional evaluation criteria from prior work and find that our privacy labels improve comprehension without sacrificing quality on these other metrics (Tables 2.5– 3.4).

3.6 Discussion

Our results highlight the value of combining disparate best practices from prior work on explaining other security and privacy (S&P) concepts to explain complex PETs such as DP [167, 72]. We find that consequences-focused explanations (i.e., privacy label explanations that

highlight information flows) to be a promising approach for promoting accurate understandings of potential data leaks in DP systems. However, to ensure that such explanations are trusted we find that it is necessary to pair such consequences-focused information with a limited amount of high-level information about mechanisms: how DP works to offer particular consequences and protections. Below we discuss potential pitfalls of privacy labels for DP as well as ways to extend our designs to explain other PETs individually or in combination.

Potential Pitfalls.

Although the nutrition label approach shows promise for setting appropriate privacy expectations, it is important to avoid pitfalls from prior deployments of nutrition labels for privacy [55]. For example, iOS privacy labels can be misleading and inaccurate [178], in part because developers struggle to create accurate labels [190]. Similarly, our labels for DP could be misleading if an organization has implemented DP incorrectly [155, 32, 48, 229, 200] or has chosen an inappropriately large privacy budget [82]. Specialized programming platforms, audits, and formal verification approaches are therefore an important complement to our work [219, 270, 371, 323, 71, 173], namely by ensuring that the communicated privacy guarantees match the implementation.

Furthermore, while privacy labels can empower individuals to make decisions that better align with their goals and values, it is also important not to overburden individuals in the same way that traditional privacy policies do [215]. As some of the participants we interviewed highlighted, it can be difficult to strike the right balance between simplicity and comprehensiveness. Such a balance is important not only for data subjects, but also for other audiences who may encounter DP. For instance, privacy labels for DP could be used to educate policymakers, advocacy organizations, or software developers to support them in various decision-making processes. For example, Mozilla’s “privacy not included” guide offers expert reviews to help buyers choose products that provide strong privacy and security, since it can be difficult for individual buyers to evaluate various data protection policies themselves. One could imagine a similar project to provide reviews

for different data collection initiatives. An advocacy organization might use privacy labels for PETs like DP to identify and recommend certain initiatives that provide good S&P guarantees.

Finally, it is crucial that privacy labels for DP be contextual. While the information disclosures our explanations highlight are ones that people care about [57], they represent a starting point which should be used to further adapt explanations for specific contexts. The information disclosures we highlight may not be comprehensive of all specific disclosures people are concerned about across contexts. For example, privacy concerns in a particular educational setting may differ from a medical setting. Future work should also study ways to supplement privacy labels for DP with contextually-appropriate communication about the choice of privacy budget [235, 27].

Privacy Labels for Other PETs.

Our approach to designing privacy labels for DP could be adapted to other PETs. We hypothesize that privacy labels that take a contextual integrity approach—emphasizing which data flows are permitted and which are prohibited—could lead to improved comprehension of a variety of PETs [239]. Our survey respondents found it more difficult to reason about the implications of local DP than central DP. This finding suggests that clearly explaining which data flows are permitted is particularly important for PETs that enable outsourced computation, such as local DP. Future work could confirm whether the techniques employed here, and the greater difficulty with mental model formation among participants, extends to other PETs that engage in outsourced computation, such as secure multi-party computation, trusted execution environments, and homomorphic encryption

Our findings suggest that people employ their known models of PETs (e.g., understandings of encryption) to reason about new PETs. A standardized approach for presenting the kinds of protection a particular PET offers could help people compare new PETs with more familiar ones. Leveraging this kind of prior knowledge could be beneficial; however, we also caution that in some cases, drawing on knowledge of other PETs could lead to confusion or overtrust. It is

important that comparisons between PETs clearly explain their differences and do not overstate the protection offered.

Finally, PETs are rarely deployed in isolation. Our qualitative data show that people are interested in learning about DP *in context*. That is, they want information about the protection offered by DP, but they also care about the other safeguards and signals of trustworthiness that might help them make better-informed holistic data-sharing decisions. Particularly in the case of the central model, users may feel more comfortable if information about DP is presented alongside information about other PETs used to secure user data. Future work should go beyond explaining PETs one at a time and study effective ways to explain the nature of the protection obtained through combinations of PETs. Since our privacy labels focus on information flows—rather than the details of how DP works—it should be straightforward to modify them to communicate the protection offered by multiple PETs in combination.

Acknowledgments

We would like to thank everyone who provided feedback on various stages of this project, especially Aaron Broukhim, participants in the Technically Private reading group, and our reviewers. All authors were supported by DARPA (contract number W911NF-21-1-0371). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government or DARPA. In addition to DARPA support, the third author was supported in part by NSF grant CNS-1942772 (CAREER), a Mozilla Research Grant, a JP-Morgan Chase Faculty Research Award, and an Apple Privacy-Preserving Machine Learning Award. The fourth author was also supported by NSF grant #2030859 to the Computing Research Association for the CIFellows Project, and the fifth author was also supported by a Google Research Scholar Award.

The material in this chapter has been submitted for publication. Smart, M. A., Nanayakkara, P., Cummings, R., Kaptchuk, G., and Redmiles, E. M. Models Matter: Setting

Accurate Privacy Expectations for Local and Central Differential Privacy. The dissertation author was the primary author of this paper.

Chapter 4

Negotiating Privacy Interdependence on Facebook

4.1 Introduction

The issue of data privacy has been a topic of concern among computer scientists, legal scholars, and the media alike—particularly in the context of social media. In the past few years, popular social networking sites have responded to privacy concerns by introducing new privacy settings for their users [53, 328]. Yet there is a growing understanding that granting users greater control over their individual data still leaves users vulnerable [316, 298, 296]. Protections of this nature that operate at the individual level fail to account for interdependence within the network. In practice, the information one shares on social media can reveal a great deal not only about oneself, but also about one’s friends [211, 278, 108, 321]. When social networking sites offer users more privacy settings and transparency tools, they may placate concerns about privacy—and give users a false sense of security—without meaningfully shifting the “*power imbalance between data collectors and the data subjects*” [76]. This work aims to expose and explore concepts of privacy interdependence with Facebook users through a probe tool that displays inferences that can be made from network data.

Previous work has explored ways to raise awareness about privacy issues by showing people the kinds of inferences that can be made about them from their data [335, 358]. But little is known about awareness of and attitudes towards algorithmic inferences that can be made

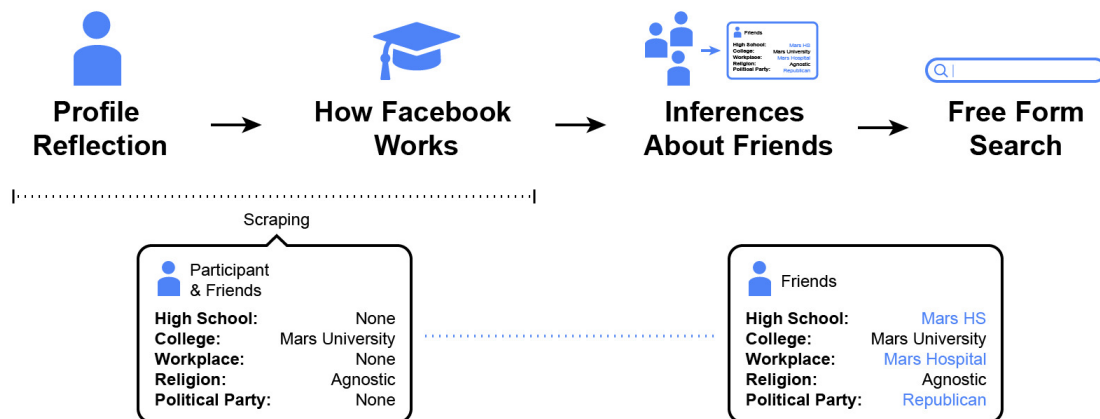


Figure 4.1. FIG’s four stages. Each of the four stages serves a particular purpose. Scraping continues in the background during the first two stages. In the final two stages, FIG displays the inferences made from the scraped data.

not from any single individual’s data but from the larger network. Our study fills this gap by interviewing 27 Facebook users, employing a probe tool to prompt users to think more deeply about interconnected privacy concerns. Through these interviews, we seek a better understanding of how Facebook users reason about and negotiate interdependent privacy.

Our probe tool—the Friend Inference Generator (FIG)—scrapes information from a user’s Facebook network. Based on the information collected, FIG makes inferences about attributes such as workplace or places lived. FIG guides users through four distinct stages: 1) a stage that shows the user their own “about section” of their profile; 2) a stage that displays examples of how Facebook leverages user data for targeted advertising; 3) a stage that displays inferences made about the user’s friends; and 4) a stage that allows for the open exploration of all inferences made by FIG (Figure 4.1). A final step deletes all scraped information.

Each stage of FIG’s interface complements our interview protocol, enabling us to probe participants’ understanding and perceptions of privacy issues on Facebook. In particular, the four stages serve the following respective functions in our interviews: 1) familiarizing the participant with the “about” section of their Facebook profile and prompting them to reflect on their information-sharing decisions; 2) determining familiarity and attitudes toward targeted

advertising; 3) introducing participants to the kind of inferences that can be made about their friends from network data; and 4) revealing how participants imagine different friends might respond to different kinds of inferences.

We find that algorithmic inferences — particularly incorrect inferences — are frequently perceived as algorithmic stereotyping. However, participants anticipate a range of reactions from their friends in response to these inferences—from anger to amusement. Our findings echo prior work on the significance of the alignment (or lack thereof) between one’s sense of identity and the “algorithmic self”; people may be upset when algorithmically generated inferences do not align with their own sense of self [87]. We also find that participants tend to rely on workarounds—rather than built-in privacy features—to manage shared privacy concerns. While participants are generally willing to honor privacy-related requests from friends out of respect, these requests can nonetheless lead to judgment, tension, or conflict. For example, even if a friend’s requests are honored, that friend may be perceived as paranoid. The understanding that others’ actions can undermine one’s own privacy may lead to resignation. Therefore, we point to the need for better tools for managing privacy collectively, that can inspire and support collective action related to privacy issues.

4.2 Background

4.2.1 Evolving Conceptions of Privacy

On November 29, 2011—the same day that Facebook settled with the Federal Trade Commission regarding charges that it had failed to keep its privacy promises to users[91]—Facebook founder, Mark Zuckerberg, wrote in a Facebook post that he was “*committed to making Facebook the leader in transparency and control around privacy*” and that “*everyone needs complete control over who they share with at all times*” [381]. Now a wide range of privacy settings, data download portals, and other tools offer individuals greater control over their data on Facebook and other popular social media.

Privacy, however, involves more than just control of information [239]. Reducing the idea of privacy to questions of individual control over information “*greatly reduces what can be private*” [316], since much of the data of our daily lives is inherently relational [37, 243, 356, 165, 146]. This is especially true within the context of social media [307]. For example, while Facebook users can choose whether or not to upload photos to the site, they cannot control the photos that their friends share [319]. Privacy conflicts related to photo sharing are quite common and can contribute to context collapse [309, 210, 78]. As another example of the relational nature of privacy, even when individuals opt-out of sharing certain kinds of information, such as location information, that information can often be inferred using data from the individual’s network [61, 31, 108, 230, 23, 218].

Recognizing the inherent limitations of individual control, Marwick and boyd (2014) propose a theory of *networked privacy* and argue that “*legal and technical regimes around information privacy must adapt to better reflect the reality of networked social information*” [211]. Similarly, *privacy interdependence* is used to discuss the fact that the “*protection of personal, relational and spatial privacy of individuals is increasingly dependent on the actions of others, rather than the individuals themselves, in the interconnected digital world*” [33]. For example, mobile apps may request access to photos, calendars, or contacts—all of which typically contain information not only about the device owner but also about other people [208]. Another closely related concept is *group privacy*. Members of a group may face threats not only to their personal privacy but to their collective privacy as a group [310, 355]. For example, genetic research involving Indigenous people poses privacy threats not only to individuals but also to the entire tribal community [324]. A social understanding of privacy opens up the space of possible interventions and invites solutions that are not purely “technical” [363, 74, 262].

4.2.2 Negotiating Privacy

Making decisions about relational data or responding to multiparty privacy violations typically requires some form of negotiation. For example, social media users may attempt to

avoid conflicts by asking for permission before sharing photos of their friends [309]. To support this kind of negotiation, systems have been proposed that might prompt users to request their friends' consent or even automatically notify friends and request their permission before photos can be posted [238, 369, 274]. Privacy issues on social media that may require negotiation extend beyond photo sharing as well. For example, social media users may expose information about their friends by tagging them in other kinds of posts [191]. Users often trust their friends to follow shared norms about what is appropriate to share on social media, but misunderstandings may still lead to conflict and require corrective action—such as deleting a post or untagging a friend from a photo [184]. A variety of tools have been suggested for preventing these kinds of conflicts, such as software agents that can negotiate privacy concerns on behalf of users [166, 308].

Many situations beyond social media also require privacy negotiation. For example, AirBnB guests may reach out to hosts about the presence of security cameras or other smart home devices [351]. In some cases, power imbalances may complicate attempts at negotiation around smart home devices; for example, nannies may worry that asking too many questions about in-home cameras could prompt suspicion [28]. Decisions about sharing genetic data offer another context in which privacy negotiation may be warranted. Educational tools for exploring how genetic data sharing affects relatives may help to prompt a family conversation [147]. Advances in big data analytics further complicate privacy negotiation, since information that one person shares may be aggregated with other data to make unexpected inferences about other people.

4.2.3 Reasoning about Inferences

A growing body work has studied the threats that algorithmic systems pose to privacy [25, 197, 286, 380], and legal scholars have pointed out the new challenges that such systems pose for data protection regulation [296, 349]. An important part of understanding and addressing these issues is understanding how people reason about these privacy threats [60, 162, 133, 132, 258, 362]. When people make decisions about whether or not to share certain kinds of data,

it is often impossible for them to reason about the kinds of inferences that may be made from the information they share and about how these inferences may threaten their privacy and the privacy of others [296, 136]. Prior work has found that people are often surprised to learn about how interests may be inferred from browsing data [358]. Certain kinds of inferences are seen as particularly creepy; for example, many people are uncomfortable with inferences related to health conditions or religion [73]. The alignment between self-perception and the *algorithmic self* also influences how people feel about the inferences algorithms make about them [87]. For example, people may react with anger when inferences seem to contradict an important aspect of their identity.

The now-defunct Data Selfie project sought to bring greater visibility to the machine learning algorithms that operate behind the scenes on Facebook [335]. While our tool is in part inspired by the Data Selfie extension, the functionality is somewhat different. Since we are interested in understanding Facebook users' attitudes towards privacy interdependence, our tool focuses on exposing inferences that can be made about someone from the data of their friends. Thus this tool helps us address our main research questions:

1. RQ1: How do Facebook users imagine that their own information sharing affects their friends?
2. RQ2: How do Facebook users negotiate privacy with their friends?

4.3 Designing the Friend Inference Generator (FIG)

Our probe tool, Friend Inference Generator (FIG), was designed to help participants think about the ways that information they share on Facebook could be used to infer things about their friends. We begin with the data collection (Section 4.3.1) and inferential process (Section 4.3.2). Next, we outline the stages of the user experience (Section 4.3.3) and conclude with a discussion of ethical considerations (Section 4.3.4) and implementation details (Section 4.3.5).

4.3.1 Scraping from Facebook

FIG first scrapes information from the “about” section of the participant’s profile as well as the full list of the participant’s friends. Next, the system proceeds through the list of friends, scraping information from the “about” section of each friend. In addition, for each of the participant’s friends, the system records a partial list of their mutual friends. This information about the social graph is what we ultimately use to make inferences. Because collecting friend names and data required iteratively scrolling and waiting, it created a significant bottleneck. Due to time constraints, we did not collect information on all of a user’s friends. Instead, we collect as many as possible in approximately 20 minutes.

Scraping user data from Facebook is difficult by design. We used Selenium to automate the process of clicking through Facebook and scraping the desired information. Deploying as a standalone website would require users to install a browser extension to disable content security policy headers. This could weaken the security of users’ browsers, so we hosted the system locally and provided access via remote control. Users would use Zoom remote control to access the researcher machine and log in. All data is stored in MySQL database on the researcher machine, before being deleted during the final phase of the system use. In the interest of participants’ privacy, we did not collect any summary statistics about participants’ networks.

4.3.2 Making Inferences

FIG makes inferences about the following kinds of information commonly shared in the “about” section of Facebook profiles: 1) workplace, 2) college/university, 3) high school, 4) places lived, 5) religion, and 6) political views. We use a majority vote system for making inferences, similar to the approach taken to infer locations of Twitter users in prior work [61]. For each friend, we check to see for which attributes no information was shared explicitly (e.g., workplace). For each missing attribute, we infer a value by looking at the majority vote among the mutual friends that we have scraped (Figure 4.2). We require at least two votes in

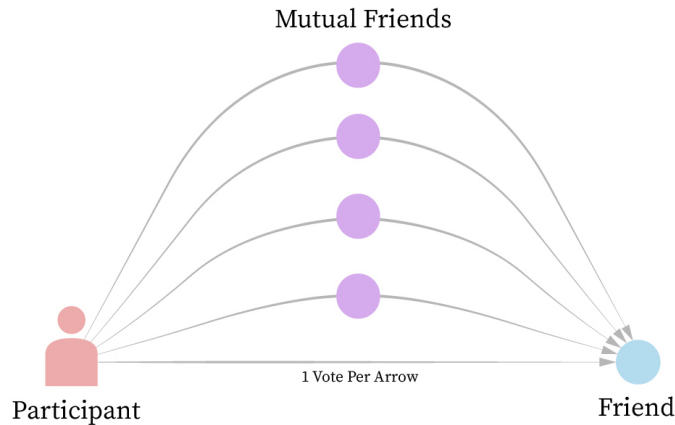


Figure 4.2. Strategy for making inferences. To make an inference about a particular friend, FIG turns to that friend’s mutual friends with the participant. A simple voting process determines which inferences FIG makes.

order to make an inference. In the case where too few mutual friends have shared the relevant information or in the case of a tie, we make no inference. This is a simplification of the process that Facebook uses to make inferences and select audiences for ad targeting, since Facebook has a much larger collection of data from which to draw. FIG’s purpose, however, is to reveal the kinds of inferences that Facebook may be able to make from networked data—not to provide details of the exact mechanisms that Facebook employs or achieve state-of-the-art accuracy. Even inaccurate inferences may provide meaningful opportunities for reflection. We acknowledge the simplifications we make in our interviews and explain that Facebook has access to large quantities of data beyond the scope of what we collect.

4.3.3 FIG’s Four Stages

FIG’s interface walks participants through four distinct stages.

Stage 1: Reflection

In the first stage (Figure 4.3), FIG directs participants to their own profile page (opened in a new tab), so that they can remind themselves what information they themselves have shared in the “about” section. This not only helps participants re-familiarize themselves with this section

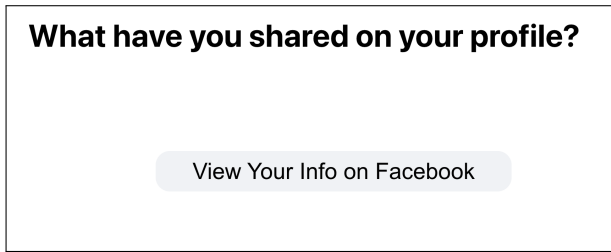


Figure 4.3. Stage 1. FIG provides an external link to the user’s actual live profile on Facebook.

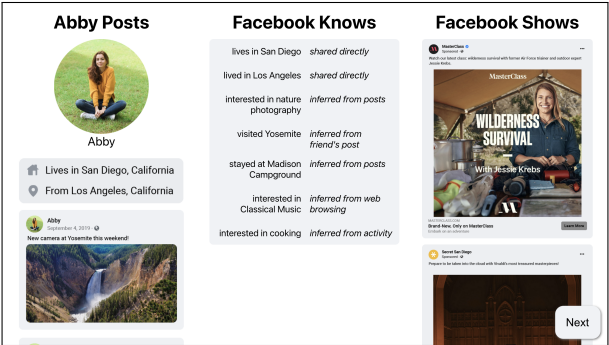


Figure 4.4. Stage 2. This page shows how Facebook might use information shared by “Abby.” of Facebook profile pages but also prompts them to think about their own past information sharing decisions.

Stage 2: Background

Next, FIG illustrates the way that Facebook uses data to make inferences about its users and to target ads. This stage is the same for every participant, and it is intended to help participants understand why Facebook makes inferences about its users—namely, so that Facebook can better target ads. This background sets the stage for the rest of the interview and helps participants understand the connections between targeted advertising, privacy, and their social network. The first page shows examples of the kind of information that Facebook uses to target ads (e.g., users’ activity on Facebook). The second page (Figure 4.4) illustrates ad targeting on Facebook with examples. The leftmost column of this page displays a sample profile for an imaginary Facebook user “Abby” with posts about recent trips. The middle column displays information that Facebook knows about Abby—such as the fact that Abby recently

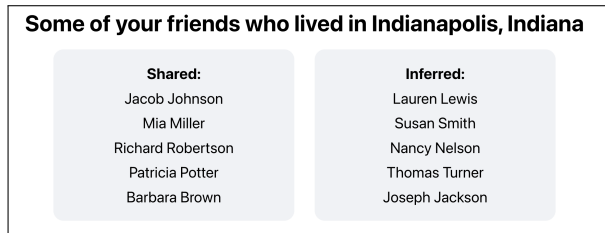


Figure 4.5. Stage 3. This page shows a list of friends who shared in their about sections that they lived in Indianapolis (left) and a list of friends that FIG inferred lived in Indianapolis even though it was not shared in their about sections (right). Names in this screenshot are fabricated.

visited Yosemite—based on the information Abby has shared in addition to other sources such as web browsing. Finally, the third column displays ads that Facebook shows Abby—such as an ad for a class on wilderness survival—based on what Facebook knows about Abby. After this introduction to targeted advertising, participants are set up to better understand and reflect on the kinds of inferences Facebook might want to make about users.

Stage 3: Inferences

In the third stage, FIG highlights some of the inferences that can be made about the participants’ Facebook friends from data about their network. Building on prior work exploring attitudes towards inferences and interdependent privacy [87, 358, 208], we wanted to prompt participants to think about how others might be affected by their data sharing decisions.

Stage 3 consists of several subpages. First, FIG shows participants a list of their friends who did not share much information on their profiles and a list of their friends who share a great deal of information on their profiles. The purpose of this stage is to prompt participants to begin thinking 1) about their friends’ information sharing preferences, 2) about why some friends might not want certain information shared on their profiles, and 3) about why others’ information sharing preferences might be different from their own. To come up with these two lists of friends, we calculate how much of their profile each friend has filled out. In particular, we look at whether each user has shared any information under the profile sections: workplace or college/university, high school, places lived, religion, and political views. In other words, we

look at the sections of the profile that FIG makes inferences about. We chose to count workplace and college/university as one category since we imagined it would be quite common for people to be missing one or the other of these pieces of information, not because they wanted to keep the information private but because they had not gone to college or because they were currently in college and had not ever had a job.

Next, FIG highlights a particular category. For example, the category might be “friends who have lived in Indianapolis”. FIG may select this category if many of the scraped profiles share having lived in Indianapolis and if FIG has inferred that at least a few additional friends have lived in Indianapolis. Initially, FIG shows a list of several friends who shared in their about sections that they have lived in Indianapolis. Then FIG adds a second list of friends (Figure 4.5). FIG has *inferred* that these friends have lived in Indianapolis. This allows participants to make a side-by-side comparison, focusing on one portion of their social network (e.g., friends made while living in Indianapolis). This comparison is valuable for several reasons. First, it reinforces the distinction between inferred and explicitly shared information—introduced earlier in the interview. Second, it helps the interviewer present a brief explanation of how inferences are made—the information explicitly shared by certain friends (displayed on the left) is used to make inferences about other friends (displayed on the right). Third, the comparison prompts participants to reflect on why their friends may have different information sharing behaviors and preferences.

Next, FIG displays a brief list of additional categories (e.g., “friends who have worked at Best Buy”) that are automatically generated based on what the user’s friends have chosen to share or not. This step provides additional opportunities for comparison. Participants can click on each category to see side-by-side lists of friends, with the friends who directly shared the information on the left and the friends about whom this information was inferred on the right. This stage is to introduce the participant to FIG’s inferences in a guided way and highlights how individuals’ information sharing decisions might affect their friends.

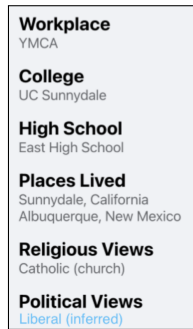


Figure 4.6. Stage 4. FIG displays information from a friend’s about section (gray) and any inferences made about that friend (blue).

Stage 4: Exploration

In the fourth stage, FIG allows the participant to engage in more open-ended exploration of their friends’ profiles and the inferences made about them. This stage allows the participant to explore their own network of friends, reflecting on particular relationships and what information Facebook might extract from these relationships. The open-ended exploration also allows the participant to test their own hypotheses about their social network. The participant can type into a search box to pull up information about specific friends (Figure 4.6). Early pilot tests revealed that for some participants — depending on the particular qualities of their network — FIG might only make inferences about a small subset of friends. In order to better highlight this subset, a blue star appears next to a friend’s name if FIG made any inferences about that friend. When the participant selects a particular friend, FIG will show the information shared in that friend’s about section. Highlighted in blue text, FIG will also show information that has been inferred about this friend.

4.3.4 Ethical Considerations

Although our institution determined that our protocol was exempt from full committee review by our institution’s review board (IRB), as other researchers have argued [248, 346, 376], IRB approval is not an ethical green light. Thus we reflect on the ethical concerns involved in this study and how we addressed them.

Our system is designed to respect Facebook users' privacy. The friends of our participants shared information about themselves with their friends through Facebook, but did not agree to share it with our research team. Since "*privacy violations might occur when information about individuals might be readily available to persons not properly or specifically authorized to have access the data,*" we do not retain the information of participants' friends [375]. Rather, we display the information back to the participant during the course of the interview. Thus the *context* in which we use this information (i.e. displaying it to the participant) aligns with the social context in which the information was shared (i.e. with Facebook friends, including the participant [239]).

Nonetheless, displaying *inferences* made from collected data may still violate Facebook users' expectations [349, 296]. For this reason, we avoid making inferences about attributes that may be especially sensitive, such as sexual orientation. Furthermore, because the the intuitive process we use for inference—a majority vote among mutual friends—is quite simplistic, it is unlikely that the inferences would reveal anything that the participant does not already know. Finally, we also make it clear to participants that not all the inferences will be correct and thus should not be taken as fact. We consider it essential to our study to display these inferences to participants, as it allows them to reflect on what the information they share may reveal about their friends.

After the interview, the participant watches the interviewer delete the collected data. Meta's privacy policy states that the company "collect[s] your activity across our Products and information you provide," meaning that the information users share on Facebook is already part of an ecosystem in which data travels far beyond its original context¹. Given these precautions, we believe that our study caused little harm to Facebook or its users.

Nevertheless, scraping user data is a violation of Facebook's terms of service (TOS). Prior work has argued that TOS violations in the context of research are not necessarily unethical [333, 100, 222]. It is potentially dangerous for Facebook to control which researchers have access

¹<https://www.facebook.com/privacy/policy?subpage=1.subpage.1-YourActivityAndInformation>

to the data needed to study Facebook; this may give Facebook the power to shape which research questions and—even more worryingly—which conclusions are acceptable [100, 138, 333]. Although Facebook does offer APIs for developers and researchers, their capabilities are limited. During the 2010s, Cambridge Analytica obtained and misused data from millions of Facebook users through a personality quiz app that gathered information not only about the app’s users but also about users’ Facebook friends [46, 273]; in response to this scandal, Facebook made significant changes to their APIs [336]. Thus, it was not possible to use Facebook’s tools for developers to obtain the data we needed for this study. The tools Facebook offers for researchers are also quite limited. After the Cambridge Analytica scandal, Facebook formed a partnership with Social Science One so that academics could study the “effect of social media on democracy and elections” [289]. Facebook has since been criticized for delays in providing access to data for researchers as well as for inaccuracies in the delivered data [134, 348, 320]. More recently, Facebook has released the Facebook Open Research and Transparency platform, available to a select group of researchers. Facebook is not the only social media platform that has been criticized for erecting barriers to academic research [183], and it has been said that researchers now operate in a ‘post-API age’ [102].

Participants were encouraged to read through a consent form before signing up for the study, so that they would clearly understand 1) that they would be asked to sign into Facebook on the researcher’s computer and 2) how their privacy would be protected. This document explained what would happen during the study, described potential risks, mentioned that scraped data would be promptly deleted, and clarified that participants could withdraw from the study or refuse to answer particular questions at any time. At the beginning of each interview, we asked participants to review this document again and ask any questions they might have.

4.3.5 Implementation

FIG scrapes information from Facebook using Selenium [291], storing the data in a local MySQL [244] database that is deleted at the end of each interview. The frontend for

FIG's interface is implemented using React [299], and a Flask [121] server acts as the backend. Everything is run locally on the interviewer's computer. Although we do make FIG's code available², we note that because of the scraping component, this code requires updating every time that Facebook's own frontend changes.

4.4 Methods

In order to investigate the research questions posed in Section 4.2.3, we conducted 27 semi-structured interviews over Zoom, using FIG as a probe tool. Each participant navigated through the stages of FIG's interface, and each stage facilitate different kinds of interview questions about targeted advertising, relationships with friends, privacy concerns, and so on.

4.4.1 Recruitment

It was important to recruit a diverse group of participants, since privacy concerns are related to cultural context [227, 192, 330]. We advertised the study as a “social media research study” and explained that we aimed to “learn about people's attitudes towards inferences that can be made using social network data.” In order to reach a diverse audience, several complementary approaches were used for participant recruitment, including posted flyers (in spaces such as coffee shops and public libraries), social media posts (on Facebook, Reddit, NextDoor, etc.), and notices in email newsletters. Care was taken to respect the rules and norms of the various social media platforms during recruitment. The eligibility criteria were having a Facebook account, being at least 18 years old, and living in the United States. In total, two authors conducted 27 interviews. In two cases where the participant had no Facebook friends, the interviews were discarded from our analysis, leaving a total of 25 interviews. Participants were paid \$15/hour, and the median interview length was about 43 minutes (range = 24–68 minutes). Participants were representative of the United States in terms of household income (median = \$50,000 - \$74,999) and education (median = *Some college, no degree*). But participants were more likely

²https://osf.io/tsjfa/?view_only=c2ccd115046348f7ab4607287a88500c

Table 4.1. Selected interview questions for each stage. This table provides examples of the kinds of questions asked during interviews, though it is not a complete list.

Stage	Sample Question
1	<i>Can you think of any way that Facebook could use your information that would make you uncomfortable? How did you decide what information to share in the about section of your Facebook profile?</i>
2	<i>Do you think that any of your Facebook posts reveal information about your friends that might be used to target ads? How do you think your friends would feel about that?</i>
3	<i>Have you ever had a friend ask you to limit what information you share about them on Facebook? Do you think that any of your friends would be upset by these inferences?</i>
4	<i>How do you think your friend would feel about these inferences? Do you think that this inference would reflect positively or negatively on your friend?</i>

to be men (58% men, 38% women, 4% nonbinary), Black, Hispanic, or Asian (50% Black, 23% Hispanic, 12% Asian, 15% White), and young (most were under the age of thirty; our two oldest participants were both in their fifties) compared to the US population. Detailed participant demographics can be found in Appendix C. Although some participants reported using Facebook only infrequently, most participants (19 of 25) reported using Facebook on at least a weekly basis. Participants were given the opportunity to read the consent form to learn about the study and any potential risks before signing up for an interview. They are also encouraged to review the consent document and ask questions at the beginning of each interview.

4.4.2 Interview Protocol

Using FIG as probe tool, interviewers guided each interview through multiple stages, prompting participants to reflect on networked privacy. Figure 4.1 outlines FIG’s four stages, and Table 4.1 provides examples of questions asked in each stage. The full protocol can be found on OSF. After submitting thorough materials for review, such as our interview protocol and screenshots of our interface, this protocol was determined to be exempt from full committee

review by our institution's review board. In total, two authors conducted 13 and 14 semi-structured interviews.

Beginning the Interview

Each participant first gave their consent to participate in the study through an online form that also included a brief demographic survey. Next, the participant logged into Facebook on the researcher's computer, so that FIG could begin running. As the participant logged in, the interviewer briefly explained how FIG would review the about sections of their friends' profiles and make inferences. Finally, the participant and the interviewer discussed a choice of pseudonym. Addressing this task at the beginning of the interview gave FIG more time to scrape data from Facebook.

The interviewer began with general questions about the participant's use of Facebook. These initial questions (e.g., frequency of usage) served as easy-to-answer warm-up questions for participants while providing additional time for FIG to scrape. The interviewer also asked how the participant perceived their own level of concern about privacy compared to their friends. This question was meant to help participants start thinking about their friends' privacy preferences.

Stage 1

FIG's first stage displayed the information in the participant's about section and thus prompted the participant to reflect on their information-sharing preferences. As scraping continued in the background, the interviewer asked about how the participant decided what information to share and about how Facebook might use this information. Next, the interviewer probed the participant's familiarity with and attitudes towards targeted advertising on Facebook, borrowing questions from prior work on targeted advertising [331].

Stage 2

Stage 2 explained how targeted advertising works on Facebook. This stage outlined kinds of data that Facebook might collect (e.g., users' web browsing) and explained to the participant

that ads may be targeted based on information that had been explicitly shared (e.g., user lives in Atlanta) and/or information that had been inferred (e.g., user is interested in nature photography). Since this distinction between explicitly shared versus inferred information would be important in the remainder of the interview, the interviewer checked the participant's understanding by asking them to explain the distinction in their own words. Follow up questions prompted the participant to reflect on their own and their friends' feelings towards targeted advertising on Facebook.

Stage 3

Stage 3 presents a walk-through of some of FIG's initial inferences. In this stage, the interviewer explains how FIG makes inferences (Section 4.3.2). Questions for this stage prompted the participant to reflect on 1) their friends' privacy preferences, 2) their perceptions of the presented inferences, and 3) their friends' likely perceptions of the presented inferences. Interviews asked about friends' privacy preferences in order to understand how each participant accounted for the fact that their friends may have privacy preferences that differed from their own. Participants were asked about the perceived accuracy of inferences and how this shaped their perceptions. For example, if a participant perceived FIG's inferences to be inaccurate, the participant might be prompted to consider situations where false inferences could be harmful. Alternatively, if a participant perceived FIG's inferences to be highly accurate, the participant might be prompted to consider the relative "creepiness" of different kinds of inferences. Finally, questions about friends' perceptions of inferences again probed how participants accounted for friends with varying privacy preferences. Since the connection between information sharing and ad targeting on Facebook remains largely opaque [19], Facebook users may not have previously considered this connection between their own behavior and the inferences that Facebook might make about their friends.

Stage 4

Finally, in stage 4, the participant could freely search for any friend whose profile FIG had scraped and view the inferences FIG had made. The participant was asked to think aloud while exploring the inferences using FIG's interface [79]. Building on the discussion from the Stage 3, the participant was asked to reflect on these inferences, how their friends might feel about these inferences, whether or not these inferences reflected important aspects of their friends' identities, and their relationships with their Facebook friends.

Ending the Interview

Participants were asked to reflect on a few final questions, such as what Facebook might be able to learn about their relationships with friends given that Facebook has access to much more data than FIG. Finally, the interviewer logged the participant out of Facebook, quit the browser, and deleted all the scraped data while screen sharing so that the participant could watch. This step was included to reassure participants that the scraped data would not be retained.

Handling Failures

In 13 of the 25 interviews, the participant had so few friends or such a sparse network that FIG made no inferences. For any failure cases, the interviewer adapted the questions as needed—for example, during stage 4, participants might click on different friends to see what information they shared in their profiles, providing an opportunity for the interviewer to ask hypothetical questions about how that friend might feel about a particular inference. These hypothetical questions still offered a way to understand how participants thought their friends might react to inferences, though participants may have seen the inferences as less feasible. It is also possible that the hypothetical examples invented by the interviewers were less relevant than those surfaced by FIG. In our analysis, we compare results for interviews where FIG was able to make inferences against those in which FIG's inferences failed.

4.4.3 Selecting Pseudonyms

In order to protect participants' privacy, quotes from interviews will be attributed using pseudonyms. A variety of common practices exist for assigning pseudonyms to participants. For example, sometimes researchers simply number the participants and then refer to them by number; in other cases, researchers may assign pseudonyms based on participants' gender and birthyear [199]. In recent years, several scholars have pointed out that these common practices are not always appropriate [43, 14, 169]. For example, Dankwa notes that these practices for assigning pseudonyms fail to recognize "*the inherent complexity of identity behind the names*" and calls for HCI researchers to "*relinquish the privilege of renaming and engage in negotiation for identity representation*" [169]. We decided to include our participants in the process of selecting pseudonyms by offering to let them the option of choosing their own pseudonym. While it is possible that a participant could choose something like a nickname that might hint at their true identity, we nonetheless choose to prioritize participants' agency.

4.5 Analysis

The interview audio recordings were uploaded to Dovetail for transcription and analysis. The recordings were automatically transcribed through Dovetail, and the research team corrected errors in the auto-generated transcripts during the analysis process. First, two researchers read through the transcripts, each focusing on different sections of the interviews. We defined a segment of analysis to be a block of text between speaker changes. Each researcher highlighted segments of the transcripts and assigned codes to each highlighted section. Multiple codes could be applied to a single segment. The other researchers provided feedback on the codes that were developed, what codes might need to be added, and how codes might be organized into higher-order themes [39]. The same two researchers made a second pass over the transcripts, since several codes had been added along the way. The final codebook, containing the list of codes is included on OSF. We then employed affinity diagramming to map the connections

between the low-level codes, higher-order themes, and our research questions [29].

4.6 Results

4.6.1 Reasoning About Inferences

Algorithmic Stereotypes: Funny or Fraught

Participants explained that certain friends were more worried about privacy than others and imagined a range of ways that friends might respond to algorithmic inferences. Some participants explained that their friends do not worry much about privacy. Sometimes this lack of concern was linked to a lack of awareness (n=6). For example, Florencia explained: “*I don’t think people are really aware. I always hear people saying, oh they’re watching us or listening to us—which is creepy but I guess we’re all just too busy to really think about it.*” Given how technology companies tend to offload the significant burden of privacy management onto individual users, it is unsurprising that some people feel that they are “*too busy*” (n=3) to think about privacy [76, 215, 8, 296]. There are indeed too many privacy policies to read and privacy settings to manage, and overly permissive default settings exacerbate this problem [34]. These feelings of resignation were common among participants (n=9). Casey expressed a similar sense of resignation about digital surveillance as follows: “*I think [...] there’s this social contract that, it’s just how it is. [...] I think people know it’s happening all the time. And I think that there’s just this understanding that this is just how the internet works now.*” For this participant, even if he knows that Facebook can learn information about his friends from his posts—and use that information to target ads—he sees this as an expected consequence of using Facebook that he and his friends have accepted.

Due in part to this lack of concern about privacy, some participants did not think their friends would care much about algorithmic inferences at all (n=8). For example, Binx did not think her friends would care that Facebook might make inferences about them unless Facebook actually published that information. In other words, the acceptability of inferences depends on

how they could be used.

Most participants n=16, however, said that they do have at least one friend who is particularly concerned about privacy. For example, Casey explained that some of his friends with left-leaning politics are more concerned about “*protecting information and [...] going offline.*” Alexander has a friend who worries about being outed as gay. Other participants have friends who worry about stalking or harassment. Andrea offered one such example: “*Well this friend of mine, she’s also a really radical feminist, and she’s kind of famous on Twitter. So I do think that she has enemies. She has haters. So I think one of them may actually be weird enough to try and go to her house.*” When asked about privacy-related requests or friends who were particularly concerned about privacy, it seems that some participants gravitated towards these cases where privacy was closely linked to physical safety (n=5). It may be that these were the most salient privacy concerns or the concerns that were easiest for participants to understand. For example, some participants mentioned friends who cared about privacy simply because they were “*private people*” (n=5). It seemed that participants did not know exactly why these people cared about privacy, but attributed it as an aspect of their personality. Andrea made a very explicit connection between her friend’s concern about privacy and this friend’s personality. She said that her friend worried about privacy because of her OCD—because it was “*part of her to be concerned about everything.*” Some of the other reasons that participants gave for why their friends care about privacy included wanting to reduce the risk of being hacked or having their identity stolen (n=4) or wanting to protect their online identity by removing content that could be embarrassing or damaging (n=7).

When prompted to reflect on the accuracy of the inferences, participants often believed that the (in)correctness of inferences would affect how their friends would feel about them. For example, Erin thought her friends would be “*weirded out*” or “*pretty upset*” by incorrect inferences. Some participants viewed incorrect inferences as inaccurate stereotypes (n=6). For example, Florencia thought her cousin would be upset by the incorrect inference that he lived in Tijuana, Mexico: “*Because he’s military so he might be like, ‘Hey, [...] I’m from the US—why*

are you guys assuming that I live in [Tijuana]? Just because my last name's Hernandez³ and I have family members there?" On the other hand, she felt that a different cousin would respond with amusement rather than anger. She speculated that this cousin might think: *"Oh wow, they just guessed that because my last name is Hernandez and my relatives in Mexico went there—soo they're just assuming that we all went there. [...] Because of my culture? Because of my last name? Because of my race?"* Although it sounds as though this cousin might also be angry, Florencia clarified that *"she would think it's funny."*

Correct inferences also inspired a range of different reactions. For example, some participants thought their friends would find these inferences *"cool"* (n=2). Participants also imagined that these inferences could be useful or beneficial for their friends (n=5). For example, some participants thought their friends might benefit from being targeted with advertisements that matched their interests. On the other hand, some participants thought their friends would find correct inferences creepy or upsetting (n=8). Thus both correct and incorrect inferences were at times perceived as harmful.

Effect of Real Inferences

The fact that FIG sometimes failed to make any inferences helped us understand how seeing real inferences affected participants' reasoning. We found two interesting patterns. First, of the six participants who brought up ways that Facebook might violate users' expectations, all were in the group of interviewees who saw real inferences generated by FIG. Seeing these inferences may provide an element of surprise that is missing when inferences are discussed in purely hypothetical terms. Second, of the ten participants who discussed tagging, all but one saw real inferences in FIG. Five participants brought up tagging specifically in the context of reasoning about inferences. For example, Elise reasoned that Facebook may be able to infer how often she spends time with a particular friend based on photos that the two of them are both tagged in. It may be that seeing real inferences made using the participants' real networks

³The individual's real last name has been replaced with a pseudonym for privacy.

helps participants reason about the kind of information algorithms might use to make inferences. Thus, while participants who did not see real inferences generated by FIG still had interesting thoughts to share, the experience of talking about hypothetical inferences does not quite match the experience of discussing real algorithmic inferences. This reinforces the value of tools like FIG for engaging social media users in conversation about algorithms and privacy.

4.6.2 Negotiating Privacy

Our participants had experienced a range of privacy requests from their friends related to Facebook. While participants were generally willing to comply with these requests, requests did sometimes lead to conflict.

Inferences Complicate Privacy Negotiation

Participants generally agreed that it was reasonable for friends to make privacy-related requests—such as requesting that a photo be removed. Some participants expressed that their willingness to comply with such requests stemmed from their respect for their friends’ autonomy. In other words, participants believed that their friends had the right to choose what information should be shared about them online (n=6). For example, Code explained that his friends have “*freedom of choice*,” and that he “*respect[s] their decisions*.” As we will see in Section 4.6.2, participants did not always understand their friends’ privacy concerns. Nevertheless, participants generally felt that even requests that they did not fully understand should still be honored.

Harsh made an interesting comparison between the way that friends treat each others’ information and the way that Facebook treats users’ information:

It kind of goes back to the idea of if I were to directly share a picture of my friend onto my Facebook page and they were to say, ‘Hey, please take that down.’ [...] I would do it. But if I shared something about myself and tagged a friend in it—and he was okay with that—but then Facebook came to a bunch of random conclusions, a bunch of inferences about him based on that, that he didn’t intend for that to happen. That feels, in the same principle, but we don’t have control.

In other words, Facebook is complicating Harsh’s desire to respect his friends’ autonomy. He

is willing to communicate with his friend and to respect his friend's wishes, but Facebook comes between them and behaves in a way that neither expects. The opacity of Facebook's algorithms limits Harsh's ability to communicate with his friend about this content and about what will happen if he shares it. Facebook takes away their agency and control over this shared content—getting in the way of effective privacy negotiation. Like Harsh, Kevin felt that it was “*wrong*” for Facebook to use content he shared to make inferences about his friends. Facebook's algorithms get in the way of the processes—however imperfect—that participants and their friends use to manage privacy.

In talking about past experiences with privacy negotiation, participants tended to discuss topics like photo-sharing and tagging rather than algorithmic inferences. This is perhaps unsurprising since Facebook offers very limited visibility into the algorithms they use to process user data. Nonetheless, a few (n=3) participants explained how their new understanding and awareness of algorithmic inferences might change how they approach privacy negotiation. For example, Casey explained that he might cut back on tagging friends on Facebook, so as not to “*carte blanche help Facebook*” learn more about his friends. Furthermore, some participants (n=5) reasoned about how photo-sharing and tagging might allow Facebook to make certain inferences about their friends.

Privacy Negotiation Requires Workarounds

Participants discussed several ways that they work around the limitations of Facebook's affordances for privacy management. In theory, Facebook offers settings to control the visibility of one's posts, so a Facebook user may decide to share a group photo only within a limited group of friends rather than posting the photo publicly. Yet our participants rarely brought up these kinds of affordances when discussing privacy negotiation and management. One reason for this may be a lack of understanding or awareness of Facebook's privacy features. For example, Rod mentioned that he sometimes decided not to add someone as a Facebook friend because he was unsure about what he would be able to hide from that individual. A more common strategy for

audience management among our participants was the use of multiple accounts. For example, Blake was more willing to share information about their sister on their “finsta” (Instagram) than on their Facebook account, since their sister could see their Facebook posts. They reasoned that if someone was not able to see a post, then the post wouldn’t “*be associated with them,*” though Blake acknowledged that this assumption may be “*naïve.*”

Many (n=10) participants discussed tagging as a way that friends might reveal information about one another, but not all participants felt that the available privacy settings related to tagging were sufficient. For example, Andrea was aware of the ability to review tagged photos and hide them from one’s timeline, but was unsure about whether it was possible to untag oneself. Raphael thought that Facebook ought to have “*a way that you can restrict people from tagging you,*” noting that Instagram has such a feature. Participants tended to rely on shared norms for preventing potential tagging-related conflicts on Facebook. For example, Guido explained that his social circle has a shared understanding that: “*You don’t tag all of your friends at the gun range on Facebook. That’s just good form to not tag them.*” This reliance on norms means that differences in users’ expectations provide opportunities for conflict.

Privacy Requests Lead to Judgment

Even when participants complied with friends’ requests, these requests nonetheless sometimes created conflict, tension, or judgment (n=5). For example, Andrea has a friend who is particularly cautious about sharing photos of her house. For example, this friend does not want people to share photos from the front of the house, because she is worried that “*people will search [her] up and try to find [her].*” Andrea and the rest of their friends therefore refrain from sharing these kinds of photos. Although Andrea and her other friends are willing to honor this request—Andrea nonetheless implies that the request may be a bit paranoid: “*Maybe if something happened to me, if [...] somebody actually went to my friend’s house. Maybe I would say, ‘People are crazier than I thought. I should stop sharing as much.’*” Andrea’s friend is comfortable making privacy-related requests even though Andrea may not believe such

precautions are necessary. In some social contexts, however, the fear of being perceived as paranoid may prevent someone from expressing their privacy concerns.

As another example of the tensions that can arise from privacy requests, Alexander has a friend who “*doesn’t want that many people to know that he’s gay.*” In particular, this friend is cautious about what photos people take and share of him. He does not want to be included in photos taken at Pride events, nor does he want to be in photos at restaurants that are popular with the local gay community. At one of these restaurants, Alexander may take photos of his friend’s food, but he would not post any photos with his friend’s face. Similarly, this friend asks not to be tagged in posts about going to eat at these restaurants. Alexander complies with these requests as a matter of respect for his friend: “*It’s his profile and I’ve respected that. [...] So I’ll respect whatever choice, decision he wants to make. I mean if he doesn’t wanna be ‘out out’ in a matter of speaking, then that’s his prerogative.*” Nevertheless, Alexander hints that he considers these requests immature, saying that “*even though he’s an adult, he still thinks that way.*” Thus for both Andrea and Alexander, while they may not wholly understand or agree with the requests their friends make, they are nonetheless willing to honor those requests.

In other cases, there is more explicit tension surrounding friends’ privacy requests. For example, Val once posted a photo of her grandchildren that her daughter-in-law asked her to remove. Val was initially uncomfortable with her daughter-in-law’s reaction:

She asked me never to do it again. At first I was kind of insulted because, you know, everybody has pictures of their children having fun [...] and I wanted that too. But then I understand it’s her privacy.

So while Val eventually came to accept her daughter-in-law’s wishes, she initially perceived the request as an insult—particularly because it seemed to go against the norms of the platform. Val wanted to share the same kinds of photos that her Facebook friends were sharing, but ultimately decided that honoring her daughter-in-law’s wishes was more important.

As another example, Blake explains that they sometimes post about shared memories and that these posts can upset their sister. Blake and their sister disagree about the focus of these posts.

Namely, Blake believes that this kind of post centers on their own personal experience, whereas Blake's sister responds: "*no, it's about me.*" In other words, there is a tension between Blake's desire for self-expression and their sister's concern about how such expression reflects on her. In response to this conflict, Blake avoids posting about their sister, with the exception of a more tightly controlled Instagram account that their sister cannot access. Their conflict is perhaps exacerbated by differences in their expectations of each other. For example, Blake's sister asks for her friends' permission before posting photos of them. While Blake says that they respect this choice, they view this practice as "*a lot of energy.*" Blake does not think it is always necessary to ask permission before posting content related to a friend—a view shared by several other participants.

Rod also had a story about a time when sharing information online had led to conflict. His son had posted about him online out of anger. Rod explained that his son "*talked about some stuff that [...] you don't say that stuff out to people.*" His son shared stories about their family that Rod described as information that one is not "*supposed to share.*" In this case, it is possible that the conflict arose from misaligned understandings of social norms—about what information one is "*supposed to share.*" It seems likely, however, that the violation of Rod's privacy was a deliberate act of anger. In either case, Rod's son eventually apologized and removed the offending post.

Just as clashes in expectations and norms can cause or exacerbate conflict, having a clear set of shared norms may help prevent it. But keeping up with norms and expectations is not necessarily trivial, in part because they can evolve over time. For example, Casey recalls playing a part in a friend's art film and explained that it was "*definitely more of a college thing.*" Although he had been happy to participate in the film at that particular time in his life, he later tried to disassociate the video from his profile by untagging himself. These kinds of preferences—and associated expectations—can change over the course of an individual's life.

4.7 Limitations

One limitation of this study is its focus on a single platform that is declining in popularity [347]. It may be more useful to study multiple platforms in parallel, since users sometimes use different platforms to communicate with different audiences—navigating between platforms as a form of privacy management [374, 68, 77]. Although our participants sometimes discussed other platforms, FIG was designed to focus specifically on Facebook. There are a few advantages of focusing on Facebook for studying privacy interdependence. For example, the networked aspect of Facebook is central to the experience of the platform, particularly compared to some other platforms—such as TikTok—that rely more heavily on algorithmic curation than on connections between users [370]. Furthermore, Facebook aims to connect people who know each other, asking users to “*provide for your account the same name that you use in everyday life*” [90]. Some other platforms place more emphasis on interactions between anonymous users. Finally, the Facebook “about” section also offers a place for users to enter structured biographical information in a way that many other platforms do not.

Another limitation is that we interview individual Facebook users. An alternative approach would be a dyadic study. Recruiting pairs of Facebook friends might offer a more holistic view of how friends navigate privacy concerns together. An ethnographic approach might also be useful in investigating how specific social groups with shared privacy concerns navigate privacy together [12]. Diary studies could also be useful for further exploration of privacy negotiation. This approach would be helpful for studying users’ reflections on content they post, content they are tagged in, privacy requests or conflicts, and inferred interests over time.

Another limitation to consider is that demand effects may have influenced participants’ responses—for example, participants may have expressed deep concern about privacy, because that was what they thought we wanted to hear [245]. One way that we mitigate this effect is by complementing questions about opinions and attitudes with questions about specific experiences or behaviors. For example, we ask: Have you ever had a friend ask you to limit what information

you share about them on Facebook? We also ask questions about the experiences and attitudes of participants' friends, which participants may be more inclined to answer honestly.

Two final limitations lie in our approach to recruitment. We focus our recruitment in the United States, although attitudes toward social media and privacy differ across cultural contexts [50, 114, 330]. Even within a U.S. context, privacy attitudes, concerns, and behaviors vary, so we sought to recruit a diverse participant population [125, 36, 196, 202, 265, 111]. While we recruited participants from a range of backgrounds, our sample nevertheless skews young and overrepresents Black Americans. Our sample also contains more men than women, despite the fact that there are more women than men among U.S. Facebook users [271, 118]. Prior work has found gender differences both in privacy beliefs and behaviors [247, 379]; for example, women are less likely to believe that it is “*possible to be completely anonymous online*” [259]. It is possible that we have missed perspectives that may be unique to particular demographic groups. Another limitation of our recruitment strategy is that we did not screen participants based on their frequency of Facebook usage or number of friends. While FIG may have been able to make more inferences if we required the number of friends to exceed some threshold, we found that even participants with small—or otherwise unusual—networks often still had interesting things to say about privacy.

4.8 Discussion

4.8.1 Challenges of Networked Information

The responsibility for managing privacy risk is generally placed on the shoulders on individuals. Unfortunately, privacy management is too large and too complex a task for any individual to manage alone [296]. Part of the problem lies in information asymmetry. People do not have the time to read the privacy policies for every website they visit and every application they use [215]. Furthermore, it is difficult to reason about the hidden algorithms that process, aggregate, and analyze our data. Even with all the time and information in the world, however,

there are real limits to the agency that individuals have to control their own information. One of these limiting factors is that, in many situations, our data is not only our own. Photos often contain more than one person in the foreground—not to mention the bystanders who may appear in the background [131]. Genetic information from a third cousin can be enough to narrow down the search and identify a wanted individual [85, 106]. In a group chat context, a single compromised device can expose the conversations of the entire group [12]. When information does not belong to a single individual, we should not expect this information to be individually managed. As Harsh pointed out, when third parties—such as Facebook—inject themselves in the middle of these relationships, making their own claims to this shared content, any efforts to assert control grow increasingly futile. This phenomenon is of course not unique to Facebook; for example, Reddit has recently entered an arrangement with Google to allow users’ content to be used to train machine learning models [322]. While a tool like FIG can start to address the information asymmetry between data subjects and data collectors, these other challenges remain. New tools and strategies are needed to support *collective* rather than *individual* privacy management.

4.8.2 Facilitating Privacy Negotiation

FIG’s design highlights the limitations of individualistic understandings of privacy. Even if a given individual chooses not to share their workplace, religion, etc., these attributes can often be inferred from information their friends have shared. Academics have critiqued the reliance on individual consent in the context of “big data” [296, 158, 324], calling for a greater focus on “rights and obligations” [25]. In other words, the burden of responsibility currently placed on individuals is untenable—some shifting of responsibility onto the shoulders of data collectors is necessary.

Platform features designed to facilitate collective privacy management could address some challenges that social media users face [211, 319]. Presently, Facebook users frequently rely on social norms to determine whether it is appropriate to post content about a friend, but

conflict can arise from disagreement about these norms. For example, Val felt it was acceptable to post photos of her grandchildren on Facebook, since “*everybody has pictures of their children having fun.*” Yet her daughter-in-law expected others to ask permission before sharing photos of her children. Similarly, Blake’s sister asks permission before posting photos of others, whereas Blake “*respect[s] it*” yet sees this as “*a lot of energy.*” Furthermore, these norms change over time [337]. Some participants in our study mentioned that their posting behavior on Facebook had changed over the years, at times causing regret. Social media platforms could offer more support for managing and communicating norms in online communities. For example, platforms could implement features that facilitate giving feedback when a user violates a friend’s privacy [262]. This approach helps empower users to construct and enforce community norms around privacy.

A number of additional platform features have been proposed for better facilitating collective privacy management. In particular, features designed to prevent privacy conflicts may be more useful than features to address conflict after the fact. For example, as mentioned by participants in our study, it is possible to untag yourself from a Facebook post, but not to block others from tagging you in the first place. Prior work has found that concerns about photo tagging can lead people to alter their behavior, so as not to be photographed in the first place; in extreme cases, conflicts over photo tagging have even led Facebook friends to cut ties [360]. A mechanism to prevent tagging before it happens could help prevent these kinds of conflicts. This kind of obstructive feature might make privacy management easier, since users could set up their preferences ahead of time rather than responding to conflicts as they arise. Prior work has proposed automated systems that incorporate individual preferences to develop group-level policies [301]. The automation approach is appealing in that it has the potential to grant users greater control over their privacy without imposing a substantial burden. Another potential intervention might be to use facial recognition to identify the people in photos before they are posted and remind the poster to ask for their consent [238]—though facial recognition algorithms pose their own threats to privacy and may do more harm than good [304]. An alternative would be forgo facial recognition and simply prompt users to seek permission every time they post a

photo—regardless of who is pictured.

4.8.3 Avoiding Privacy Theater

New privacy features would need to be carefully designed, particularly since existing privacy tools and settings on Facebook have well-documented usability issues [144, 198, 126]. As in our study, prior work has found that social media users often rely on multiple accounts across various platforms to manage who sees what content—even though platforms like Facebook offer built-in tools for audience management [374, 68, 312]. One reason that users may turn to these kinds of workarounds is that they are confused about platforms’ existing privacy features [77]. Introducing new privacy features without investing in improving the usability and discoverability of these features would only offer privacy theater—the illusion of privacy rather than meaningful improvements. Prior work on Facebook’s advertising controls found that “contextual menus located directly within an advertisement can effectively supplement advertising controls within settings pages” [126]. In a similar vein, contextual settings menus for tagged content are an important opportunity to improve the usability and discoverability of features for collective privacy management.

4.8.4 Utility of Probe Tools

Our findings suggest that probe tools can be useful for understanding and facilitating reasoning about algorithms—in particular, we find evidence that using real data from participants’ own social networks facilitates participants’ reasoning about inferences. Therefore, even though privacy negotiation on social media has been studied extensively, FIG supports deeper reflection from participants and offers fresh insights. Stable APIs or consistent data export formats would make it easier to build educational tools to help people reason about inferential privacy risks. Such tools could even help guide conversations among friends, family, or community members about how to address collective privacy concerns. Building tools that help communities understand inferential privacy risks may even be a strategy for supporting collective action on privacy issues.

Many researchers have advocated for collective action related to data collection and privacy issues [76, 362, 59, 189, 344, 342, 143, 44]. A number of factors make collective action in this context challenging. One challenge is a lack of transparency. Tisné [321] explains:

Perhaps urgency has been lacking so far because the nature of the collective harms — much like CO2 pollution — is invisible to the average person. Algorithms are cloaked in secrecy, their effects omnipresent but invisible [...] Collective action is therefore also less likely to take place.

Tools like FIG offer one way to demystify these invisible algorithms. Indeed, FIG might be expanded as a transparency-enhancing tool to provide additional insights about the kinds of data Facebook is collecting and the kinds of inference Facebook is making. Tools that both provide transparency and suggest opportunities for action are particularly compelling. For example, AdNauseam is an ad-blocking browser extension; in addition to “*providing insight into the algorithmic profiles created by advertising networks*”, AdNauseam also simulates clicking on ads as a form of sabotage—seeking to drown out any signal in the noise and reduce the effectiveness of targeted advertising [143].

Another approach for offering better transparency is through algorithmic audits. For example, auditing studies have uncovered how Facebook’s advertising platform could be used to uncover sensitive information (such as phone numbers) about individuals [341] or to enable discrimination [13]. These audits can help chip away at the information asymmetry between data collectors and data subjects. Everyday users of social media sometimes draw attention to algorithmic harms through informal audits [284]. Developing new tools to support this kind of participatory auditing is another way to encourage collective action.

Another challenge is overcoming feelings of resignation. Inspiring collective action requires us to “*disrupt the rhetorical strategies, tactics, and technical tools that industries use to [...] convince the public about the inevitability of data use*” [76]. In other words, increasing transparency and awareness of digital surveillance is not enough, since greater awareness can increase resignation [275, 326]. We saw this in our own study when participants like

Casey expressed the idea that digital surveillance is inevitable—“*this is just how the internet works.*” Inspiring and scaffolding collective action thus requires attending to the emotions people experience surrounding these issues [264, 362]. Again, tools that offer both increased transparency and a route for collective resistance appear particularly promising. The sense of belonging to a larger community can mitigate feelings of helplessness [264].

4.9 Conclusion

Through interviews conducted using FIG, we investigate how participants perceive and navigate privacy interdependence on Facebook. Participants explained that different friends would respond differently to various kinds of inferences. For example, Florencia explained that two of her family members would view certain inferences as inaccurate stereotypes. Yet she anticipated that one would be amused by this mistake while the other would likely be offended. We also observed a range of privacy requests that participants fielded from friends. While participants sought to comply with these requests out of respect for their friends’ autonomy, these requests nonetheless still sometimes involved tension and conflict. We point to the need for better tools for managing privacy collectively and for scaffolding collective action related to privacy.

Acknowledgements

We thank everyone who helped with this project by providing advice, volunteering as pilot testers, etc. We especially thank Dzhangir Bayandarov, Casey Meehan, and Lu Sun.

The material in this chapter has been submitted for publication. Smart, M. A., Broukhim, A., Sekhon, B., Tan, S., and Vaccaro, K. Negotiating Privacy Interdependence on Facebook. The dissertation author was the primary author of this paper.

Chapter 5

Selling Privacy: Privacy Conceptualizations Promoted Through Advertisements

5.1 Introduction

Although the word privacy does not have a singular, universally agreed-upon meaning [295, 20, 122], tech companies often benefit from framing privacy in a narrow, individualistic way. Tech companies' rhetoric can “*influence [...] beliefs and general assumptions held by the public*” about privacy [49]. For example, Facebook has “*reconceptualize[d] privacy within a rhetoric of control*” [103]. Through this framing, Facebook's privacy issues can be solved by introducing additional control settings—no reexamination of the underlying business model is necessary. A related strategy adopted by tech companies like Microsoft is to frame privacy as a purely technical problem, which they are well equipped to solve with their own technical expertise [302]. This framing shields Microsoft from privacy-related critiques that might threaten Microsoft's business model.

One of the ways that tech companies promote these narratives and communicate to members of the general public about privacy is through advertisements. Several big tech companies have launched major ad campaigns centering around privacy. For example, Apple launched their “Privacy. That's iPhone.” campaign in 2019 [364], and WhatsApp launched a campaign touting the benefits of end-to-end encryption in 2021 [120]. But despite their prevalence, privacy-related advertisements remain an understudied form of corporate communication. Prior work has studied

public statements, blog posts, and other web pages from tech companies [139, 103, 302, 276]. Akgul et al. studied YouTube influencer advertisements for VPNs, finding that many advertisements contain misleading claims [9]. In our work, we analyze privacy-related videos more broadly, uncovering underlying assumptions, identifying rhetorical strategies, and interrogating the political dimensions of the privacy conceptualizations communicated in these advertisements.

We study privacy-related videos from six tech companies to better understand how these companies promote certain ideas about privacy through their advertising. Our dataset consists of 61 YouTube videos from six companies: three ‘big tech’ companies—Apple, Google, and WhatsApp (owned by Meta)—as well as three privacy or security-focused companies—DuckDuckGo, NordVPN, and Brave. We conducted a content analysis of the full dataset, which informed our selection of a subset of six videos for critical discourse analysis. We analyze these videos in more depth. In particular, we study how these companies 1) represent victims, users, or viewers; 2) conceptualize and represent adversaries; and 3) seek to position themselves in relationship to their users.

The six companies we study use a range of techniques to construct narratives about privacy that benefit their bottom line, and we reflect on the limitations of these narratives. For example, the narrative that everyone is a potential victim evokes viewers’ fears and encourages them to purchase products or services that can protect them from such ever-present threats. Stereotypical portrayals of hackers and cybercriminals similarly call upon viewers’ fears while keeping the details of who these adversaries are intentionally vague. Finally, claiming to offer viewers stronger privacy through control and choice presents the illusion of meaningful privacy without threatening the underlying business model that requires mass collection of user data [103, 380]. Although the narratives promoted in these advertisements can in some cases be harmful, some advertisements nevertheless offer insights that may be useful for privacy education and interaction design.

5.2 Related Work

5.2.1 Politics of Privacy-Enhancing Technologies

Privacy-enhancing technologies (PETs) are more than mere technical tools—PETs also have a political dimension [272]. If PETs are political, then companies’ choices about which PETs to deploy and advertise are political as well. For example, in 2016, WhatsApp completed their project of ensuring all messages on the platform would be end-to-end encrypted by default [150]. This did not occur in a vacuum. WhatsApp had faced a national blackout in Brazil after refusing to hand suspects’ messages over to the Brazilian government; Santos and Faure argue that the launch of end-to-end encryption on WhatsApp was a “*techno-political statement*” in the context of an ongoing “*power struggle between corporate and state power*” [276]. As tech companies like WhatsApp invest their time and money in marketing PETs, it is worth investigating what these advertisements have to say about privacy and power.

5.2.2 Critical Discourse Analysis

In recent years, scholars have called for greater attention to power dynamics in privacy studies [209, 269]. We examine power dynamics using multimodal critical discourse analysis; critical discourse analysis allows us to “*reveal [...] ideas, absences and taken-for-granted assumptions*” in privacy-related advertisements and, in turn, to reveal “*the kinds of power interests buried in*” these advertisements [213]. We are not the first to bring critical discourse analysis to privacy studies. Greene and Shilton have studied how privacy is defined and legitimized by mobile app developers in their studies of developer forums [285, 119]. Stahl studied “*the question of how ideological assumptions shape discourses on privacy and security and vice versa,*” finding that Microsoft’s messaging reduces privacy to an engineering problem, and thus, a “*matter of technical expertise, which MS has without doubt*” [302]. In our work, we similarly employ critical discourse analysis to understand how privacy is conceptualized—in our case, focusing on advertisements.

5.3 Methods

5.3.1 Data Collection

We chose to collect and analyze videos from six companies’ YouTube channels. Different companies advertise in different places—for example, some companies erect billboards, some advertise on Facebook, and others air advertisements on television. Most tech companies, however, have a YouTube channel through which they share promotional content. We chose six companies in particular: three big tech companies—WhatsApp (owned by Meta), Google, and Apple—as well as three companies that explicitly focus on privacy or security—NordVPN (a popular virtual private network service), DuckDuckGo (best known for their search engine), and Brave Software (best known for their web browser).

In March 2023, we used the YouTube API [1] and youtube-dl [2] to collect and download 61 videos from these six companies. Two of these videos were duplicates of other videos in the dataset, merely posted with different titles. We included short, recent, English-language videos about privacy from the aforementioned six YouTube channels in our dataset. Specifically, we include videos with the words “privacy” or “private” in the title or description. We focus on a five-year window and exclude any videos posted before 2019—a big year for privacy in the news, since this was the year that the New York Times introduced their privacy project [318] and the year of a historic FTC settlement with Facebook [329]. Since we focus on short, promotional videos, we excluded videos that were longer than six minutes in length. This allows for the inclusion of promotional videos that are somewhat longer than traditional advertisements but excludes lengthier videos such as podcasts or conference keynotes. Some of the videos in our dataset are high production-value advertisements that also aired on television; others are clearly designed for social media. The view counts vary, but all six of our selected YouTube channels have at least tens of thousands of subscribers, indicating substantial reach. Some videos were also posted on other platforms or aired on television.

5.3.2 Analysis

Our analysis consisted of two phases. The first phase involved a content analysis of the entire dataset, with a focus on broad patterns of the advertisements. A second phase employed multimodal critical discourse analysis to identify *how* companies represent themselves, adversaries, and consumers with a focus on the political dimensions of the privacy conceptualizations they endorse. In reporting our results, we focus on the critical discourse analysis; we include the results of the initial content analysis—which informed the subsequent analysis—in our supplementary materials.

Content Analysis

The content analysis used an iterative approach to identify and quantify broad patterns in the data [35]. In the first pass through the dataset, we conducted a deductive coding using Solove’s taxonomy of privacy harms [294]—a widely-adopted taxonomy in privacy research [241, 104, 176]. The segment of analysis is a single video, and each video could be assigned multiple codes. For each harm that appears in a given video, we record at least one timestamp. As a first step, to ensure that the coders had a shared understanding of the taxonomy and the coding process, three authors coded 16 videos together—noting when any of the harms occurred in each video, while taking notes on other interesting features of the videos. Then two of the authors separately coded and took notes on the remaining videos. The original group of three reconvened at regular intervals to discuss and resolve any disagreements. This first pass helped us form our initial understanding of how privacy was being conceptualized in these videos.

In our second pass, three authors took notes on three dimensions for each video: 1) the representations of users, victims, or viewers; 2) the representations of adversaries (e.g., hackers); and 3) the representation of the company’s relationship to its users. Analyzing the representation of social actors is an important part of critical discourse analysis [339]; we chose these three dimensions to encompass what we felt were the most important social actors in all of our videos. We group users, victims, and viewers, because they are often grouped in the videos themselves.

That is, the viewer may be encouraged to see themselves as a potential user of some product or victim of some threat. Likewise, someone who at first is portrayed as a victim of some threat may later be protected against this threat by becoming a user of some product.

Discourse Analysis

For the second phase, we employed critical discourse analysis. At its core, critical discourse analysis exposes “*strategies that appear normal or neutral on the surface but which may in fact be ideological and seek to shape the representations of events and persons for particular ends*” [213]. In our case, we seek to expose the strategies used in video advertisements that promote certain narratives about privacy that further the aims of tech companies.

For each of the three dimensions outlined above, we selected one video that was in some way typical or representative and one video that was atypical or noteworthy in some way; the first phase of analysis guided this selection. We then conducted a deeper analysis of each of these six videos, employing critical discourse analysis. Focusing on a small subset of our data allows us to think deeply about the decisions made in each video. Why does the narrator address the viewer directly in the second person instead of discussing their users in third person? Why were particular choices made about how characters are dressed or positioned? Even choices that may be dismissed as commonsense are worthy of study, as “*commonsense practices are the most deeply ideological of all*” [339].

5.4 Results

The videos in our dataset depict a wide range of privacy harms, though surveillance is particularly common [294]. Even when discussing the same harms, however, videos may differ in their representation of adversaries and victims as well as in the way the company is positioned in relation to its users. These choices about representation are used to construct narratives about privacy that align with the interests of the companies we study. Below we describe our in-depth analysis of six videos selected as case studies.



Figure 5.1. Diverse genders and racial backgrounds in “A New Era of Personal Privacy with Default End-to-Encryption”.

5.4.1 Viewers, Users, and Victims

We group viewers, users, and victims because they are often grouped in the videos we study. For example, some videos show how users of a product may take advantage of that product’s features prevent themselves from becoming victims of a privacy threat. Viewers of these videos may be encouraged to relate to the victims or users depicted on screen. A frequent strategy in our advertisements is to appeal to viewers’ fears by arguing that everyone is a potential victim—and thus that everyone could benefit by becoming a user of the advertised product. The first video we analyze is an exemplar of this narrative. The second video we analyze takes a quite different approach, portraying their users as an exclusive in-group.

Everyone is a potential victim.

A common strategy for portraying users or victims is to portray multiple individuals of different backgrounds. The underlying message is that everyone is a potential victim of some kind of threat and thus that everyone could benefit from the advertised product. WhatsApp’s video, “A New Era of Personal Privacy with Default End-to-Encryption,” exemplifies this strategy. The overarching narrative is that various individuals enter what appears to be a post office in order to send messages. To the shock and dismay of these individuals, their messages are attached to passenger pigeons. The video then makes a comparison between this insecure “pigeon-mail” and unencrypted messaging. The text on the screen at the end of the video reads: *“5.5 billion unencrypted texts are sent every day. With WhatsApp, your messages won’t be one of them.”* By showing us several different individuals of different genders and racial backgrounds (Figure 5.1)

and by telling us that billions of unencrypted texts are sent every day, the video emphasizes the fact that WhatsApp can help anyone and everyone communicate more securely. On the one hand, this maximizes the size of WhatsApp's potential user base; on the other hand, this disguises the unequal burden of surveillance faced by some groups.

When the characters hear that most messaging is unencrypted, they express surprise, dismay, and distress. While it is true that the insecurity of SMS texting affects billions of people, this insecurity does not affect everyone equally. This video does not directly discuss the fact that “*surveillance disproportionately affects marginalized communities and peoples,*” meaning that privacy risks are unevenly distributed [124]. Through its portrayal of potential victims, WhatsApp's advertisement disguises this unequal distribution. Marwick argues that it is indeed “*in the entire technology industry's best interest to keep these [power relations] obscured*” [209]. Obscuring these power relations allows companies to present privacy and security as matters of individual responsibility, offering their own products and services as the responsible choice.

One interesting moment in the video calls attention to the interdependent nature of privacy. In reference to a man standing in line, the office employee says, “*That SMS text that he's sending? It's unencrypted.*” The man in question is holding a framed photograph of a child while sending his text message. This scene reminds viewers that an unencrypted message implicates not only the sender but also the receiver. While companies often appeal to individualistic notions of privacy to serve their interests, in this case, WhatsApp appeals to viewers' concerns not only about their own privacy, but also that of their friends and family.

Our users are in-the-know.

One of NordVPN's videos (“I am a fighter, but you gotta help me here, guys! Online privacy is no joke #shorts”) offers a very different understanding of who their users are. This video—which was also cross-posted on TikTok and Instagram—follows the format of a popular meme that is based on the comments of former UK Prime Minister, Liz Truss. Liz Truss is the shortest-serving prime minister in British history and was mocked for publicly proclaiming

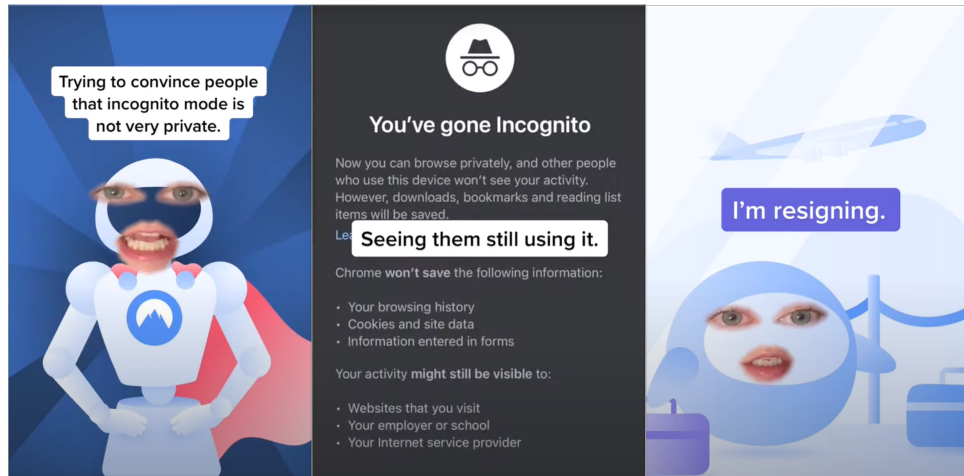


Figure 5.2. The three frames from NordVPN’s “I am a fighter, but you gotta help me here, guys! Online privacy is no joke #shorts”.

herself a “*fighter and not a quitter*” only to announce her resignation the following day [256, 350]—this is the origin of the popular meme. The first frame of the video portrays NordVPN as a superhero—a robot branded with NordVPN’s logo stands with hands on hips, wearing a red cape (Figure 5.2). Someone’s mouth and eyes are superimposed over the robot, in the style of a popular TikTok filter. Text on the screen reads: “*Trying to convince people that incognito mode is not very private.*” The audio that plays is a soundbite of Liz Truss: “*I am a fighter and not a quitter.*” The next frame shows a screenshot of an incognito tab in Chrome, with overlaid text that reads “*Seeing them still using it.*” In the final frame, a figure stands in line holding a bag—seemingly waiting to board a plane. Again, someone’s eyes and mouth are superimposed over the figure. A second Truss soundbite plays: “*I am resigning.*” The overlaid text reads: “*Seeing them still using it.*”

The overarching message is that people who rely on incognito mode to protect their privacy are ignorant at best or unintelligent at worst. One of the functions of humor is to “*make both alliances and distinctions,*” often by nurturing “*feeling[s] of superiority over those being ridiculed*” and feelings of unity among those doing the ridiculing [223]. The viewer is encouraged to feel “in on the joke” at the expense of people who use incognito mode. The video suggests that NordVPN users are part of an exclusive in-group of people who are privacy-

conscious and tech-savvy. The blame for any invasions of privacy that occur in incognito mode is placed squarely on the individual user, who persists in using inadequate protection strategies despite the ambiguous educational efforts of NordVPN. The video does not place any blame on Chrome for misleading its users, although such blame would not be misplaced. A class-action lawsuit filed in 2020 accused Google of misleading users by gathering data even when users had turned on incognito mode; in response, Google has recently agreed to delete billions of data records [42]. Prior work has documented a range of common reasons that people use private browsing; some of these reasons (e.g., hiding activity from other users of shared devices) align with the actual functionality of private browsing, while others (e.g., protecting against malware or hackers) reveal misconceptions [125, 361]. This video ignores the fact the private browsing may indeed be sufficient for some users' needs—dismissing all users of private browsing as ignorant and positioning NordVPN's tech-savvy users as members of an exclusive in-group.

5.4.2 Adversaries

Rogaway has criticized “*cutesy*” depictions of adversaries in the cryptography community, arguing that such representations shape the way cryptographers identify important problems in their field [272]. He observes that “*the adversary as a \$53-billion-a-year military-industrial-surveillance complex and the adversary as a red-devil-with-horns induce entirely different thought processes.*” Similarly, the way that adversaries are depicted in advertisements can shape the public's understanding of the importance of different kinds of privacy threats. As case studies, we select 1) a video that draws upon stereotypes in its depiction of cybercriminals and 2) a video that uses the metaphor of an auction house to draw attention to issues with online tracking and targeted advertising.

Hackers and Snoopers

Stereotypical depictions of hackers and cybercriminals abound in our set of videos, particularly in the NordVPN videos. The video titled “7 Tips to Secure your Wi-Fi Router —

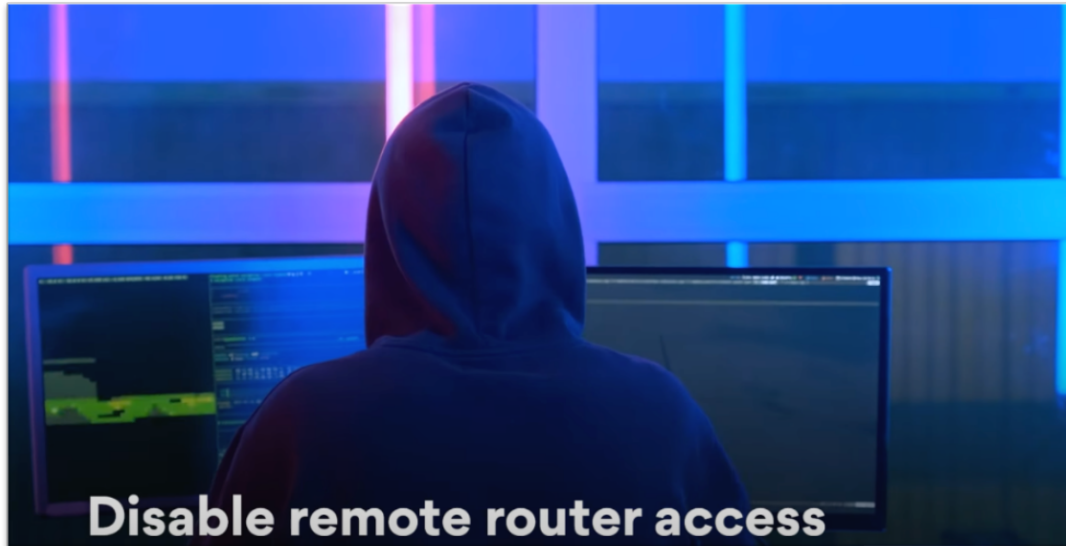


Figure 5.3. Stereotypical hacker depicted in “7 Tips to Secure your Wi-Fi Router — NordVPN”.

NordVPN” offers a typical example. This video describes several ways to enhance the security of one’s router, including installing a VPN. One of the tips in the video is to “*disable remote router access,*” explaining that remote router access “*leaves your router’s backdoor open to hackers.*” The video depicts a person in a black hoodie, facing away from the camera so that we cannot see the person’s face (Figure 5.3). The hidden face serves at least two purposes. One, the fact that this individual keeps their face hidden suggests malicious intent and encourages feelings of suspicion. Two, the lack of any individuating characteristics solidifies this figure as an instance of a generic archetype. The individual has two displays set up, with activity happening on both screens at the same time—implying technical sophistication and suggesting that equally sophisticated tools (i.e., those offered by NordVPN) are needed to protect against this kind of attacker. This stereotypical depiction of the lone cybercriminal is of course only one kind of hacker. In some cases, hackers may be operating in a larger team—for example, as part of a national or criminal organization. Such a hacker may very well be wearing a suit and tie while working in an office cubicle. Depicting a large, organized team of hackers, however, may overwhelm viewers and leave them feeling helpless against such a powerful threat. NordVPN needs viewers to believe that their products are capable of delivering protection; thus, the depicted adversary must be

threatening enough that the viewer needs NordVPN's help, yet not so threatening as to leave the viewer skeptical of NordVPN's effectiveness.

Later, the video suggests disabling Wi-Fi Protected Setup (WPS), since WPS PINs can be recovered through brute-force attacks. Again, the video depicts a hacker implementing such a brute-force attack. This depiction is somewhat atypical in that it actually displays part of the hacker's face. The video is zoomed on the hacker's glasses, which show a reflection of the computer screen where the hacker is trying out different passcodes. The rapid sequences that we see flashing across the screen are typical of dramatized hacking scenes in movies. Finally, the third depiction of a cybercriminal accompanies a recommendation to change the router's frequency band to 5 GHz. The video explains that "*if there's a criminal lurking around your neighborhood, they won't be able to sniff out your wifi*" once you've changed your router's frequency. Here the "criminal" is depicted again in a dark hoodie, facing away from the viewer. Again, NordVPN chose this neighborhood criminal as the face of Wi-Fi sniffing attacks, but there are other ways that this adversary could have been depicted. For example, in 2010, Google was caught using Street View vehicles for illegal Wi-Fi sniffing [175]. The "lurking cybercriminal," however, is a clear-cut enemy, whereas viewers' relationships and perceptions of Google may not be so black-and-white.

In addition to the depictions of hackers and criminals described above, the video mentions a few other potential adversaries. For example, the video mentions setting up a guest network, and shows what appears to be a group of friends huddled together on their phones and computers. In this case, the adversary is someone you invite over to your home. As is typical of many of our videos, this video also includes vague references to "*intruders*" or "*snoopers*." Being vague about who the adversary is makes the threat more abstract, but it also suggests that the protection being offered is universal—able to protect against any adversary.



Figure 5.4. The auctioneer gestures toward a hologram of Ellie in “Privacy on iPhone — Data Auction — Apple”.

Advertisers and Data Brokers

Apple’s video, “Privacy on iPhone — Data Auction — Apple, ” depicts a distinctly different set of adversaries. Instead of talking about hackers or cybercriminals, this video focuses on data brokers and advertisers. The video’s protagonist is a young woman named Ellie. She spots a door labeled “ELLIE’S DATA AUCTION”, and ominous music plays as she approaches the door. She opens the door to an ongoing auction. The auctioneer is an older man dressed in a suit, who stands up on a stage in front of the larger audience. The name of the auction house is “DUBIOUS,” suggesting that the legitimacy of the ongoing auction is dubious. A hologram of Ellie stands on the front of the stage on an auction block (Figure 5.4). Various aspects of Ellie’s life are then auctioned off, such as her “*wonderfully personal*” emails, her drugstore purchases (as her medicine cabinet appears on stage), her location data, and her contacts—including “*sweet Nana*” (Figure 5.5). Near the end of the video, Ellie opens her iPhone and sees a choice to click “*Ask App Not to Track*” or “*Allow*”. When she selects “*Ask App Not to Track*”, everyone and everything in the auction house begins to disappear in puffs of smoke.



Figure 5.5. “Sweet Nana” sits on the auction stage.

The adversaries in this video are the auctioneer and the buyers, representing data brokers and advertisers. Both the auctioneer and the buyers appear wealthy, as they are well-dressed, wearing nice jewelry, or, in the case of one buyer, using opera glasses to better see the stage. These displays of wealth reinforce the fact that the adversaries seek to profit from Ellie’s data. Most of the buyers, like the auctioneer, appear significantly older than Ellie. These differences in class and age serve to create distance between Ellie and the adversaries. Over the course of the video, Ellie and the adversaries do not directly interact with each other. Ellie appears confused when she first enters the auction house, suggesting that she does not know the people in the room who are bidding on her data. Furthermore, none of the buyers acknowledge Ellie as she stands watching near the door. While it is sometimes argued that targeted advertising benefits consumers by presenting them with more relevant advertisements, this video contradicts that narrative. None of the buyers care about Ellie or even bother to acknowledge her when she enters the room—they only want her data.

In two particularly dramatic moments, a person stands on stage to be auctioned. In the first instance, a hologram of Ellie stands before the audience atop an auction block. In the second instance, “*sweet Nana*” appears sitting in an armchair, looking confused as if she has been snatched up from her living room and deposited on stage (Figure 5.5). Both of these moments are

reminiscent of slave auctions, although depicting Nana sitting in an armchair and depicting Ellie as a hologram are choices that create a bit of distance from this history and make the reference to slave auctions indirect. Nevertheless, this subcontext highlights the same point being made by naming the auction house “*Dubiou’s*.” While the auctioneer at one point declares, “*It’s not creepy—it’s commerce!*”, the video nonetheless paints data auctions as distinctly “*creepy*.”

On the one hand, the extended metaphor of the physical auction house is useful; it helps viewers understand the notoriously complicated digital advertising ecosystem and the privacy issues involved. On the other hand, the video oversells the effectiveness of Apple’s privacy features in their failure to acknowledge the various workarounds available to dedicated adversaries. While the “Ask App Not to Track” feature can prevent application developers from accessing system advertising identifiers (IDFAs), other forms of tracking and targeting are still possible [178]. In fact, Apple itself has been accused of tracking its users—in Apple’s Stocks app, for example—even when users had selected the “Ask App Not to Track” option [283, 113]. While it would be nice for all the trackers to disappear in a puff of smoke at the click of a button, the reality is more complicated.

5.4.3 Companies

Finally, we examine how the companies that produced these videos position themselves in relation to their users. Many videos convey the message that the company empowers or protects their users. As a case study, we analyze one of Google’s videos that exemplifies this empowerment narrative. As the second case study, we analyze a video that positions the role of the company quite differently; rather than empowering its users, DuckDuckGo (DDG) makes the case that it leave its users alone.

We empower our users.

In many of our videos, the company positions itself as empowering their users to protect themselves against some privacy threat. Google’s “Guest Mode on Google Assistant” is one

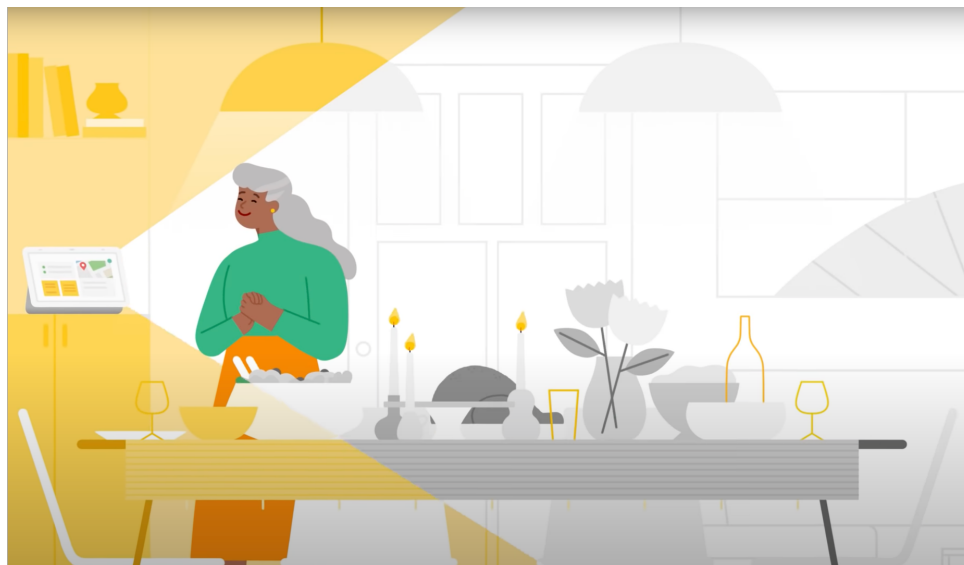


Figure 5.6. The colors change from yellow to black-and-white when guest mode is activated in “Guest Mode on Google Assistant”.

such video. The video follows three different animated characters who employ privacy features on their devices for different reasons. A woman activates guest mode as she prepares to have guests over for dinner. A man activates guest mode before asking Google Assistant where to buy an engagement ring. Finally, a man asks his Google Assistant how many points you get for a touch down and then requests that this question be deleted, seemingly embarrassed to have asked this question in front of his friends.

Control appears as a recurring theme throughout the video’s narration. For example, the narrator says (emphasis ours):

*With the new, easy-to-use Guest Mode, you have even more **control** over how your data is used and saved in your Google account. [...] And now with Guest Mode and other privacy actions available on Google speakers and displays, we’re giving you even more simple tools to **control** your privacy and data.*

The passive voice in the above fragment—“*how your data is used and saved*”—hides Google as an actor with agency in how data is managed. The visual animations also serve to support the point that the user has the ultimate agency. When the woman in the video activates guest mode, the color gradually leeches from the setting—a yellow background fading to black and white

(Figure 5.6). The cheerful color palette used to represent the area under Google’s surveillance contrasts sharply with the dark blues pervading NordVPN’s color palette in their portrayal of cybercriminals (Figure 5.3)—Google’s surveillance is depicted as benign and even friendly. As in Apple’s video, where Ellie makes the auction participants disappear into puffs of smoke, this leeching of color serves to visceralize the user’s control and agency. The colors fading to black and white signify that the Google Assistant is no longer “watching.” The implication is that Google’s surveillance is useful and, like color, enhances the user’s home—in its absence, something important is missing.

One mechanism that Google offers for empowering users is choice. The narrator explains that with Guest Mode, “*it’s even easier to make privacy choices that are best for you.*” In practice, however, more choice does not necessarily mean more control. For example, offering choice can mislead individuals into overestimating the amount of control they have over a situation [185, 188]. Furthermore, offering too many choices can lead to *choice overload* and feelings of frustration [151]. Indeed, many tech companies have introduced new privacy settings and features that claim to offer choice and control but, in fact, suffer from serious usability issues [7, 198, 204, 112, 206].

Scholars have often criticized this conception of privacy as “control of information.” The first problem is that “*defining the concept of privacy in terms of individual control of information [...] greatly reduces what can be private*” [316]. In the age of big data, individuals, in truth, are quite limited in what they can control. Thus, when companies like Google offer improved control through new settings or features, they do “*little to shift the power imbalance between data collectors and the data subjects*” [76]. While it may benefit tech companies to paint privacy as the responsibility of individual users, the truth is that this responsibility is too much for any individual to bear alone [296]. In response to this challenge, scholars have called for a greater focus on “*rights and obligations*” as opposed to “*notice and consent*” [25]. In other words, the burden of responsibility currently placed on individuals is untenable—some shifting of responsibility onto the shoulders of data collectors is necessary.

This video does not view surveillance through smart devices as a problem for which Google offers a solution. On the contrary, Google highlights how they are offering “*even more control*” and making privacy choice “*even easier*.” This repetition of the word “*even*” positions Google as a privacy-conscious company that constantly strives to prioritize privacy, rather than as a company plagued by critical press coverage surrounding issues of privacy. Furthermore, the narrator is quite vague about the nature of the ongoing surveillance. The narrator promises that “*your personalized information will stay private*,” without defining what is meant by “*personalized information*” or “*private*” in this context. This framing depoliticizes the surveillance issue and encourages users “*to allow surveillance as long as it appears to be something they control*” [103].

We leave our users alone.

DDG’s video, “DuckDuckGo: None of Our Business” positions DDG quite differently in relation to their users. This video shows different people engaged in a variety of activities, such as a man carrying a child on his shoulders, a man jumping off a boat, and two children playing with a dog. While we glimpse these snapshots of the characters’ lives, the narrator explains that DDG protects users’ privacy by leaving them alone. The narration begins:

Your life. Your hopes. Your passions. These things are none of our business. Your deodorant rash. Your favorite cheese. And whatever you were searching for at 1:15AM. That’s really none of our business.

The repeated use of the first-person plural to refer to DDG (e.g., “*our business*”) and the second-person singular to refer to the viewer (e.g., “*your privacy*”) creates the illusion of a direct conversation between DDG and the viewer. This imagined relationship between DDG and the viewer suggests a certain level of trust.

The narration continues by explaining why DDG leaves their users alone: “*Because your life is private, and unlike other tech companies, we think your internet should be too. [...] Protect your privacy online for free with DuckDuckGo*”. The video is quite vague about what DDG actually does; for example, we never see an individual actually using DDG on screen nor does

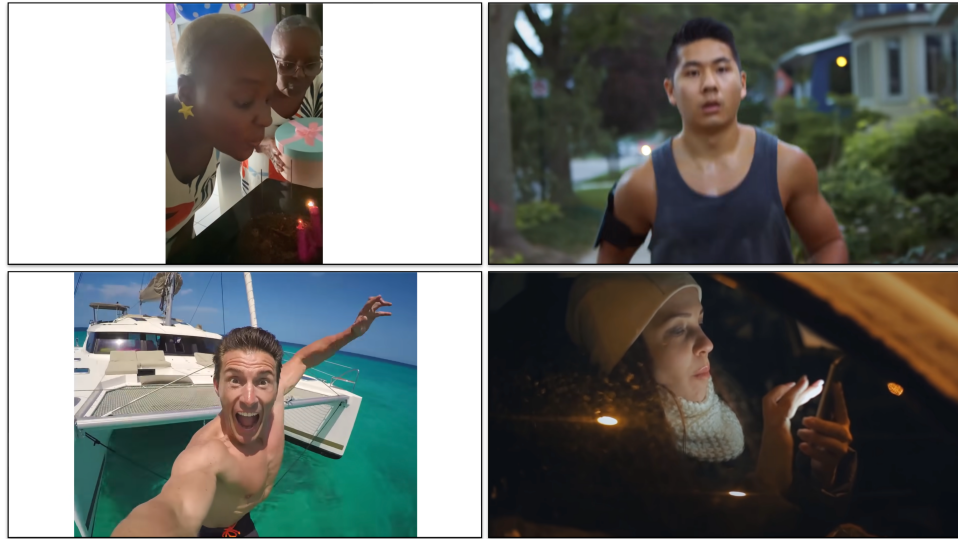


Figure 5.7. Varied aspect ratios in “DuckDuckGo: None of Our Business”.

the narrator discuss any specific features. Rather, the video offers a vague mention of “*private internet*” and a promise that DDG will mind their own business.

While showing people engaged in a variety of activities, the video switches between different aspect ratios to set up a contrast between content that is willingly shared and information that is obtained unwillingly through surveillance (Figure 5.7). For example, one snippet showing people celebrating a birthday switches to a vertical orientation, as if one of the party attendees is taking a video on their phone in order to share it later with friends. Similarly, a snippet of a man jumping off a boat is framed in a square and appears to have been taken as a selfie—as if he plans to share the video on Instagram. In contrast, other scenes have a full-screen orientation, with no interaction between the characters on-screen and the camera. For example, one such scene depicts a woman sitting in a car at night. The overall effect is creepy, as if we are staring into the window of the car to spy on her. Some of the video snippets capture moments that might be embarrassing. For example, we see a man out for a run and hear the narrator say “*your deodorant rash.*” The implication is that this runner may be searching online for advice on how to treat his deodorant rash and that he may consider this search private or embarrassing. In an even more personal moment, we view a woman in bed with her phone, as the narrator says: “*whatever*

you were searching for at 1:15AM.” The woman’s eyes widen in shock, as if she is viewing something particularly scandalous. DDG is promising that “*unlike other tech companies,*” they will refrain from spying on these kinds of personal moments.

The conception of privacy espoused by this advertisement aligns with the idea of privacy as “*the right to be let alone*” [353]. Indeed, in arguing that “*your internet should be*” private, DDG suggests a conceptualization of privacy as a right that they commit to respect. Rather than focusing on *how* they protect privacy—as Google does in the video described above—this video focuses on the *why*. DDG leaves their users alone, because users have a right to private browsing. This is a natural approach for a company like DDG that differentiates itself from its competitors almost exclusively based on its core avoidance of tracking.

DDG insists that they leave their users alone, unlike companies whose business models depend on extensive user tracking. Despite this, DDG does use information from search queries to target advertisements [357]. The difference is that DDG does not retain a history of these search queries and does not track or profile their users the way other companies do. While this video offers a binary understanding of privacy—either users are left alone or they are tracked—the truth lies somewhere in the middle. Nissenbaum has argued that privacy is not about *stopping* the flow of information—rather, privacy is about the *appropriate flow* of information [239]. DDG’s approach is perhaps better understood as a restriction of certain data flows that DDG considers inappropriate. Talking about privacy as if there exists a clear-cut binary, however, keeps matters simple and makes for a more dramatic message.

5.5 Discussion

5.5.1 Harmful Narratives

Tech companies invest a great deal of time and money to develop advertisements that shape public perceptions of privacy and its relationship to technology. In this study, we expose the underlying narratives about privacy that companies tell through advertisements. These narratives

serve the interests of the companies that promote them, but narratives that are beneficial to some may nonetheless be harmful to others.

For example, when companies promote the idea of privacy as an individual responsibility, they are promoting their own products and services. Yet this message harms people who may not have the time, resources, or knowledge to keep up-to-date with the latest PETs and best practices. What would it look like to instead view privacy as a collective responsibility? While the “privacy as individual responsibility” narrative often leaves people feeling overwhelmed or resigned [76, 128], sharing the burden through a collective effort may be energizing [263]. One datapoint in a vacuum is rarely valuable—power lies “big data”—in the aggregation and linking of large datasets; if our most pressing privacy problems, then, are collective, it makes sense that any effective solution must be collective as well [321]. While NordVPN’s video about the inefficacy of incognito mode does build a sense of community, it does so at the expense of people viewed as less tech-savvy. Embracing the idea of collective responsibility might instead require building solidarity with marginalized communities who are disproportionately burdened by surveillance.

This is just one example of how narrow conceptions of privacy that benefit tech companies may constrain the way we think about privacy problems and solutions. Understanding the limits of the stories tech companies tell about privacy and challenging unstated assumptions can open up new ways to think about what privacy means and how to best protect it. Exploring new metaphors, stories, and perspectives on privacy may present opportunities to explore new solutions as well.

5.5.2 Useful Techniques

Some of the techniques used in these videos may nevertheless be useful for privacy education and organizing. Part of what makes many of these advertisements effective is how they connect online privacy invasions that may feel abstract or intangible to people’s lived experiences. Stark has argued that the emotional context of information privacy is too often overlooked in privacy scholarship and “*advocate[s] for making privacy visceral through interaction and*

form-factor design across multiple senses” [303]. While emotional appeals in advertising may be perceived as manipulative, emotional appeals can at times be educational and productive [70]. In the data auction video, Apple depicts Ellie’s “*sweet Nana*” sitting on stage to tug at viewers’ heartstrings. Imagine if this image were displayed anytime an app asked for permission to access a user’s contacts—this would certainly inspire a different kind of reflection and consideration from users who are habituated to clicking through intentionally confusing permissions menus mindlessly. Thus some of the techniques used in these videos may be useful in interaction design or in awareness-raising efforts related to privacy.

The use of metaphors linking online and offline privacy is a technique used in advertisements that may be particularly useful for facilitating reflection and conversation on privacy issues. In our dataset, these metaphors often serve to visceralize or dramatize privacy violations. For example, in “DuckDuckGo: None of Our Business,” online tracking is implicitly compared to peering into a woman’s car at night to spy on her (Figure 5.7). Metaphors can also be explanatory. For example, the metaphor of the physical auction in Apple’s data auction video serves not only to dramatize the privacy issues at play but also to clarify these issues by relating the complicated landscape of digital advertising to something more familiar. Such metaphors can provide launching points for conversations about online privacy norms by drawing provocative comparisons to cultural norms for offline privacy. For example, the scene in which Ellie’s medicine cabinet appears on stage to be auctioned off evokes the cultural norm that snooping around in someone’s medicine cabinet is particularly invasive. This scene asks the question of whether this same norm applies in cyberspace.

5.5.3 Beyond Advertisements

Advertisements are just one form of corporate communication on privacy. Many other kinds of communication remain understudied. For example, job postings—particularly for positions in the growing field of privacy engineering—play a role in constructing a company’s self-definition and communicating privacy-related values. As another example, academic publications

from industry research groups reveal what kinds of privacy issues a company deems important—even choosing whether to present certain problems as privacy issues is a way of promoting particular perspectives on privacy. As a third example, analyzing texts that are aimed at computing students—such as advertisements for internship programs—may be particularly interesting, as these texts play a role in how the next generation of technologists think about privacy.

Acknowledgements

We thank everyone who provided suggestions and feedback on various stages of this project, including Katie Shilton and Lilly Irani.

The material in this chapter has been submitted for publication. Smart, M. A., Li, S., Liu, T., and and Vaccaro, K. Selling Privacy: Privacy Conceptualizations Promoted Through Advertisements. The dissertation author was the primary author of this paper.

Conclusion

At best, poor privacy communication leaves its audience unengaged and overwhelmed. At worst, it can be actively misleading and cause dangerous misconceptions. The current state of privacy communication—through blogs, advertisements, privacy policies, etc.—leaves much room for improvement. In this dissertation, I have critiqued existing privacy communication and offered improved alternatives. Several key tensions have emerged in this work, which point toward promising directions for future research.

Tensions in Privacy Communication

Completeness versus brevity

When explaining any kind of privacy issue or technology, there always exists a tension between the audience's desire for completeness and the competing desire for brevity. A thorough, carefully-written privacy policy will be as useless as one written in intentionally obfuscatory language if it is too long to hold readers' attention [242, 220]; however, a privacy policy that is too brief may fail to convey all the important information [115]. This tension is particularly evident in the case of developing explanations of differential privacy; while many readers appreciate brevity, others want more comprehensive information about differential privacy and how it works. Some of our study participants (Chapter3) suggested creating adaptive explanations that allow interested readers to choose to view additional details. In a similar vein, prior work has suggested personalizing explanations of differential privacy depending on an individual's particular privacy concerns [57]. Further study is needed to explore how adaptive, interactive, or customized

communication could help manage the inherent tradeoff between completeness and brevity.

Accuracy versus simplicity

A related tension exists between accuracy and simplicity. Metaphors can help people leverage familiar concepts to understand new ones, yet such metaphorical mappings are always imperfect. Thus, while metaphors might help people grasp key ideas quickly, they may also create misconceptions. For example, Apple’s video, “Privacy on iPhone — Data Auction — Apple,” uses the metaphor of the auction house to explain the modern digital advertising ecosystem. While this metaphor helps clarify the roles of the key actors—the data broker as the auctioneer, the buyers as advertisers, etc.—it overstates the effectiveness of Apple’s privacy features by underselling the capabilities of these adversaries.

Achieving the right balance between accuracy and simplicity requires reflection on the communicative goals. For example, prior work has developed explanations of privacy and security technologies with the (implicit or explicit) goal of increasing adoption of said technologies [267, 72]. In this case, a perfectly accurate understanding of the technology in question is not necessary; the reader simply needs to understand the technology well enough to appreciate its value. Nonetheless, it is important to ensure that explanations do not encourage dangerous misconceptions that could lead to overtrust.

Generalizability versus specificity

While this dissertation focuses on communication aimed at a general audience, there is value in focusing on specific audiences to design contextually appropriate communication. For example, while generic explanations of differential privacy are convenient in that they can be easily adapted to new contexts, explanations are most useful when they are specifically tailored to a particular context. My work on explaining differential privacy attempts to take a middle road, incorporating essential contextual information while offering opportunities for generalization, although more work is needed to evaluate how well these explanations generalize in practice.

There would also be value, however, in developing context-specific explanations—particularly in participatory approaches that develop explanations alongside the community that is adopting differential privacy. For example, researchers might work alongside statisticians to develop contextually-appropriate explanations that support statisticians in conducting differentially-private data analysis. These explanations would need to include mathematical details in order to be useful for their intended audience, whereas for other audiences, mathematical equations might be more intimidating than elucidating. Audiences whose responses to privacy communication may be particularly worthy of study include software developers, computing students, and policymakers.

Learning from other domains

None of the tensions described above are entirely unique to privacy communication. Similar tensions arise in other contexts, and strategies from other domains, therefore, may also be useful in privacy communication. For example, in the context artificial intelligence, there often exists a tradeoff between model accuracy and interpretability [332]. In other words, while complex models may offer better accuracy, simpler models are often easier for people to understand. At the same time, there is also a tradeoff between the simplicity of an explanation and the *accuracy of the explanation*. In this case, we are concerned not with the accuracy of the model itself but with the accuracy of a reader’s understanding of the model’s behavior. In the context of explainable artificial intelligence, this second notion of accuracy has also been referred to as “*mental model faithfulness*” [3].

A number of important insights from research on explainable and interpretable artificial intelligence are also relevant to privacy communication. For example, just as rigorous measures of cognitive load have been useful in the study of explainable artificial intelligence [3], such measures would be useful and are too often excluded in studies of privacy communication. Another insight is that expert’s intuitions about what makes a model interpretable do not necessarily align with experimental findings [254]; this underscores the importance of rigorous experimental

evaluation with the desired population and the risk of trusting blindly in expert's ideas of what it means for an explanation to be simple, easy-to-understand, or actionable. Future work on privacy communication should take advantage of insights from fields such as explainable artificial intelligence, cognitive psychology, and education.

Future of Privacy

Despite persistent challenges, privacy is not dead. Nevertheless, changes are needed in order to safeguard privacy and rectify power imbalances between data subjects and data collectors. Privacy communication can be manipulative; for example, confusing privacy policies actively encourage disengagement with privacy issues [76]. Yet privacy communication can also be empowering—for example, by helping people better understand privacy risks, privacy-enhancing technologies, privacy news, etc. Improved communication by itself will not solve all our problems, but effective privacy communication complements strategies such as building better technologies, encouraging collective action, or supporting privacy legislation. A combination of strategies, driven by cross-disciplinary collaboration, is better suited for making progress on protecting privacy, curtailing inappropriate data practices, and resolving power imbalances.

Appendix A

A.1 Demographics

Table A.1 displays participant demographics for the study described in Chapter 2. Some participants declined to state some of their demographic information, and participants were permitted to select multiple answers for race/ethnicity. Therefore, the total counts may not match.

Table A.1. Participant Demographics

Demographic Category		# of Participants	
		Survey 1	Survey 2
Age	18-29	70	53
	30-39	93	61
	40-49	55	72
	50-59	46	47
	60-69	64	52
	70+	34	61
Gender	Male	144	169
	Female	218	176
	Non-binary/third gender	2	1
Race/Ethnicity	White	245	244
	Hispanic, Latino or Spanish	64	63
	Black or African American	49	40
	Asian	22	12
	American Indian or Alaska Native	5	3
	Other	2	5
Education	High school degree or less	152	151
	Technical, trade or vocational school after high school	13	16
	Some college, no degree	63	44
	2 year degree	29	29
	4 year degree or more	106	107
Income	Less than \$25,000	96	68
	\$25,000 - \$74,999	174	177
	\$75,000 or more	86	88

A.2 Replication Study

A.2.1 Experimental Design

The second experiment described in the main paper investigated the relationship between different explanations of differential privacy, different choices for the privacy parameter, and participants' willingness to share their information. We were surprised to find both that 1) most participants did not change their willingness to share after reading the explanation of privacy protections and 2) many participants were willing to share their information even without any privacy protection. These findings suggest that most people are not too concerned about the privacy of their browsing history data. Nevertheless, for the small set of users who were concerned about sharing this data with us—namely, those who were not willing to share before reading about the privacy protection we would offer—we noticed an interesting pattern. None of the participants in this subgroup who saw our explanation in the low privacy setting were willing to share their information. The size of this subset was too small to draw meaningful conclusions, so we decided to replicate a streamlined version of the original experiment with many more participants. This would allow us to get a clearer picture of the subset of people who are not initially willing to share their browsing data. For this replication study, we had the exact same 3^2 factorial design with three explanations and three different privacy levels.

Recruitment

In order to minimize overhead costs, we recruited participants through Prolific instead of using Qualtrics. We intended to recruit 1500 participants, expecting that only about 540 (about 60 in each of the 9 conditions) of these would be initially unwilling to share their information and that only 360 would change their minds (from no to yes or yes to no) after reading the explanation of privacy protections. Our sample was designed to be representative of the US population in terms of age, sex, and ethnicity; Prolific collects demographic information about participants on

the platform and uses this information to form representative samples. We ended up with a total of 1505 participants with an average survey completion time of less than 5 minutes. Participants were paid \$0.60 for completing the survey—about \$9 per hour on average.

Experimental Protocol

In order to collect information from over 1000 participants, we needed to streamline the survey; therefore, some elements of the original experiment were cut from this version.

First, the survey explains the general concept of internet browsing histories and prompts participants to reflect on their browser usage. Then participants are asked how harmful it would be if this information were leaked—of the two sensitivity questions in the original survey, this one is kept because it had higher variance. Next, participants are asked whether or not they would be willing to share this data with us.

Next, participants read the explanation of privacy protections. They cannot advance until they have spent at least 45 seconds reading the explanation. An attention check follows. Participants must answer two easy questions related to the explanation. They may reread the explanation if necessary. If both questions are not answered correctly on the first try, they are given a second attempt before being screened out of the survey.

Next, participants are asked again whether or not they are willing to share their browsing data, after having learned about the privacy protection that would be offered. Participants were reminded that they would be able to complete the survey regardless of their answer and that if they selected “yes,” the upload of their data would occur at the end of the survey.

Next, participants are asked to rate their agreement with the statement: “I’m concerned that online companies are collecting too much personal information about me.” Of the IUIPC questions, this one had the highest variance; it is also directly related to data collection. Finally, participants share their demographic information.

A.2.2 Analysis

We fit a logistic regression model that models willingness to share as a function of 1) the choice of explanation, 2) the choice of privacy parameter ϵ , 3) the interaction between these two variables, 4) sensitivity, and 5) privacy concern. Participants who declined to answer questions for the covariates were dropped from this analysis. We fit one regression model on the full set of responses for the second survey and one regression model on a subset of responses, excluding participants who had been willing to share their data even before reading the description of differential privacy.

A.2.3 Results

We did not find any evidence that the choice of explanation, the choice of privacy parameter, or the interaction between these two variables affected willingness to share. However, we found that the proportions of participants who were and were not willing to share—and who did and did not change their minds—remained very consistent. In summary, we found that our hypotheses were not supported by the data.

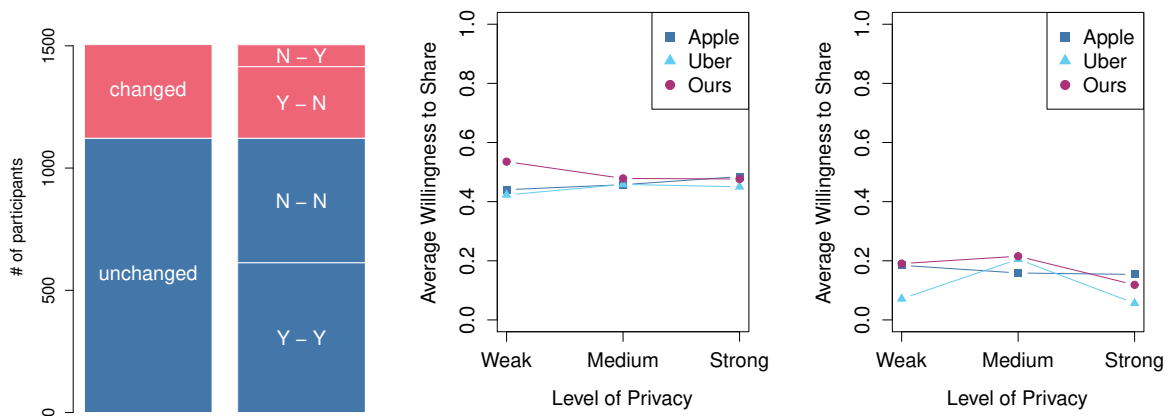


Figure A.1. Replication study results. The proportion of participants who were willing to share their data closely matched the results of the second experiment (left). There were no significant differences in willingness to share for the different explanations and privacy levels (middle), even when excluding participants who were willing to share without any privacy protection (right).

A.3 Codes

Below we list each code and provide an example.

Concerns About Scope of Data Collection

Example: *I really don't think there is anything that's truly privacy protected. If a company wants your data, information companies can get it anyways just how technology is these days.*

Trust of researchers / platform

Example: *Confident that you will not misuse the data*

Trust of mechanisms

Example: *I decided to share because you shared the rules that you use to keep my data private and because the rules that you use are simple enough and clear, I did not feel threatened*

Distrust of researchers / platform

Example: *I do not trust my information being shared with qmee*

Distrust of mechanisms

Example: *The process doesn't seem safe. I'm not sure what information would be changed. This seems like a hoax*

Helping researchers

Example: *You are doing a research study and want information. I have nothing to hide in sharing the websites I have visited. Though the process will not be accurate, it gives you an idea of the data you are searching for.*

Sensitivity

Example: *Because is very private and I don't like to share.*

Scared of getting information stolen

Example: *I dont want to be hacked*

Shared device / company device

Example: *Because I have no idea what is on here. This isn't just my phone. It is shared so I can't consent.*

Completion of survey

Example: *So can get the credit for survey if possible*

Confusion

Example: *I really have no idea what is going to happen. All the stuff about changing the information was just confusing.*

Nothing to hide

Example: *Have nothing to hide nor be ashamed of.*

Discomfort

Example: *I don't feel comfortable doing it right now*

Curiosity

Example: *I decided to share to know how the data sharing works*

Indifference

Example: *Sure why not*

Appendix B

B.1 Codes

Based on the interview data, the research team collectively developed this set of twenty low-level codes, grouped into higher-order themes:

Additional Information Requested

- How questions

Example: *Just like how the barrier works, a little more detail.*

- What questions

Example: *I would like to know exactly what information from medical records would be shown.*

- Who questions

Example: *I don't know what the who the nonprofit is partnering with or why.*

Design Feedback

- Alternative presentations

- Links to more detailed information

Example: *I'd have a link or something to explain what general patterns means, what's the full detail, maybe as a side if they really were interested in knowing.*

- Terms of use / consent documents

Example: *I think I would definitely start with the thing that comes to mind first are informed consents that we sign as participants, and they're very clear about how will your data will be stored and who has access, how will it be de-identified.*

- Video or animation

Example: *What you could do is some sort of like animation type thing with a video-like format.*

- Icons

- Color

Example: *The green and red doesn't work for me.*

- Locks

Example: *I like the look of the lock.*

- Privacy barrier

Example: *A label of some sort beyond privacy barrier might be helpful.*

- Things people liked

Example: *I like things that make it faster to read.*

Participant Understanding

- Did not understand

Example: *So I'm not clear as to what the protection actually does.*

- Misconception

Example: *It allows the people who utilize the information, say the law enforcement and medical professionals, it would allow them to share that information amongst themselves in a secure network without allowing the people who want to get that information to abuse that information.*

- DP as anonymization

Example: *The only way I could explain it would be that an individual's personally identifying details would not be included with their medical records.*

- DP as fake data

Example: *It's basically saying that we might put fake data in some parts of it.*

- Other PETs

Example: *So basically there's like some kind of firewall that keeps my privacy safe.*

- User-generated metaphors

Example: *It's kind of like an egg. You know, you crack it open and you don't know if it's going to be rotten inside or not. But I don't know what chicken it came from, so I can't blame the chicken.*

Reasoning About Data Sharing

- Benefits

Example: *I actually think that people like data analysts or employee university employees probably want to see my information. Like in that case, that's when it's okay for privacy to be breached. Because it's for the purpose of the study.*

- Concerns

- Concerns or skepticism about adequacy of protection

Example: *It sounds good, but I just read too many things about the Internet not being so secure as we would like.*

- Data disclosure risks (or lack thereof)

Example: *Especially like insurance companies, I would want to make sure that it's not being shared without my knowledge.*

- Lack of concern about privacy in general

Example: *I don't care about my personal information being released.*

Sets of codes were also developed for the open-text survey responses. The following codes are related to respondents' reasoning about data sharing.

- Relationship with doctor

Example: *I believe that if the doctors office is working with the non profit, I believe I trust them, there would also be massive repercussions if they were to do anything wrong with the records.*

- Want more info

Example: *before i say yes, i would need more info such as-will they see my name, do they want my entire medical history, what kind of boundaries in medicine are they pushing and do they align with my beliefs and morals*

- Too risky or too private

Example: *I think with all that's been going with abortion in the USA I'd be extremely wary of sharing medical data with a third party. Even if they have an extra layer of privacy protection they could still get hacked or the government could decide it has a right to that data.*

- Nothing to hide

Example: *I would share my medical records with anyone who wanted to see them. This would not be an issue for me. I have nothing to hide.*

- Benefits of data sharing

Example: *Yes, yes and absolutely yes. If this will help just ONE person who needs it, I would gladly share what I can to help them as long as my privacy was protected. Heck, even if it wasn't protected if it could still help then yes. I'm seeing commercials talking about wanting cancer Institutions to start doing this. This could have helped my dad perhaps. And if anyone would need to see his records to help others, I'd say yes.*

- Trust

Example: *I want to help them with their research and I trust that they will be able to keep my information private.*

- Distrust

Example: *i dont trust them*

- Money

Example: *In todays society where information is money, I have a hard time trusting organizations or institutions with very private information such as medical records.*

- Frequency of data breaches

Example: *Reassurances about security technology are hollow. Everything is breached eventually. It's just an arms race with the hackers.*

- Laws and Regulations

Example: *Medical record information should be protected and private. That is what HIPPA is for.*

- Deletion

Example: *Too loose in management, no note of when data will be deleted (which is the basic requirement for data collection in modern times)*

- Data already out there

Example: *Because most medical information is public*

This final set of codes is related to feedback obtained through the online survey.

- Simplify

Example: *Make it more simplified and shorter*

- More info about how protection works

Example: *There needs to be more explanation about how the privacy protection works.*

- Other info requests

Example: *need to be informed on where my data is going.*

- Positives

Example: *I found the explanations of the privacy protection to be clear, concise, and easy to understand.*

- Confusion

Example: *The picture is confusing to me. I don't understand why it needs two different sections*

- Nothing is foolproof

Example: *Anything can be hacked. No one can be trusted*

- Distrust

Example: *For me it's more of a feeling that I don't trust what is being presented as far as the safety of my information.*

B.2 Survey Instrument

B.2.1 Instructions

In this survey we are going to ask you a series of questions about a hypothetical scenario. Please do your best to imagine yourself in this scenario and answer the questions as if you were actually making the decisions about which you will be asked.

B.2.2 Scenario Description

Imagine that during your next doctor's visit, your primary care doctor informs you that they are part of a non-profit organization trying to push the boundaries of medical research. The non-profit is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. Your doctor, with your permission, can facilitate the non-profit getting the information they need.

B.2.3 Privacy Description

The non-profit organization will use an extra layer of privacy technology to protect your information. [Explanation inserted here.]

B.2.4 Comprehension Check

What kind of information does the non-profit want to collect? [Choice order randomized.]

- Medical records
- Music videos
- Book titles
- Location histories

B.2.5 Trust

Please indicate your agreement with the following statement: I trust the non-profit organization to protect my personal information privacy.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree
- Prefer not to answer

B.2.6 Self-Efficacy

How confident are you that you have enough information to decide whether to share your medical record with the non-profit?

- Very confident
- Confident
- Moderately confident
- Slightly confident
- Not at all confident
- Prefer not to answer

How confident are you about deciding whether to share your medical record with the non-profit?

- Very confident
- Confident
- Moderately confident
- Slightly confident
- Not at all confident
- Prefer not to answer

B.2.7 Share

Would you be willing to share your medical record with the non-profit?

- Yes
- No

- Prefer not to answer

Please explain your decision. [Text entry.]

B.2.8 Objective Comprehension

For each of the following statements, please indicate if you expect the following to be true or false if you share your medical record with the non-profit.

An employee working for the non-profit, such as a data analyst, could be able to see my exact medical history.

- True
- False
- I don't know
- Prefer not to answer

A criminal or foreign government that hacks the non-profit could learn my medical history.

- True
- False
- I don't know
- Prefer not to answer

A law enforcement organization could access my medical history with a court order requesting this data from the non-profit.

- True
- False
- I don't know

- Prefer not to answer

Graphs or informational charts created using information given to the non-profit could reveal my medical history.

- True
- False
- I don't know
- Prefer not to answer

Data that the non-profit shares with other organizations doing medical research could reveal my medical history.

- True
- False
- I don't know
- Prefer not to answer

B.2.9 Thoroughness

Please indicate your agreement with the following statement: I feel that it was explained thoroughly to me how the non-profit protects patient privacy.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree
- Prefer not to answer

B.2.10 Subjective Understanding

How confident are you in your understanding of the privacy protection?

- Very confident
- Confident
- Moderately confident
- Slightly confident
- Not at all confident
- Prefer not to answer

B.2.11 Feedback

What feedback (if any) would you like to share about the explanations of privacy protection?

[Text entry.]

B.2.12 PETs

Have you ever heard of the following technologies? (select all that apply) [Choice order randomized.]

- Differential privacy
- End-to-end encryption
- Secure multi-party computation
- Deliquescent security
- None of the above
- Prefer not to answer

Which of these technologies do you think was described in the survey? [Choice order randomized.]

- Differential privacy
- End-to-end encryption
- Secure multi-party computation
- Deliquescent security
- None of the above
- Prefer not to answer

Please explain your reasoning. [Text entry.]

B.2.13 Background

How familiar are you with the following computer and Internet-related items? Please choose a number between 1 and 5, where 1 represents no understanding and 5 represents full understanding of the item.

- Advanced Search
1 (No Understanding) - 5 (Full understanding)
or Prefer not to answer
- PDF
1 (No Understanding) - 5 (Full understanding)
or Prefer not to answer
- Spyware
1 (No Understanding) - 5 (Full understanding)
or Prefer not to answer

- Wiki
1 (No Understanding) - 5 (Full understanding)
or Prefer not to answer
- Cache
1 (No Understanding) - 5 (Full understanding)
or Prefer not to answer
- Phishing
1 (No Understanding) - 5 (Full understanding)
or Prefer not to answer

In what year were you born? (four digits please) [Text entry.]

What is your gender? [Multiselect.]

- Man
- Woman
- Non-binary
- Prefer to self describe: [Text entry.]
- Prefer not to answer

Please specify your race/ethnicity (select all that apply).

- Hispanic, Latino, or Spanish
- Black or African American
- White
- American Indian or Alaska Native

- Asian, Native Hawaiian, or Pacific Islander
- Prefer to self describe: [Text entry.]
- Prefer not to answer

What is the highest level of school you have completed or the highest degree you have received?

- Less than high school degree
- High school graduate (high school diploma or equivalent including GED)
- Some college but no degree
- Associate's degree
- Bachelor's degree
- Advanced degree (e.g., Master's, doctorate)
- Prefer not to answer

Which of the following best describes your educational background or job field?

- I have an education in, or work in, the field of computer science, computer engineering or IT.
- I DO NOT have an education in, nor do I work in, the field of computer science, computer engineering or IT.
- Prefer not to answer

Which one of the following includes your total HOUSEHOLD income for last year, before taxes?

- Less than \$10,000
- \$10,000 to under \$20,000

- \$20,000 to under \$30,000
- \$30,000 to under \$40,000
- \$40,000 to under \$50,000
- \$50,000 to under \$65,000
- \$65,000 to under \$80,000
- \$80,000 to under \$100,000
- \$100,000 to under \$125,000
- \$125,000 to under \$150,000
- \$150,000 to under \$200,000
- \$200,000 or more
- Prefer not to answer

B.3 Demographics

Table B.1 describes the demographic makeup of the 24 participants in the main interview study. Table B.2 describes the demographic makeup of the 10 participants who participated in the follow-up interviews. Table B.3 summarizes the demographic makeup of the survey respondents. Note that respondents could select multiple values for race/ethnicity and gender and that many respondents selected multiple options for race/ethnicity but did not explicitly describe themselves as multiracial.

Table B.1. Participant Demographics: Initial Interviews

Demographic Attribute		Count
<i>Gender</i>	Female	10
	Male	14
<i>Age</i>	< 20	2
	20-29	9
	30-39	6
	40-49	4
	50+	3
<i>Race</i>	Asian	1
	Black or African American	4
	Mixed, Multiracial, or Biracial	3
	White or Caucasian	16
<i>Education</i>	Secondary education (e.g. GED / GCSE)	1
	High school diploma / A-levels	11
	Technical / community college	4
	Undergraduate degree (BA / BSc / other)	5
	Graduate degree	2
	Doctorate degree (PhD / other)	1

Table B.2. Participant Demographics: Follow-up Interviews

Demographic Attribute		Count
<i>Gender</i>	Female	5
	Male	5
<i>Age</i>	< 20	1
	20-29	3
	30-39	1
	40-49	1
	50+	4
<i>Race</i>	Asian	3
	Black or African American	1
	Mixed, Multiracial, or Biracial	2
	White or Caucasian	3
	Native American	1
<i>Education</i>	High school diploma / A-levels	2
	Technical / community college	2
	Undergraduate degree (BA / BSc / other)	5
	Doctorate degree (PhD / other)	1

Table B.3. Respondent Demographics

Demographic Attribute		Count
Gender	Woman	343
	Man	335
	Non-binary	15
	Agender / Gender-fluid afab / genderqueer / they	5
Age	< 20	12
	20-29	249
	30-39	219
	40-49	98
	50+	119
Race/Ethnicity	Hispanic, Latino, or Spanish	83
	Black or African American	68
	White	478
	American Indian or Alaska Native	12
	Asian, Native Hawaiian, or Pacific Islander	110
	Multiracial or Mixed race	4
Education	High school or less	124
	Some college	233
	Bachelor's or above	337
Income	Less than \$10,000	41
	\$10,000 to under \$20,000	53
	\$20,000 to under \$30,000	79
	\$30,000 to under \$40,000	68
	\$40,000 to under \$50,000	65
	\$50,000 to under \$65,000	85
	\$65,000 to under \$80,000	88
	\$80,000 to under \$100,000	51
	\$100,000 to under \$125,000	57
	\$125,000 to under \$150,000	30
	\$150,000 to under \$200,000	25
	\$200,000 or more	32
Tech	Education or work in CSE/IT	148
	No education nor work in CSE/IT	527

B.4 Designs

Figure B.1 shows an example of the kind of Miro board with which participants in our follow-up interviews would have interacted. Table B.4 lists the original metaphor texts. Figure B.2 shows representative examples of our diagrams, and figure B.3 shows our original privacy labels.

Figure B.4 shows how our designs evolved over time. In the initial interviews, we evaluated multiple versions of each explanation type for both the local and central models. Based on participant feedback, we dropped the diagram explanations and modified the privacy labels. During the follow-up interviews, we developed a new metaphor and introduced a text with

information about the data protection process. Next, we compared the two privacy labels through a survey, and dropped the version with locks. Finally, we evaluated the disguise metaphor, the process text, and combinations of these texts and the privacy label with arrows.

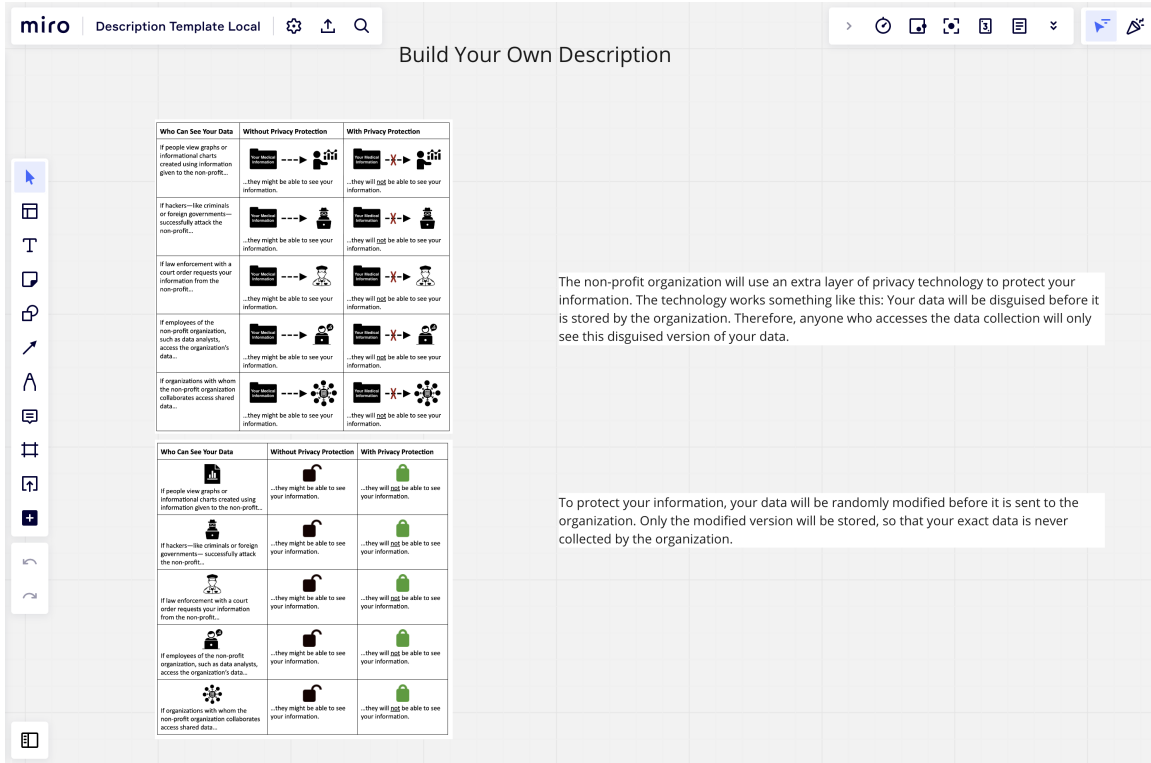


Figure B.1. Example of the Miro board setup used for the follow-up interviews.

Table B.4. Original Metaphor Descriptions

Local	Central
<i>Sharing data with the protection of this technology is like donating a penny to a crowdfunding campaign. No one will know with certainty that you donated. The sum of the donations from a large group of people will be valuable to our data analysts.</i>	<i>Publishing statistics, graphs, or tables using this technology is like publishing a blurry photo of the database that allows the viewer to see general patterns while hiding individual details. However, someone who obtained access to the database would be able to see all of the collected information in full detail.</i>
<i>The technology works something like this: Imagine that we are collecting photographs, but instead of collecting the raw images, we blur the images, and only collect the blurry images, so that little is revealed about you as an individual. Anyone who accesses our collected data will only see the blurry images, rather than the originals.</i>	<i>Publishing statistics, graphs, or tables using this technology is like publishing a photo of a mosaic, taken from a distance. People viewing this photo would not be able to see the individual tiles—in other words, individuals' data—yet they would still be able to see the overall picture. However, someone with direct access to the mosaic would be able to discern the individual tiles.</i>

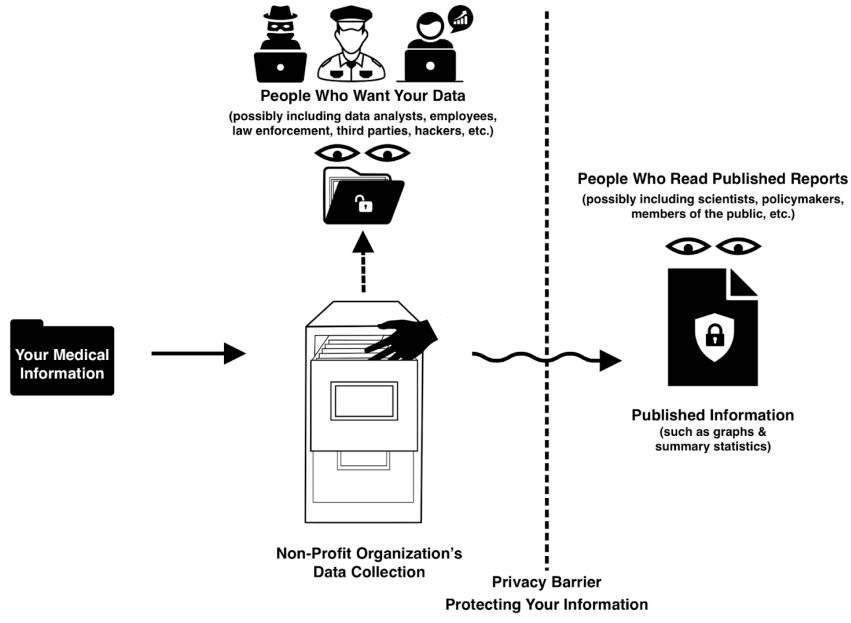
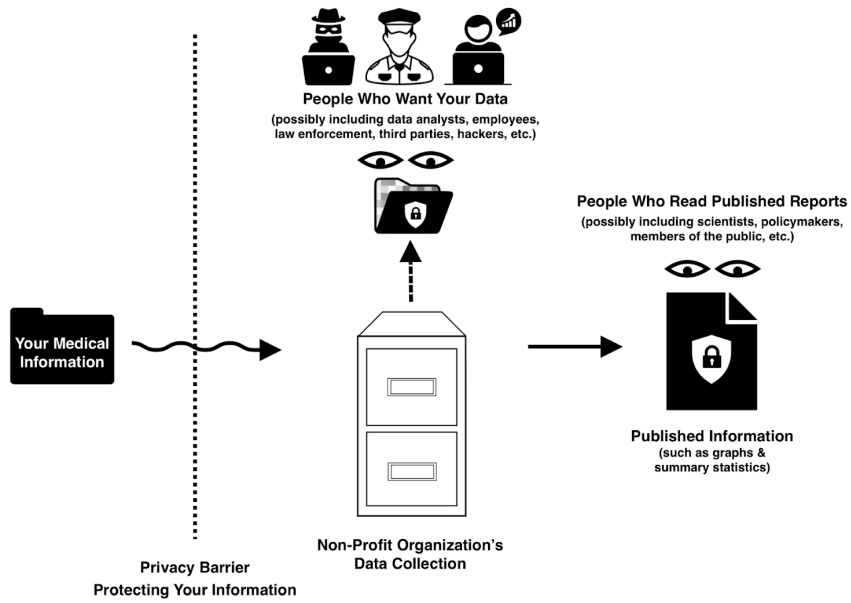






































Figure B.2. Top: Diagram for local model. Bottom: Diagram for central model.

Privacy protection	
	A person looking at graphs or informational charts created using information given to the non-profit will not be able to see your information.
	A criminal or foreign government that hacks the non-profit will not be able to see your information.
	A law enforcement organization with a court order requesting this data from the non-profit will not be able to see your information.
	Employees, such as data analysts, working for the non-profit organization will not be able to see your information.
	Other organizations doing medical research with whom the non-profit organization shares data will not be able to see your information.

Privacy protection	
	A person looking at graphs or informational charts created using information given to the non-profit will not be able to see your information.

Who Can See Your Data	Without Privacy Protection	With Privacy Protection
 If people view graphs or informational charts created using information given to the non-profit...	 ...they might be able to see your information.	 ...they will <u>not</u> be able to see your information.
 If hackers—like criminals or foreign governments—successfully attack the non-profit...	 ...they might be able to see your information.	 ...they will <u>not</u> be able to see your information.
 If law enforcement with a court order requests your information from the non-profit...	 ...they might be able to see your information.	 ...they will <u>not</u> be able to see your information.
 If employees of the non-profit organization, such as data analysts, access the organization's data...	 ...they might be able to see your information.	 ...they will <u>not</u> be able to see your information.
 If organizations with whom the non-profit organization collaborates access shared data...	 ...they might be able to see your information.	 ...they will <u>not</u> be able to see your information.

Who Can See Your Data	Without Privacy Protection	With Privacy Protection
 If people view graphs or informational charts created using information given to the non-profit...	 ...they might be able to see your information.	 ...they will <u>not</u> be able to see your information.
 If hackers—like criminals or foreign governments—successfully attack the non-profit...	 ...they might be able to see your information.	 ...they might be able to see your information.
 If law enforcement with a court order requests your information from the non-profit...	 ...they might be able to see your information.	 ...they might be able to see your information.
 If employees of the non-profit organization, such as data analysts, access the organization's data...	 ...they might be able to see your information.	 ...they might be able to see your information.
 If organizations with whom the non-profit organization collaborates access shared data...	 ...they might be able to see your information.	 ...they might be able to see your information.

(a) Local

(b) Central

Figure B.3. Original Privacy Labels.

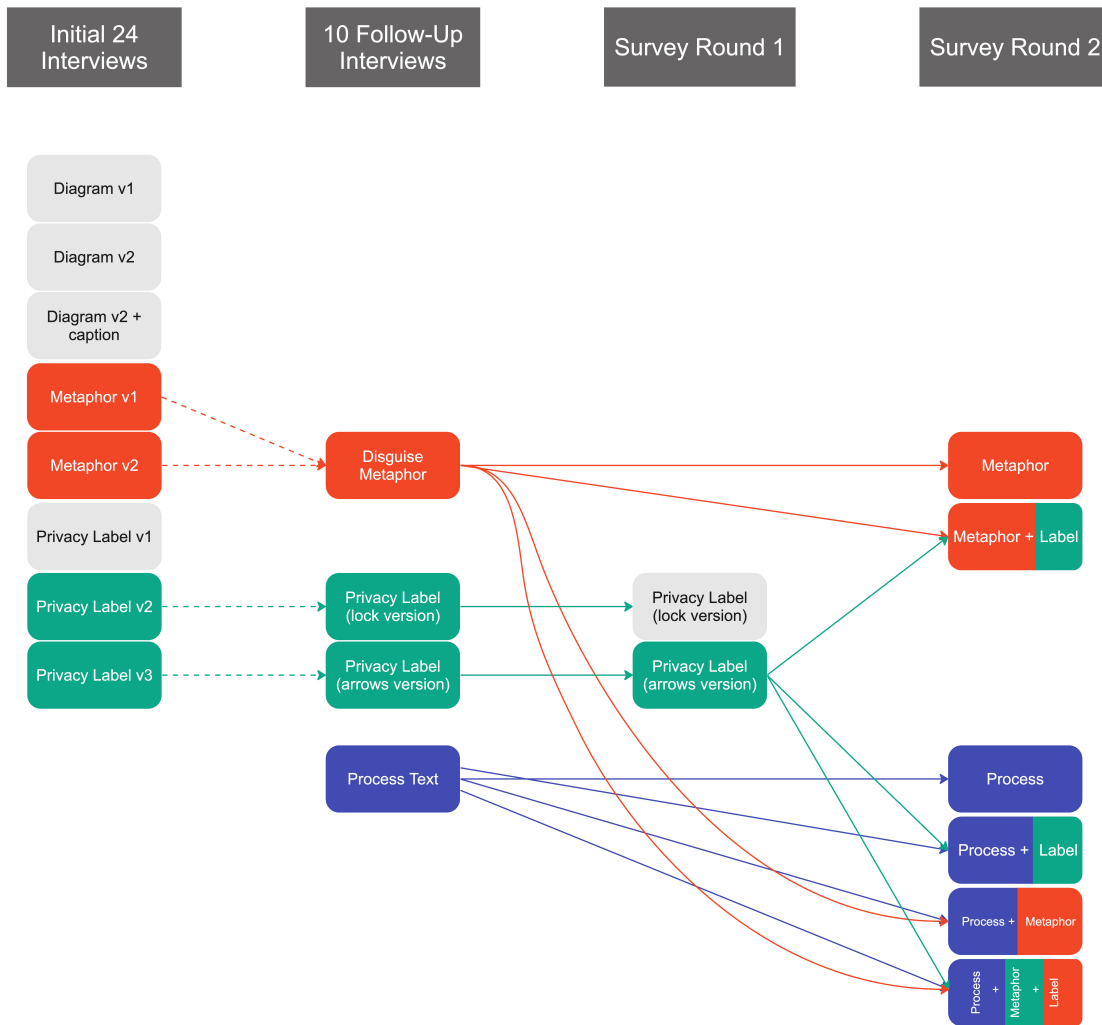


Figure B.4. Evolution of designs over time.

B.5 Descriptive Statistics

Table B.5 displays the proportion of respondents per condition who answered each comprehension correctly (+) and incorrectly (-). Since some respondents selected ‘I don’t know,’ these percentages may not add to 1.

Table B.5. Accuracy of Privacy Expectations

Model	Explanation	Hack		Law		Org		Graph		Share	
		+	-	+	-	+	-	+	-	+	-
Central	Metaphor	0.78	0.05	0.54	0.14	0.89	0.03	0.41	0.46	0.54	0.22
Local	Metaphor	0.26	0.47	0.24	0.45	0.37	0.37	0.39	0.42	0.26	0.47
Central	Process	0.50	0.28	0.50	0.17	0.52	0.17	0.42	0.32	0.45	0.32
Local	Process	0.28	0.48	0.3	0.50	0.38	0.42	0.32	0.28	0.28	0.52
Central	Process+Metaphor	0.79	0.11	0.76	0.05	0.92	0.08	0.61	0.34	0.50	0.37
Local	Process+Metaphor	0.36	0.38	0.33	0.33	0.44	0.41	0.49	0.31	0.38	0.41
Central	ArrowLabel	0.92	0.03	0.95	0.00	0.92	0.05	0.58	0.26	0.82	0.00
Local	ArrowLabel	0.51	0.32	0.44	0.37	0.46	0.32	0.63	0.24	0.63	0.27
Central	Label+Metaphor	0.85	0.13	0.82	0.05	0.82	0.08	0.64	0.26	0.77	0.15
Local	Label+Metaphor	0.72	0.22	0.58	0.33	0.58	0.33	0.69	0.19	0.61	0.31
Central	Label+Process	0.85	0.12	0.82	0.05	0.65	0.15	0.6	0.25	0.62	0.22
Local	Label+Process	0.59	0.28	0.44	0.31	0.67	0.21	0.62	0.18	0.67	0.28
Central	Label+Process+Metaphor	0.85	0.05	0.82	0.03	0.80	0.15	0.45	0.40	0.70	0.15
Local	Label+Process+Metaphor	0.56	0.26	0.59	0.18	0.56	0.26	0.69	0.15	0.49	0.26
Central	Xiong	0.91	0.03	0.44	0.12	0.62	0.15	0.35	0.44	0.44	0.32
Local	Xiong	0.33	0.31	0.23	0.44	0.33	0.41	0.56	0.23	0.41	0.46

Appendix C

The table below provides a detailed breakdown of participant demographics. Note that participants had the option to decline to state any of their demographic information.

Table C.1. Participant Demographics

Demographic Category		Count
Age	18-29	17
	30-39	2
	40-49	3
	50-59	2
Gender	Male	14
	Female	9
	Non-binary/third gender	1
Race/Ethnicity	Asian/Native Hawaiian/Pacific Islander	3
	Black/African American	11
	Hispanic/Latino/Spanish	6
	White	4
Education	High school degree or equivalent	1
	Some college, no degree	10
	2 year degree	1
	4 year degree or more	12
Income	Less than \$25,000	3
	\$25,000 - \$49,999	6
	\$50,000 - \$74,999	6
	\$75,000 - \$99,999	5
	\$100,000 or more	3

Bibliography

- [1] YouTube Data API. <https://developers.google.com/youtube/v3>.
- [2] youtube-dl, 2021. <https://github.com/yt-dl-org/youtube-dl>.
- [3] A. Abdul, C. von der Weth, M. Kankanhalli, and B. Y. Lim. COGAM: measuring and moderating cognitive load in machine learning model explanations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.
- [4] J. M. Abowd. The U.S. Census Bureau Adopts Differential Privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18*, page 2867. Association for Computing Machinery, 2018. ISBN 9781450355520. URL <https://doi.org/10.1145/3219819.3226070>.
- [5] J. M. Abowd and I. M. Schmutte. An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, 109(1):171–202, 2019.
- [6] R. Abu-Salma, E. M. Redmiles, B. Ur, and M. Wei. Exploring user mental models of {End-to-End} encrypted communication tools. In *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*, 2018.
- [7] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58. Springer, 2006.
- [8] A. Acquisti, L. Brandimarte, and G. Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.
- [9] O. Akgul, R. Roberts, M. Namara, D. Levin, and M. L. Mazurek. Investigating influencer vpn ads on youtube. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 876–892. IEEE, 2022.
- [10] A. S. Alaqra, F. Karegar, and S. Fischer-Hübner. Structural and functional explanations for informing lay and expert users: the case of functional encryption. *Proceedings on Privacy Enhancing Technologies*, 4:359–380, 2023.

- [11] A. S. Alaqra, F. Karegar, and S. Fischer-Hübner. Communicating the privacy functionality of PETs to eHealth stakeholders. 2023.
- [12] M. R. Albrecht, J. Blasco, R. B. Jensen, and L. Mareková. Collective information security in large-scale urban protests: the case of hong kong. In *USENIX Security Symposium*, pages 3363–3380, 2021.
- [13] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke. Discrimination through optimization: How Facebook’s ad delivery can lead to biased outcomes. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), 2019. URL <https://doi.org/10.1145/3359301>.
- [14] R. E. Allen and J. L. Wiles. A rose by any other name: Participants choosing research pseudonyms. *Qualitative Research in Psychology*, 13(2):149–165, 2016.
- [15] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15*, pages 787–796, Seoul, Republic of Korea, Apr. 2015. Association for Computing Machinery. ISBN 978-1-4503-3145-6. URL <https://doi.org/10.1145/2702123.2702210>.
- [16] Amazon Web Services. Nitro Enclaves. <https://aws.amazon.com/ec2/nitro/nitro-enclaves/>, 2019. Accessed 9/14/2023.
- [17] M. Ananny and K. Crawford. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society*, 20(3): 973–989, 2018.
- [18] E. B. Andrade, V. Kaltcheva, and B. Weitz. Self-disclosure on the web: The impact of privacy policy, reward, and company reputation. *ACR North American Advances*, 2002.
- [19] A. Andreou, G. Venkatadri, O. Goga, K. P. Gummadi, P. Loiseau, and A. Mislove. Investigating ad transparency mechanisms in social media: A case study of Facebook’s explanations. In *NDSS 2018-Network and Distributed System Security Symposium*, pages 1–15, 2018.
- [20] P. Arora. Decolonizing privacy studies. *Television & New Media*, 20(4):366–378, 2019.
- [21] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of security risks. In S. Dietrich and R. Dhamija, editors, *Financial Cryptography and Data Security*, pages 367–377, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-77366-5.
- [22] B. Auxier, L. Rainie, M. Anderson, A. Perrin, M. Kumar, and E. Turner. Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal

Information. Pew Research Center, 2019. URL <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>.

- [23] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, pages 61–70, 2010.
- [24] W. Bai, M. Namara, Y. Qian, P. G. Kelley, M. L. Mazurek, and D. Kim. An Inconvenient Trust: User Attitudes toward Security and Usability Tradeoffs for Key-Directory Encryption Systems. pages 113–130, 2016. ISBN 978-1-931971-31-7. URL <https://www.usenix.org/conference/soups2016/technical-sessions/presentation/bai>.
- [25] S. Barocas and H. Nissenbaum. Big data’s end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement*, 1:44–75, 2014.
- [26] S. Barth and M. D. De Jong. The privacy paradox—investigating discrepancies between expressed privacy concerns and actual online behavior—a systematic literature review. *Telematics and informatics*, 34(7):1038–1058, 2017. URL <https://doi.org/10.1016/j.tele.2017.04.013>.
- [27] S. Benthall and R. Cummings. Integrating differential privacy and contextual integrity. Santa Clara, CA, June 2022. USENIX Association.
- [28] J. Bernd, R. Abu-Salma, J. Choy, and A. Frik. Balancing power dynamics in smart homes: nannies’ perspectives on how cameras reflect and affect relationships. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 687–706, 2022.
- [29] H. Beyer and K. Holtzblatt. Contextual design. *interactions*, 6(1):32–42, 1999.
- [30] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta. Discovering frequent patterns in sensitive data. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 503–512, 2010.
- [31] B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel. Inferring the demographics of search users: Social data meets search queries. In *Proceedings of the 22nd international conference on World Wide Web*, pages 131–140, 2013.
- [32] B. Bichsel, S. Steffen, I. Bogunovic, and M. Vechev. Dp-sniper: Black-box discovery of differential privacy violations using classifiers. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 391–409, 2021. doi: 10.1109/SP40001.2021.00081.
- [33] G. Biczók and P. H. Chia. Interdependent privacy: Let me share your data. In *International conference on financial cryptography and data security*, pages 338–353. Springer, 2013.

- [34] C. Bösch, B. Erb, F. Kargl, H. Kopp, and S. Pfattheicher. Tales from the dark side: privacy dark strategies and privacy dark patterns. *Proc. Priv. Enhancing Technol.*, 2016 (4):237–254, 2016.
- [35] R. E. Boyatzis. *Transforming qualitative information: Thematic analysis and code development*. 1998.
- [36] d. boyd and E. Hargittai. Facebook privacy settings: Who cares? *First Monday*, 2010.
- [37] D. Boyd and J. Heer. Profiles as conversation: Networked identity performance on friendster. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, volume 3, pages 59c–59c, 2006. doi: 10.1109/HICSS.2006.394.
- [38] boyd, danah and Sarathy, Jayshree. Differential perspectives: Epistemic disconnects surrounding the us census bureau’s use of differential privacy. *Harvard Data Science Review*, 2022. URL <https://doi.org/10.1162/99608f92.66882f0e>.
- [39] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [40] C. Bravo-Lillo, L. F. Cranor, J. S. Downs, and S. Komanduri. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *IEEE Security & Privacy*, 9 (2):18–26, Mar. 2011.
- [41] B. Brown. Studying the internet experience. *HP laboratories technical report HPL*, 49, 2001.
- [42] Brown v. Google LLC. Case No. 4:20-cv-03664-YGR-SVK, 2024. URL <https://www.courtlistener.com/docket/17216783/1096/brown-v-google-llc/>.
- [43] A. Bruckman, K. Luther, and C. Fiesler. When should we use real names in published accounts of internet research. *Digital research confidential: The secrets of studying behavior online*, page 243, 2015.
- [44] F. Brunton and H. Nissenbaum. *Obfuscation: A user’s guide for privacy and protest*. MIT Press, 2015.
- [45] B. Bullek, S. Garboski, D. J. Mir, and E. M. Peck. Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, pages 3833–3837, Denver, Colorado, USA, May 2017. Association for Computing Machinery. ISBN 978-1-4503-4655-9. URL <https://doi.org/10.1145/3025453.3025698>.
- [46] C. Cadwalladr and E. Graham-Harrison. Revealed: 50 million facebook profiles harvested

for cambridge analytica in major data breach. *The Guardian*, March 2018. URL <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

- [47] L. J. Camp. Mental models of privacy and security. *IEEE Technology and society magazine*, 28(3):37–46, 2009.
- [48] S. Casacuberta, M. Shoemate, S. P. Vadhan, and C. Wagaman. Widespread underestimation of sensitivity in differentially private libraries and how to fix it. In H. Yin, A. Stavrou, C. Cremers, and E. Shi, editors, *ACM CCS 2022*, pages 471–484. ACM Press, Nov. 2022. doi: 10.1145/3548606.3560708.
- [49] G. Cheney, L. T. Christensen, C. Conrad, and D. J. Lair. Corporate rhetoric as organizational discourse. *The Sage handbook of organizational discourse*, pages 79–103, 2004.
- [50] H. Cho, B. Knijnenburg, A. Kobsa, and Y. Li. Collective privacy management in social media: A cross-cultural validation. *ACM Trans. Comput.-Hum. Interact.*, 25(3), jun 2018. ISSN 1073-0516. URL <https://doi.org/10.1145/3193120>.
- [51] Chromium Blog. An update on the lock icon, May 2023.
- [52] J. Colnago, Y. Feng, T. Palanivel, S. Pearman, M. Ung, A. Acquisti, L. F. Cranor, and N. Sadeh. Informing the design of a personalized privacy assistant for the internet of things. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.
- [53] K. Corcoran. Facebook is overhauling its privacy settings in response to the cambridge analytica scandal. *Business Insider*, 2018.
- [54] K. J. W. Craik. *The nature of explanation*, volume 445. CUP Archive, 1967.
- [55] L. F. Cranor. Mobile-app privacy nutrition labels missing key ingredients for success. *Communications of the ACM*, 65(11):26–28, 2022.
- [56] L. F. Cranor. How Everyone Can Get the Online Privacy They Want. *The Wall Street Journal*, June 2022. URL <https://www.wsj.com/articles/online-privacy-consent-11654540664>.
- [57] R. Cummings, G. Kaptchuk, and E. M. Redmiles. “I need a better description”: An investigation into user expectations for differential privacy. In G. Vigna and E. Shi, editors, *ACM CCS 2021*, pages 3037–3052. ACM Press, Nov. 2021. doi: 10.1145/3460120.3485252.

- [58] F. K. Dankar and K. El Emam. Practicing differential privacy in health care: A review. *Trans. Data Priv.*, 6(1):35–67, 2013.
- [59] S. Das, W. K. Edwards, D. Kennedy-Mayo, P. Swire, and Y. Wu. Privacy for the people? exploring collective action as a mechanism to shift power to consumers in end-user privacy. *IEEE Security & Privacy*, 19(5):66–70, 2021.
- [60] V. Das Swain, L. Gao, W. A. Wood, S. C. Matli, G. D. Abowd, and M. De Choudhury. Algorithmic power or punishment: Information worker perspectives on passive sensing enabled ai phenotyping of performance and wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2023.
- [61] C. A. Davis Jr, G. L. Pappa, D. R. R. De Oliveira, and F. de L. Arcanjo. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15(6): 735–751, 2011.
- [62] I. Dekel, R. Cummings, O. Heffetz, and K. Ligett. The privacy elasticity of behavior: Conceptualization and application. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC ’23, 2023.
- [63] A. Demjaha, J. M. Spring, I. Becker, S. Parkin, and M. A. Sasse. Metaphors considered harmful? an exploratory study of the effectiveness of functional metaphors for end-to-end encryption. In *Proc. USEC*, volume 2018. Internet Society, 2018.
- [64] L. Dencik. Surveillance realism and the politics of imagination: Is there no alternative. *Krisis: Journal for Contemporary Philosophy*, 1:31–43, 2018.
- [65] L. Dencik and J. Cable. The advent of surveillance realism: Public opinion and activist responses to the Snowden leaks. *International Journal of Communication*, 11:763–781, 2017. ISSN 1932-8036.
- [66] D. Desfontaines. A list of real-world uses of differential privacy. <https://desfontain.es/privacy/real-world-differential-privacy.html>, 10 2021. Ted is writing things (personal blog).
- [67] D. Desfontaines and B. Pejó. Sok: differential privacies. *Proceedings on privacy enhancing technologies*, 2020(2):288–313, 2020.
- [68] M. A. DeVito, A. M. Walker, and J. Birnholtz. ‘Too Gay for Facebook’ Presenting LGBTQ+ Identity Throughout the Personal Social Media Ecosystem. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–23, 2018.
- [69] Differential Privacy Team at Apple. Learning with privacy at scale. *Apple Machine Learning Journal*, 1(8), 2017.

- [70] C. D’Ignazio and L. Klein. On rational, scientific, objective viewpoints from mythical, imaginary, impossible standpoints. *Data Feminism*, 2020.
- [71] Z. Ding, Y. Wang, G. Wang, D. Zhang, and D. Kifer. Detecting violations of differential privacy. In D. Lie, M. Mannan, M. Backes, and X. Wang, editors, *ACM CCS 2018*, pages 475–489. ACM Press, Oct. 2018. doi: 10.1145/3243734.3243818.
- [72] V. Distler, C. Lallemand, and V. Koenig. Making encryption feel secure: Investigating how descriptions of encryption impact perceived security. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 220–229. IEEE, 2020.
- [73] C. Dolin, B. Weinshel, S. Shan, C. M. Hahn, E. Choi, M. L. Mazurek, and B. Ur. Unpacking perceptions of data-driven inferences underlying online targeting and personalization. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–12, 2018.
- [74] P. Dourish and K. Anderson. Collective information practice: Exploring privacy and security as social and cultural phenomena. *Human-computer interaction*, 21(3):319–342, 2006.
- [75] P. Dourish, R. E. Grinter, J. D. De La Flor, and M. Joseph. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, 2004.
- [76] N. A. Draper and J. Turow. The corporate cultivation of digital resignation. *New media & society*, 21(8):1824–1839, 2019.
- [77] B. E. Duffy and N. K. Chan. “you never really know who’s looking”: Imagined surveillance across social media platforms. *New Media & Society*, 21(1):119–138, 2019.
- [78] S. Duguay. “he has a way gayer Facebook than i do”: Investigating sexual identity disclosure and context collapse on a social networking site. *New media & society*, 18(6): 891–907, 2016.
- [79] K. Duncker and L. S. Lees. On problem-solving. *Psychological monographs*, 58(5):i, 1945.
- [80] C. Dwork. Differential Privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II, ICALP’06*, pages 1–12, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 978-3-540-35907-4. URL http://dx.doi.org/10.1007/11787006_1.
- [81] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer,

2006.

- [82] C. Dwork, N. Kohli, and D. Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), 2019. URL <https://journalprivacyconfidentiality.org/index.php/jpc/article/view/689>.
- [83] N. Ebert, K. A. Ackermann, and P. Heinrich. Does Context in Privacy Communication Really Matter? — A Survey on Consumer Concerns and Preferences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–11, Honolulu, HI, USA, Apr. 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. URL <https://doi.org/10.1145/3313831.3376575>.
- [84] P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal, and L. F. Cranor. An informative security and privacy “nutrition” label for internet of things devices. *IEEE Security & Privacy*, 20(02):31–39, 2022.
- [85] Y. Erlich, T. Shor, I. Pe’er, and S. Carmi. Identity inference of genomic data using long-range familial searches. *Science*, 362(6415):690–694, 2018.
- [86] Ú. Erlingsson, V. Pihur, and A. Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1054–1067, 2014.
- [87] M. Eslami, S. R. Krishna Kumaran, C. Sandvig, and K. Karahalios. Communicating algorithmic process in online behavioral advertising. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.
- [88] V. Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press, 2018.
- [89] B. Fabian, T. Ermakova, and T. Lentz. Large-scale readability analysis of privacy policies. In *Proceedings of the International Conference on Web Intelligence*, WI '17, page 18–25. Association for Computing Machinery, 2017. ISBN 9781450349512. URL <https://doi.org/10.1145/3106426.3106427>.
- [90] Facebook. Terms of service. 2022. URL <https://m.facebook.com/legal/terms>.
- [91] Federal Trade Commission. Facebook settles FTC charges that it deceived consumers by failing to keep privacy promises, 2011. URL <https://www.ftc.gov/news-events/press-releases/2011/11/facebook-settles-ftc-charges-it-deceived-consumers-failing-keep>.
- [92] Federal Trade Commission. Snapchat settles FTC charges that promises of disappearing messages were false, 2014.

- [93] Federal Trade Commission. FTC imposes \$5 billion penalty and sweeping new privacy restrictions on facebook, 2019.
- [94] A. P. Felt, R. W. Reeder, H. Almuhiemedi, and S. Consolvo. Experimenting at Scale with Google Chrome’s SSL Warning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’14, page 2667–2670. Association for Computing Machinery, 2014. ISBN 9781450324731. URL <https://doi.org/10.1145/2556288.2557292>.
- [95] A. P. Felt, R. W. Reeder, A. Ainslie, H. Harris, M. Walker, C. Thompson, M. E. Acer, E. Morant, and S. Consolvo. Rethinking connection security indicators. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 1–14, 2016.
- [96] C. Fennell and R. Wash. Do Stories Help People Adopt Two-factor Authentication? page 5, 2019.
- [97] P. M. Fernbach, T. Rogers, C. R. Fox, and S. A. Sloman. Political Extremism Is Supported by an Illusion of Understanding. *Psychological Science*, 24(6):939–946, 2013. ISSN 0956-7976, 1467-9280. URL <http://journals.sagepub.com/doi/10.1177/0956797612464058>.
- [98] A. Ferreira and G. Lenzini. An analysis of social engineering principles in effective phishing. In *2015 Workshop on Socio-Technical Aspects in Security and Trust*, pages 9–16, 2015. doi: 10.1109/STAST.2015.10.
- [99] D. S. Fidler and R. E. Kleinknecht. Randomized response versus direct questioning: Two data-collection methods for sensitive information. *Psychological Bulletin*, 84(5):1045, 1977.
- [100] C. Fiesler, N. Beard, and B. C. Keegan. No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 187–196, 2020.
- [101] D. Franzen, S. N. von Voigt, P. Sörries, F. Tschorsch, and C. Müller-Birn. Am I private and if so, how many?: Communicating privacy guarantees of differential privacy with risk communication formats. In H. Yin, A. Stavrou, C. Cremers, and E. Shi, editors, *ACM CCS 2022*, pages 1125–1139. ACM Press, Nov. 2022. doi: 10.1145/3548606.3560693.
- [102] D. Freelon. Computational research in the post-api age. *Political Communication*, 35(4): 665–668, 2018.
- [103] R. L. Freishtat and J. A. Sandlin. Shaping youth discourse about technology: Technological colonization, manifest destiny, and the frontier myth in facebook’s public pedagogy. *Educational Studies*, 46(5):503–523, 2010.

- [104] A. Frikk, L. Nurgalievaa, J. Bernd, J. Lee, F. Schaub, and S. Egelman. Privacy and security threat models and mitigation strategies of older adults. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 21–40, Santa Clara, CA, Aug. 2019. USENIX Association. ISBN 978-1-939133-05-2. URL <https://www.usenix.org/conference/soups2019/presentation/frik>.
- [105] A. Frikk, J. Bernd, and S. Egelman. A model of contextual factors affecting older adults’ information-sharing decisions in the us. *ACM Transactions on Computer-Human Interaction*, 30(1):1–48, 2023.
- [106] M. Gafni and L. M. Krieger. Here’s the ‘open-source’ genealogy dna website that helped crack the golden state killer case, 2018. URL <https://www.mercurynews.com/2018/04/26/ancestry-23andme-deny-assisting-law-enforcement-in-east-area-rapist-case/>.
- [107] G. S. Gaglione Jr. The equifax data breach: an opportunity to improve consumer protection and cybersecurity efforts in america. *Buff. L. Rev.*, 67:1133, 2019.
- [108] D. Garcia, M. Goel, A. K. Agrawal, and P. Kumaraguru. Collective aspects of privacy in the twitter social network. *EPJ Data Science*, 7:1–13, 2018.
- [109] P. Garcia, T. Sutherland, M. Cifor, A. S. Chan, L. Klein, C. D’Ignazio, and N. Salehi. No: Critical refusal as feminist data practice. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, pages 199–202, 2020.
- [110] C. S. Gates, J. Chen, N. Li, and R. W. Proctor. Effective Risk Communication for Android Apps. *IEEE Transactions on Dependable and Secure Computing*, 11(3):252–265, May 2014. ISSN 1941-0018. doi: 10.1109/TDSC.2013.58.
- [111] C. Geeng, M. Harris, E. Redmiles, and F. Roesner. ”like lesbians walking the perimeter”: Experiences of U.S. LGBTQ+ folks with online security, safety, and privacy advice. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 305–322, Boston, MA, Aug. 2022. USENIX Association. ISBN 978-1-939133-31-1. URL <https://www.usenix.org/conference/usenixsecurity22/presentation/geeng>.
- [112] P. Gerber, M. Volkamer, and K. Renaud. Usability versus privacy instead of usable privacy: Google’s balancing act between usability and privacy. *Acm Sigcas Computers and Society*, 45(1):16–21, 2015.
- [113] T. Germain. Apple Is Tracking You Even When Its Own Privacy Settings Say It’s Not, New Research Says. November 2022. URL <https://gizmodo.com/apple-iphone-analytics-tracking-even-when-off-app-store-1849757558>.
- [114] R. Ghaiumy Anaraky, Y. Li, and B. Knijnenburg. Difficulties of measuring culture in

- privacy studies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–26, 2021.
- [115] J. Gluck, F. Schaub, A. Friedman, H. Habib, N. Sadeh, L. F. Cranor, and Y. Agarwal. How short is too short? implications of length and framing on the effectiveness of privacy notices. In *Twelfth symposium on usable privacy and security (SOUPS 2016)*, pages 321–340. USENIX Association, 2016.
- [116] M. Golla, M. Wei, J. Hainline, L. Filipe, M. Dürmuth, E. Redmiles, and B. Ur. ” what was that site doing with my facebook password?” designing password-reuse notifications. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1549–1566, 2018.
- [117] M. S. Goodstadt and V. Gruson. The randomized response technique: A test on drug use. *Journal of the American Statistical Association*, 70(352):814–818, 1975.
- [118] J. Gottfried. Americans’ social media use. 2024.
- [119] D. Greene and K. Shilton. Platform privacies: Governance, collaboration, and the different meanings of “privacy” in ios and android development. *new media & society*, 20(4):1640–1657, 2018.
- [120] A. Griffin. WhatsApp launches new privacy campaign after attacks from users and governments. June 2021. URL <https://www.independent.co.uk/tech/whatsapp-privacy-encryption-campaign-ad-b1865461.html>.
- [121] M. Grinberg. *Flask web development: developing web applications with python*. O’Reilly Media, Inc., 2018.
- [122] E. Guo and T. Ryan-Mosley. Computer scientists designing the future can’t agree on what privacy means, 2023. URL <https://www.technologyreview.com/2023/04/03/1070665/cm-u-university-privacy-battle-smart-building-sensors-mites/>.
- [123] S. Gürses and B. Berendt. Pets in the surveillance society: a critical review of the potentials and limitations of the privacy as confidentiality paradigm. *Data Protection in a Profiled World*, pages 301–321, 2010.
- [124] S. Gürses, A. Kundnani, and J. Van Hoboken. Crypto and empire: The contradictions of counter-surveillance advocacy. *Media, Culture & Society*, 38(4):576–590, 2016.
- [125] H. Habib, J. Colnago, V. Gopalakrishnan, S. Pearman, J. Thomas, A. Acquisti, N. Christin, and L. F. Cranor. Away from prying eyes: Analyzing usage and understanding of private browsing. In *Fourteenth symposium on usable privacy and security (SOUPS 2018)*, pages 159–175, 2018.

- [126] H. Habib, S. Pearman, E. Young, I. Saxena, R. Zhang, and L. F. Cranor. Identifying user needs for advertising controls on facebook. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–42, 2022.
- [127] E. Hargittai and Y. P. Hsieh. Succinct survey measures of web-use skills. *Social Science Computer Review*, 30(1):95–107, 2012.
- [128] E. Hargittai and A. Marwick. “What can I really do?” Explaining the privacy paradox with online apathy. *International journal of communication*, 10:21, 2016.
- [129] E. Hargittai and M. Micheli. Internet skills and why they matter. *Society and the internet: How networks of information and communication are changing our lives*, 109, 2019.
- [130] W. Hartzog. What is privacy? that’s the wrong question. *U. Chi. L. Rev.*, 88:1677, 2021.
- [131] R. Hasan, D. Crandall, M. Fritz, and A. Kapadia. Automatically detecting bystanders in photos to reduce privacy risks. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 318–335. IEEE, 2020.
- [132] S. Hautea, S. Dasgupta, and B. M. Hill. Youth perspectives on critical data literacies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI ’17*, page 919–930. Association for Computing Machinery, 2017. ISBN 9781450346559. URL <https://doi.org/10.1145/3025453.3025823>.
- [133] S. Hautea, A. Munasinghe, and E. Rader. ‘That’s Not Me’: Surprising Algorithmic Inferences. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pages 1–7, 2020.
- [134] S. Hegelich. Facebook needs to share more with researchers. *Nature*, 579(7800):473–474, 2020.
- [135] A. Herzberg and H. Leibowitz. Can johnny finally encrypt? evaluating e2e-encryption in popular im applications. In *Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust*, pages 17–28, 2016.
- [136] K. Hill. How target figured out a teen girl was pregnant before her father did. *Forbes, Inc.*, 2012.
- [137] K. Hill. ‘Do Not Track,’ the Privacy Tool Used by Millions of People, Doesn’t Do Anything. *Gizmodo*, 2021. URL <https://gizmodo.com/do-not-track-the-privacy-tool-used-by-millions-of-peop-1828868324>.
- [138] A. L. Hoffmann and A. Jonas. Recasting justice for internet and online industry research ethics. *Internet Research Ethics for the Social Age: New Cases and Challenges. M.*

Zimmer and K. Kinder-Kuranda (Eds.), *np Bern, Switzerland: Peter Lang, Forthcoming*, 2016.

- [139] A. L. Hoffmann, N. Proferes, and M. Zimmer. “Making the world more open and connected”: Mark Zuckerberg and the discursive construction of Facebook and its users. *New media & society*, 20(1):199–218, 2018.
- [140] C. P. Hoffmann, C. Lutz, and G. Ranzini. Privacy cynicism: A new approach to the privacy paradox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(4), 2016.
- [141] S. Holland, A. Hosny, S. Newman, J. Joseph, and K. Chmielinski. The dataset nutrition label. *Data Protection and Privacy, Volume 12: Data Protection and Democracy*, 12:1, 2020.
- [142] D. C. Howe. Surveillance countermeasures: Expressive privacy via obfuscation. *Datafied Research*, 4(1):88–98, 2015.
- [143] D. C. Howe and H. Nissenbaum. Engineering privacy and protest: A case study of adnauseam. In *IWPE@ SP*, pages 57–64, 2017.
- [144] S. Hsu, K. Vaccaro, Y. Yue, A. Rickman, and K. Karahalios. Awareness, navigation, and use of feed control settings online. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.
- [145] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one*, 10(11): e0142444, 2015.
- [146] M. Humbert, B. Trubert, and K. Huguenin. A survey on interdependent privacy. *ACM Computing Surveys (CSUR)*, 52(6):1–40, 2019.
- [147] M. Humbert, D. Dupertuis, M. Cherubini, and K. Huguenin. Kgp meter: Communicating kin genomic privacy to the masses. In *2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P)*, pages 410–429. IEEE, 2022.
- [148] H. Ibrek and M. G. Morgan. Graphical communication of uncertain quantities to nontechnical people. *Risk analysis*, 7(4):519–529, 1987.
- [149] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Čapkun. Home is safer than the cloud! privacy concerns for consumer cloud storage. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, pages 1–20, 2011.
- [150] M. Isaac. WhatsApp Introduces End-to-End Encryption. April 2016. URL <https://>

//www.nytimes.com/2016/04/06/technology/whatsapp-messaging-service-introduces-full-encryption.html.

- [151] S. S. Iyengar and M. R. Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995, 2000.
- [152] M. Jagielski, J. Ullman, and A. Oprea. Auditing Differentially Private Machine Learning: How Private is Private SGD? In *Advances in Neural Information Processing Systems*, volume 33, pages 22205–22216. Curran Associates, Inc., 2020. URL <https://proceedings.neurips.cc/paper/2020/file/fc4ddc15f9f4b4b06ef7844d6bb53abf-Paper.pdf>.
- [153] M. Janic, J. P. Wijnbenga, and T. Veugen. Transparency enhancing tools (tets): An overview. In *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*, pages 18–25, 2013. doi: 10.1109/STAST.2013.11.
- [154] B. Jayaraman and D. Evans. Evaluating differentially private machine learning in practice. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 1895–1912, 2019.
- [155] J. Jin, E. McMurtry, B. I. P. Rubinstein, and O. Ohrimenko. Are we there yet? timing and floating-point attacks on differential privacy systems. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 473–488, 2022. doi: 10.1109/SP46214.2022.9833672.
- [156] B. Johnson. Privacy no longer a social norm, says Facebook founder. *The Guardian*, 2010. URL <https://www.theguardian.com/technology/2010/jan/11/facebook-privacy>.
- [157] N. Johnson, J. P. Near, J. M. Hellerstein, and D. Song. Chorus: a programming framework for building scalable differential privacy mechanisms. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 535–551. IEEE, 2020.
- [158] Jonathan Zong and J. Nathan Matias. Building Collective Power to Refuse Harmful Data Systems, Aug. 2020. URL <https://citizensandtech.org/2020/08/collective-refusal/>.
- [159] J. Jordon, J. Yoon, and M. Van Der Schaar. Pate-gan: Generating synthetic data with differential privacy guarantees. In *International conference on learning representations*, 2018.
- [160] V. Juárez Ramos. *Analyzing the role of cognitive biases in the decision-making process*. IGI Global, 2018.
- [161] B. Kacsmar, V. Duddu, K. Tilbury, B. Ur, and F. Kerschbaum. Comprehension from chaos: What users understand and expect from private computation. *arXiv preprint arXiv:2211.07026*, 2022.
- [162] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler. my data just goes everywhere:” user

- mental models of the internet and implications for privacy and security. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*, pages 39–52. Ottawa, 2015.
- [163] F. Karegar, A. S. Alaqra, and S. Fischer-Hübner. Exploring {User-Suitable} metaphors for differentially private data analyses. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 175–193, 2022.
- [164] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.
- [165] Katrina Ligett. The Elephant in the Room: The Problems that Privacy-Preserving ML Can’t Solve, Dec. 2020. URL <https://slideslive.com/38938421/the-elephant-in-the-room-the-problems-that-privacypreserving-ml-can-t-solve>.
- [166] D. Kekulluoglu, N. Kokciyan, and P. Yolum. Preserving privacy as social responsibility in online social networks. *ACM Transactions on Internet Technology (TOIT)*, 18(4):1–22, 2018.
- [167] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A “nutrition label” for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, pages 1–12, 2009.
- [168] P. G. Kelley, L. Cesca, J. Bresee, and L. F. Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 1573–1582, 2010. URL <https://doi.org/10.1145/1753326.1753561>.
- [169] N. Kesewaa Dankwa. “All Names are Pseudonyms”: A Critical Reflection on Pseudonymizing Names in HCI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. URL <https://doi.org/10.1145/3411763.3450376>.
- [170] R. Khare. Privacy theater: Why social networks only pretend to protect you. *TechCrunch*, 2009.
- [171] D. Kifer and A. Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 193–204, 2011.
- [172] D. Kifer and A. Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1):1–36, 2014.
- [173] D. Kifer, S. Messing, A. Roth, A. Thakurta, and D. Zhang. Guidelines for implementing and auditing differentially private systems. *arXiv preprint arXiv:2002.04049*, 2020.
- [174] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom. Derivation of new

readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch, 1975.

- [175] J. Kiss. Google admits collecting Wi-Fi data through Street View cars. May 2010. URL <https://www.theguardian.com/technology/2010/may/15/google-admits-storing-private-data>.
- [176] A. Kitkowska, E. Wästlund, J. Meyer, and L. A. Martucci. Is It Harmful? Re-examining Privacy Concerns. In M. Hansen, E. Kosta, I. Nai-Fovino, and S. Fischer-Hübner, editors, *Privacy and Identity Management. The Smart Revolution : 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers*, volume AICT-526 of *IFIP Advances in Information and Communication Technology*, pages 59–75. Springer International Publishing, 2018. doi: 10.1007/978-3-319-92925-5\5. URL <https://inria.hal.science/hal-01883632>. Part 3: Privacy in the Era of the Smart Revolution.
- [177] S. Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*, 64:122–134, 2017. URL <https://doi.org/10.1016/j.cose.2015.07.002>.
- [178] K. Kollnig, A. Shuba, M. Van Kleek, R. Binns, and N. Shadbolt. Goodbye tracking? Impact of iOS app tracking transparency and privacy labels. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 508–520, 2022.
- [179] K. Krombholz, K. Busse, K. Pfeffer, M. Smith, and E. Von Zezschwitz. “If HTTPS Were Secure, I Wouldn’t Need 2FA”-End User and Administrator Mental Models of HTTPS. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 246–263. IEEE, 2019.
- [180] J. A. Krosnick. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3):213–236, 1991.
- [181] J. A. Krosnick. Questionnaire design. In *The Palgrave handbook of survey research*, pages 439–455. Springer, 2018.
- [182] P. Kühtreiber, V. Pak, and D. Reinhardt. Replication: The effect of differential privacy communication on german users’ comprehension and data sharing attitudes. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 117–134, 2022.
- [183] K. Kupferschmidt. Twitter’s plan to cut off free data access evokes ‘fair amount of panic’ among scientists. *Science*, February 2023.
- [184] A. Lampinen, V. Lehtinen, A. Lehmuskallio, and S. Tamminen. We’re in it together: interpersonal management of disclosure in social network services. In *Proceedings of the*

SIGCHI conference on human factors in computing systems, pages 3217–3226, 2011.

- [185] E. J. Langer. The illusion of control. *Journal of personality and social psychology*, 32(2): 311, 1975.
- [186] A. Lapets, F. Jansen, K. D. Albab, R. Issa, L. Qin, M. Varia, and A. Bestavros. Accessible privacy-preserving web-based data analysis for assessing and addressing economic inequalities. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS '18*. Association for Computing Machinery, 2018. ISBN 9781450358163. URL <https://doi.org/10.1145/3209811.3212701>.
- [187] J. Lee and C. Clifton. How much is enough? Choosing ϵ for differential privacy. In X. Lai, J. Zhou, and H. Li, editors, *ISC 2011*, volume 7001 of *LNCS*, pages 325–340. Springer, Heidelberg, Oct. 2011.
- [188] L. A. Leotti, S. S. Iyengar, and K. N. Ochsner. Born to choose: The origins and value of the need for control. *Trends in cognitive sciences*, 14(10):457–463, 2010.
- [189] H. Li, N. Vincent, J. Tsai, J. Kaye, and B. Hecht. How do people change their technology use in protest? understanding. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–22, 2019.
- [190] T. Li, K. Reiman, Y. Agarwal, L. F. Cranor, and J. I. Hong. Understanding challenges for developers to create accurate privacy nutrition labels. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–24, 2022.
- [191] Y. Li, Y. Li, Q. Yan, and R. H. Deng. Privacy leakage analysis in online social networks. *Computers & Security*, 49:239–254, 2015.
- [192] Y. Li, A. Kobsa, B. P. Knijnenburg, M.-H. C. Nguyen, et al. Cross-cultural privacy prediction. *Proc. Priv. Enhancing Technol.*, 2017(2):113–132, 2017.
- [193] B. Y. Lim, Q. Yang, A. M. Abdul, and D. Wang. Why these explanations? selecting intelligibility types for explanation goals. In *IUI Workshops*, 2019.
- [194] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: understanding users’ mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 501–510, 2012.
- [195] H. R. Lipford, G. Hull, C. Latulipe, A. Besmer, and J. Watson. Visible flows: Contextual integrity and the design of privacy mechanisms on social network sites. In *2009 International Conference on Computational Science and Engineering*, volume 4, pages 985–989. IEEE, 2009.

- [196] E. Litt. Understanding social network site users' privacy tool use. *Computers in Human Behavior*, 29(4):1649–1656, 2013.
- [197] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin. When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)*, 54(2):1–36, 2021.
- [198] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11*, page 61–70. Association for Computing Machinery, 2011. ISBN 9781450310130. URL <https://doi.org/10.1145/2068816.2068823>.
- [199] D. Loviglio. Picking pseudonyms for your research participants, 2012. URL <https://blog.mozilla.org/ux/2012/05/picking-pseudonyms-for-your-research-participants/>.
- [200] M. Lyu, D. Su, and N. Li. Understanding the sparse vector technique for differential privacy. *Proc. VLDB Endow.*, 10(6):637–648, feb 2017. ISSN 2150-8097. URL <https://doi.org/10.14778/3055330.3055331>.
- [201] Z. Ma, J. Reynolds, J. Dickinson, K. Wang, T. Judd, J. D. Barnes, J. Mason, and M. Bailey. The impact of secure transport protocols on phishing efficacy. In *12th USENIX Workshop on Cyber Security Experimentation and Test (CSET 19)*, 2019.
- [202] M. Madden. Privacy, security, and digital inequality. 2017.
- [203] M. Madden, M. Gilman, K. Levy, and A. Marwick. Privacy, poverty, and big data: A matrix of vulnerabilities for poor americans. *Wash. UL Rev.*, 95:53, 2017.
- [204] M. Madejski, M. Johnson, and S. M. Bellovin. A study of privacy settings errors in an online social network. In *2012 IEEE international conference on pervasive computing and communications workshops*, pages 340–345. IEEE, 2012.
- [205] N. K. Malhotra, S. S. Kim, and J. Agarwal. Internet Users' Information Privacy Concerns (UIIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15(4):336–355, Dec. 2004. ISSN 1047-7047. URL <https://pubsonline.informs.org/doi/abs/10.1287/isre.1040.0032>. Publisher: INFORMS.
- [206] N. Malkin, J. Deatrck, A. Tong, P. Wijesekera, S. Egelman, and D. Wagner. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies*, 2019 (4), 2019.
- [207] J. Markoff. Growing compatibility issue: Computers and user privacy. *The New York Times*, 1999. URL <https://www.nytimes.com/1999/03/03/business/growing-compatibility>

-issue-computers-and-user-privacy.html.

- [208] M. Marsch, J. Grossklags, and S. Patil. Won't you think of others?: Interdependent privacy in smartphone app permissions. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021. URL <https://doi.org/10.1145/3479581>.
- [209] A. Marwick. Privacy without power: What privacy research can learn from surveillance studies. *Surveillance & Society*, 20(4):397–405, 2022.
- [210] A. E. Marwick and D. Boyd. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1):114–133, 2011.
- [211] A. E. Marwick and D. Boyd. Networked privacy: How teenagers negotiate context in social media. *New media & society*, 16(7):1051–1067, 2014.
- [212] A. E. Marwick and danah boyd. Privacy at the margins— understanding privacy at the margins—introduction. *International Journal of Communication*, 12:9, 2018.
- [213] A. Mayr and D. Machin. How to do critical discourse analysis: A multimodal introduction. *How to Do Critical Discourse Analysis*, pages 1–240, 2012.
- [214] A. Mazzia, K. LeFevre, and E. Adar. The PViz Comprehension Tool for Social Network Privacy Settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12*. Association for Computing Machinery, 2012. ISBN 9781450315326. URL <https://doi.org/10.1145/2335356.2335374>.
- [215] A. M. McDonald and L. F. Cranor. The cost of reading privacy policies. *Isjlp*, 4:543, 2008.
- [216] N. McDonald and A. Forte. The politics of privacy theories: Moving from norms to vulnerabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.
- [217] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019. URL <https://doi.org/10.1145/3359174>.
- [218] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.
- [219] F. D. McSherry. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, SIGMOD '09*, page 19–30. Association for Computing Machinery, 2009. ISBN 9781605585512. URL <https://doi.org/10.1145/1559845.1559850>.

- [220] Y. Meier, J. Schäwel, and N. C. Krämer. The shorter the better? effects of privacy policy length on online privacy decision-making. *Media and Communication*, 8(2):291–301, 2020.
- [221] S. Messing, C. DeGregorio, B. Hillenbrand, G. King, S. Mahanti, Z. Mukerjee, C. Nayak, N. Persily, B. State, and A. Wilkins. Facebook Privacy-Protected Full URLs Data Set, 2020. URL <https://doi.org/10.7910/DVN/TDOAPG>.
- [222] D. Metaxa, J. S. Park, R. E. Robertson, K. Karahalios, C. Wilson, J. Hancock, C. Sandvig, et al. Auditing algorithms: Understanding algorithmic systems from the outside in. *Foundations and Trends® in Human–Computer Interaction*, 14(4):272–344, 2021.
- [223] J. C. Meyer. Humor as a double-edged sword: Four functions of humor in communication. *Communication theory*, 10(3):310–331, 2000.
- [224] Microsoft. Microsoft SEAL: Fast and Easy-to-Use Homomorphic Encryption Library. <https://www.microsoft.com/en-us/research/project/microsoft-seal/>. Accessed 9/14/2023.
- [225] D. B. Miele and D. C. Molden. Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General*, 139(3):535–557, 2010. ISSN 1939-2222, 0096-3445. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0019745>.
- [226] G. Miklau. How Tumult Labs helped the IRS support educational accountability with differential privacy, July 2021. URL <https://www.tmlt.io/research/how-tumult-labs-helped-irs-support-educational-accountability-with-differential-privacy>.
- [227] S. J. Milberg, S. J. Burke, H. J. Smith, and E. A. Kallman. Values, personal information privacy, and regulatory approaches. *Communications of the ACM*, 38(12):65–74, 1995.
- [228] A. R. Miller. *The Assault on Privacy: computers, data banks, and dossiers*. Berichte der Universitaet von Michigan. Univ. of Michigan Press, Ann Arbor, 1971. ISBN 978-0-472-65500-7.
- [229] I. Mironov. On significance of the least significant bits for differential privacy. In T. Yu, G. Danezis, and V. D. Gligor, editors, *ACM CCS 2012*, pages 650–661. ACM Press, Oct. 2012. doi: 10.1145/2382196.2382264.
- [230] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260, 2010.
- [231] D. K. Mulligan, C. Koopman, and N. Doty. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Philosophical Transactions of the Royal*

- Society A: Mathematical, Physical and Engineering Sciences*, 374(2083):20160118, 2016.
- [232] P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. F. Cranor, and N. Sadeh. Privacy expectations and preferences in an IoT world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 399–412. USENIX Association Santa Clara, 2017.
- [233] P. Nanayakkara and J. Hullman. What’s Driving Conflicts Around Differential Privacy for the US Census. *IEEE Security & Privacy*, (01):2–11, 2022.
- [234] P. Nanayakkara, J. Bater, X. He, J. R. Hullman, and J. Duggan. Visualizing privacy-utility trade-offs in differentially private data releases. *Proceedings on Privacy Enhancing Technologies*, 2022:601 – 618, 2022.
- [235] P. Nanayakkara, M. A. Smart, R. Cummings, G. Kaptchuk, and E. M. Redmiles. What are the chances? explaining the epsilon parameter in differential privacy. In *32nd USENIX Security Symposium (USENIX Security 23)*, 2023. URL <https://www.usenix.org/conference/usenixsecurity23/presentation/nanayakkara>.
- [236] J. Near and D. Darais. Threat Models for Differential Privacy , 2020. URL <https://www.nist.gov/blogs/cybersecurity-insights/threat-models-differential-privacy>.
- [237] A. Ng. Teens have figured out how to mess with Instagram’s tracking algorithm, 2021. URL <https://www.cnet.com/news/teens-have-figured-out-how-to-mess-with-instagram-s-tracking-algorithm/>.
- [238] K. S. Niksirat, E. Anthoine-Milhomme, S. Randin, K. Huguenin, and M. Cherubini. “I thought you were okay”: Participatory Design with Young Adults to Fight Multiparty Privacy Conflicts in Online Social Networks. In *Designing Interactive Systems Conference (DIS)*, 2021.
- [239] H. Nissenbaum. Privacy in context: Technology, policy, and the integrity of social life. In *Privacy in Context*. Stanford University Press, 2009.
- [240] P. A. Norberg, D. R. Horne, and D. A. Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs*, 41(1):100–126, 2007. URL <https://doi.org/10.1111/j.1745-6606.2006.00070.x>.
- [241] M. Oates, Y. Ahmadullah, A. Marsh, C. Swoopes, S. Zhang, R. Balebako, and L. F. Cranor. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration. *Proceedings on Privacy Enhancing Technologies*, 2018(4):5–32, 2018.
- [242] J. A. Obar and A. Oeldorf-Hirsch. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information*,

Communication & Society, 23(1):128–147, 2020.

- [243] A.-M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, and J.-P. Hubaux. Quantifying interdependent privacy risks with location data. *IEEE Transactions on Mobile Computing*, 16(3):829–842, 2017. doi: 10.1109/TMC.2016.2561281.
- [244] Oracle. Mysql 8.0 reference manual, 2023. URL <https://dev.mysql.com/doc/refman/8.0/en/>.
- [245] M. T. Orne. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. In *Sociological methods*, pages 279–299. Routledge, 2017.
- [246] L. Padilla, M. Kay, and J. Hullman. *Uncertainty Visualization*, pages 1–18. 2021. ISBN 9781118445112. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat08296>.
- [247] Y. J. Park. Do men and women differ in privacy? gendered privacy and (in) equality in the internet. *Computers in Human Behavior*, 50:252–258, 2015.
- [248] J. Pater, C. Fiesler, and M. Zimmer. No humans here: Ethical speculation on public data, unintended consequences, and the limits of institutional review. 6(GROUP), jan 2022. URL <https://doi.org/10.1145/3492857>.
- [249] A. Pearce and E. Jiang. How randomized response can help collect sensitive information responsibly, 2020. URL <https://pair.withgoogle.com/explorables/anonymization/>.
- [250] M. Peterson. New iPhone privacy ad takes shots at other smartphones oversharing information, 2021. URL <https://appleinsider.com/articles/20/09/03/new-iphone-privacy-ad-takes-shots-at-other-smartphones-oversharing-information>.
- [251] Pew Internet Project. Trust and Privacy Online, Aug. 2000. URL <https://www.pewresearch.org/internet/2000/08/20/trust-and-privacy-online/>.
- [252] Pew Research Center. Mobile fact sheet. <https://www.pewresearch.org/internet/fact-sheet/mobile/>, Apr 2021. Accessed 7/29/2022.
- [253] I. Pollach. A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. *Journal of Business Ethics*, 62(3):221–235, 2005.
- [254] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Wortman Vaughan, and H. Wallach. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–52, 2021.

- [255] A. Puig. Equifax Data Breach Settlement: What You Should Know. *Federal Trade Commission Consumer Advice*, 2019. URL <https://consumer.ftc.gov/consumer-alerts/2019/07/equifax-data-breach-settlement-what-you-should-know>.
- [256] B. Quinn. ‘I’m a fighter not a quitter’: Truss channels Peter Mandelson at PMQs. October 2022. URL <https://www.theguardian.com/politics/2022/oct/19/im-a-fighter-not-a-quitter-truss-channels-peter-mandelson-at-pmqs>.
- [257] E. Rader, R. Wash, and B. Brooks. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS ’12. Association for Computing Machinery, 2012. ISBN 9781450315326. URL <https://doi.org/10.1145/2335356.2335364>.
- [258] E. Rader, S. Hautea, and A. Munasinghe. “I have a narrow thought process” constraints on explanations connecting inferences and self-perceptions. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, pages 457–488, 2020.
- [259] L. Rainie, S. Kiesler, R. Kang, M. Madden, M. Duggan, S. Brown, and L. Dabbish. Anonymity, privacy, and security online. *Pew research center*, 5, 2013.
- [260] F. Raja, K. Hawkey, S. Hsu, K.-L. C. Wang, and K. Beznosov. A brick wall, a locked door, and a bandit: a physical security metaphor for firewall warnings. In *Proceedings of the seventh symposium on usable privacy and security*, pages 1–20, 2011.
- [261] A. Rao, F. Schaub, N. Sadeh, A. Acquisti, and R. Kang. Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online. In *Symposium on Usable Privacy and Security*, SOUPS ’16, pages 77–96, Denver, Colorado, USA, July 2016. USENIX.
- [262] Y. Rashidi, A. Kapadia, C. Nippert-Eng, and N. M. Su. “It’s easier than causing confrontation”: Sanctioning Strategies to Maintain Social Norms and Privacy on Social Media. *Proceedings of the ACM on human-computer interaction*, 4(CSCW1):1–25, 2020.
- [263] S. J. Ray. Claim your calling and scale your action. In *A Field Guide to Climate Anxiety*, pages 52–79. University of California Press, 2020.
- [264] S. J. Ray. *A field guide to climate anxiety: how to keep your cool on a warming planet*. University of California Press Oakland, 2020.
- [265] E. Redmiles. Net benefits: Digital inequities in social capital, privacy preservation, and digital parenting practices of us social media users. In *proceedings of the international AAAI conference on web and social media*, volume 12, 2018.
- [266] E. M. Redmiles, S. Kross, and M. L. Mazurek. How I Learned to Be Secure: A Census-

- Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16*, page 666–677. Association for Computing Machinery, 2016. ISBN 9781450341394. URL <https://doi.org/10.1145/2976749.2978307>.
- [267] E. M. Redmiles, E. Liu, and M. L. Mazurek. You want me to do what? a design study of two-factor authentication messages. In *SOUPS*, volume 57, page 93, 2017.
- [268] E. M. Redmiles, Z. Zhu, S. Kross, D. Kuchhal, T. Dumitras, and M. L. Mazurek. Asking for a friend: Evaluating response biases in security user studies. In D. Lie, M. Mannan, M. Backes, and X. Wang, editors, *ACM CCS 2018*, pages 1238–1255. ACM Press, Oct. 2018. doi: 10.1145/3243734.3243740.
- [269] E. M. Redmiles, M. M. Bennett, and T. Kohno. Power in computer security and privacy: A critical lens. *IEEE Security & Privacy*, 21(2):48–52, 2023.
- [270] J. Reed and B. C. Pierce. Distance makes the types grow stronger: A calculus for differential privacy. *SIGPLAN Not.*, 45(9):157–168, sep 2010. ISSN 0362-1340. URL <https://doi.org/10.1145/1932681.1863568>.
- [271] F. N. Ribeiro, F. Benevenuto, and E. Zagheni. How biased is the population of facebook users? comparing the demographics of facebook users with census data to generate correction factors. In *Proceedings of the 12th ACM Conference on Web Science*, pages 325–334, 2020.
- [272] P. Rogaway. The moral character of cryptographic work. 2015. URL <https://eprint.iacr.org/2015/1162>.
- [273] M. Rosenberg, N. Confessore, and C. Cadwalladr. How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*, March 2018.
- [274] K. Salehzadeh Niksirat, D. Korka, H. Harkous, K. Huguenin, and M. Cherubini. On the potential of mediation chatbots for mitigating multiparty privacy conflicts - a wizard-of-oz study. 7(CSCW1), apr 2023. URL <https://doi.org/10.1145/3579618>.
- [275] I. Sander. What is critical big data literacy and how can it be implemented? *Internet Policy Review*, 9(2):1–22, 2020.
- [276] M. Santos and A. Faure. Affordance is power: Contradictions between communicational and technical dimensions of whatsapp’s end-to-end encryption. *Social Media+ Society*, 4(3):2056305118795876, 2018.
- [277] J. Sarathy. From algorithmic to institutional logics: the politics of differential privacy. Available at SSRN, 2022.

- [278] E. Sarigol, D. Garcia, and F. Schweitzer. Online privacy as a collective phenomenon. In *Proceedings of the second ACM conference on Online social networks*, pages 95–106, 2014.
- [279] P. Schaar. Privacy by design. *Identity in the Information Society*, 3(2):267–274, 2010.
- [280] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor. A Design Space for Effective Privacy Notices. In *Symposium on Usable Privacy and Security, SOUPS '15*, pages 1–17, Ottawa, Canada, July 2015. USENIX.
- [281] E.-M. Schomakers, C. Lidynia, D. Müllmann, and M. Ziefle. Internet users' perceptions of information sensitivity – insights from germany. *International Journal of Information Management*, 46:142–150, 2019. ISSN 0268-4012. doi: <https://doi.org/10.1016/j.ijinfo mgt.2018.11.018>. URL <https://www.sciencedirect.com/science/article/pii/S0268401218307692>.
- [282] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [283] Serrano v. Apple Inc. Case No. 23-cv-70, 2023. URL https://regmedia.co.uk/2023/01/10/apple_pa_complaint_wiretap.pdf.
- [284] H. Shen, A. DeVos, M. Eslami, and K. Holstein. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–29, 2021.
- [285] K. Shilton and D. Greene. Linking platforms, practices, and developer ethics: Levers for privacy discourse in mobile application development. *Journal of Business Ethics*, 155: 131–146, 2019.
- [286] P. Skeba and E. P. Baumer. Informational friction as a lens for studying algorithmic aspects of privacy. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2): 1–22, 2020.
- [287] M. A. Smart, D. Sood, and K. Vaccaro. Understanding risks of privacy theater with differential privacy. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), 2022. URL <https://doi.org/10.1145/3555762>.
- [288] E. Snowden. *Permanent Record*. Henry Holt and Company, 2019. ISBN 978-1-250-23724-8.
- [289] Social Science One. Our Facebook Partnership. <https://socialscience.one/our-facebook-partnership>, 2018.

- [290] K. L. Soeken and G. B. Macready. Respondents' perceived protection when using randomized response. *Psychological bulletin*, 92(2):487, 1982.
- [291] Software Freedom Conservancy. Selenium, 2021. URL <https://www.selenium.dev/documentation/>.
- [292] C. Soghoian. An end to privacy theater: Exposing and discouraging corporate disclosure of user data to the government. *Minn. JL Sci. & Tech.*, 12:191, 2011.
- [293] D. J. Solove. Conceptualizing privacy. *Calif. L. Rev.*, 90:1087, 2002.
- [294] D. J. Solove. A taxonomy of privacy. *University of Pennsylvania law review*, pages 477–564, 2006.
- [295] D. J. Solove. *Understanding privacy*. Harvard University Press, May, 2008.
- [296] D. J. Solove. Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.*, 126:1880, 2012.
- [297] D. J. Solove and C. J. Hoofnagle. A model regime of privacy protection. *U. Ill. L. Rev.*, page 357, 2006.
- [298] A. Solow-Niederman. Information privacy and the inference economy. *Nw. UL Rev.*, 117:357, 2022.
- [299] M. O. Source. React, 2023. URL <https://react.dev/>.
- [300] S. Spiekermann and L. F. Cranor. Engineering privacy. *IEEE Transactions on software engineering*, 35(1):67–82, 2008.
- [301] A. C. Squicciarini, M. Shehab, and F. Paci. Collective privacy management in social networks. In *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, page 521–530. Association for Computing Machinery, 2009. ISBN 9781605584874. URL <https://doi.org/10.1145/1526709.1526780>.
- [302] B. C. Stahl. Privacy and security as ideology. *IEEE Technology and Society Magazine*, 26(1):35–45, 2007.
- [303] L. Stark. The emotional context of information privacy. *The Information Society*, 32(1):14–27, 2016.
- [304] L. Stark. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students*, 25(3):50–55, 2019.

- [305] G. Stewart and D. Lacey. Death by a thousand facts: Criticising the technocratic approach to information security awareness. *Information Management & Computer Security*, 20(1): 29–38, 2012.
- [306] C. Stransky, D. Wermke, J. Schrader, N. Huaman, Y. Acar, A. L. Fehllhaber, M. Wei, B. Ur, and S. Fahl. On the limited impact of visualizing encryption: Perceptions of e2e messaging security. In *Seventeenth Symposium on Usable Privacy and Security*, pages 437–454, 2021.
- [307] J. M. Such and N. Criado. Multiparty privacy in social media. *Communications of the ACM*, 61(8):74–81, 2018.
- [308] J. M. Such and M. Rovatsos. Privacy policy negotiation in social media. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 11(1):1–29, 2016.
- [309] J. M. Such, J. Porter, S. Preibusch, and A. Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 3821–3832, 2017.
- [310] J. J. Suh, M. J. Metzger, S. A. Reid, and A. El Abbadi. Distinguishing group privacy from personal privacy: The effect of group inference technologies on privacy perceptions and behaviors. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), Nov. 2018. URL <https://doi.org/10.1145/3274437>.
- [311] S. Suh, S. Lamorea, E. Law, and L. Zhang-Kennedy. Privacytoon: Concept-driven storytelling with creativity support for privacy concepts. In *Designing Interactive Systems Conference*, pages 41–57, 2022.
- [312] L. Taber and S. Whittaker. “On Finsta, I can say ‘Hail Satan’”: Being Authentic but Disagreeable on Instagram. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.
- [313] J. Tan, K. Nguyen, M. Theodorides, H. Negrón-Arroyo, C. Thompson, S. Egelman, and D. Wagner. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 91–100, Toronto, Ontario, Canada, 2014. ACM Press. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557400. URL <http://dl.acm.org/citation.cfm?doid=2556288.2557400>.
- [314] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang. Privacy Loss in Apple’s Implementation of Differential Privacy on MacOS 10.12. *arXiv:1709.02753 [cs]*, Sept. 2017. URL <http://arxiv.org/abs/1709.02753>. arXiv: 1709.02753.
- [315] J. Tang, E. Birrell, and A. Lerner. Replication: How well do my results generalize now?

- The external validity of online privacy and security surveys. In *Eighteenth symposium on usable privacy and security (SOUPS 2022)*, pages 367–385, 2022.
- [316] H. T. Tavani and J. H. Moor. Privacy protection, control of information, and privacy-enhancing technologies. *ACM Sigcas Computers and Society*, 31(1):6–11, 2001.
- [317] A. G. Thakurta, A. H. Vyrros, U. S. Vaishampayan, G. Kapoor, J. Freudiger, V. R. Sridhar, and D. Davidson. Learning new words, Mar. 14 2017. US Patent 9,594,741.
- [318] The New York Times Opinion. The Privacy Project. April 2019. URL <https://www.nytimes.com/interactive/2019/opinion/internet-privacy-project.html>.
- [319] K. Thomas, C. Grier, and D. M. Nicol. unfriendly: Multi-party privacy risks in social networks. In *Privacy Enhancing Technologies: 10th International Symposium, PETS 2010, Berlin, Germany, July 21-23, 2010. Proceedings 10*, pages 236–252. Springer, 2010.
- [320] C. Timberg. Facebook made big mistake in data it provided to researchers, undermining academic work. *The Washington Post*, September 2021.
- [321] M. Tisné and M. Schaake. The data delusion: Protecting individual data isn’t enough when the harm is collective. *Luminate*, July, 2020.
- [322] A. Tong, E. Wang, and M. Coulter. Exclusive: Reddit in AI content licensing deal with Google. *Reuters*, February 2024.
- [323] F. Tramer, A. Terzis, T. Steinke, S. Song, M. Jagielski, and N. Carlini. Debugging differential privacy: A case study for privacy auditing. *arXiv preprint arXiv:2202.12219*, 2022.
- [324] K. S. Tsosie, J. M. Yracheta, and D. Dickenson. Overvaluing individual consent ignores risks to tribal participants. *Nature Reviews Genetics*, 20(9):497–498, 2019.
- [325] J. Turow, L. Feldman, and K. Meltzer. Open to exploitation: America’s shoppers online and offline. *Departmental Papers (ASC)*, page 35, 2005.
- [326] J. Turow, M. Hennessy, and N. Draper. The tradeoff fallacy: How marketers are misrepresenting american consumers and opening them up to exploitation. *Available at SSRN 2820060*, 2015.
- [327] J. Turow, M. Hennessy, and N. Draper. Persistent misperceptions: Americans’ misplaced confidence in privacy policies, 2003–2015. *Journal of Broadcasting & Electronic Media*, 62(3):461–478, 2018.
- [328] TwitterSupport. Your privacy matters. so does having the resources to understand and

manage your privacy settings. check out the updated ‘settings and privacy’ page on web and follow this thread to adjust your settings and make the twitter experience more your own. [tweet], October 2020. URL <https://twitter.com/TwitterSupport/status/1316455868225912833>.

- [329] United States of America v. Facebook, Inc. Case No. 19-cv-2184, 2019. URL https://www.ftc.gov/system/files/documents/cases/182_3109_facebook_order_filed_7-24-19.pdf.
- [330] B. Ur and Y. Wang. A cross-cultural framework for protecting user privacy in online social media. In *Proceedings of the 22nd International Conference on World Wide Web, WWW '13 Companion*, page 755–762. Association for Computing Machinery, 2013. ISBN 9781450320382. URL <https://doi.org/10.1145/2487788.2488037>.
- [331] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12*. Association for Computing Machinery, 2012. ISBN 9781450315326. URL <https://doi.org/10.1145/2335356.2335362>.
- [332] B. Ustun and C. Rudin. Supersparse linear integer models for optimized medical scoring systems. *Machine Learning*, 102:349–391, 2016.
- [333] K. Vaccaro, K. Karahalios, C. Sandvig, K. Hamilton, and C. Langbort. Agree or cancel? research and terms of service compliance. In *ACM CSCW Ethics Workshop: Ethics for Studying Sociotechnical Systems in a Big Data World*, 2015.
- [334] K. Vaccaro, C. Sandvig, and K. Karahalios. “At the End of the Day Facebook Does What It Wants” How Users Experience Contesting Algorithmic Content Moderation. *Proc. CSCW*, 2020.
- [335] F. Van Der Vlist and A. Helmond. Speculative data selfies. *Internet Policy Review*, 2017. URL <https://policyreview.info/articles/news/speculative-data-selfies/449>.
- [336] F. N. van der Vlist, A. Helmond, M. Burkhardt, and T. Seitz. API governance: the case of Facebook’s evolution. *Social Media+ Society*, 8(2):20563051221086228, 2022.
- [337] M. Van Kleek and K. O’Hara. The future of social is personal: The potential of the personal data store. *Social collective intelligence: Combining the powers of humans and machines to build a smarter society*, pages 125–158, 2014.
- [338] M. Van Kleek, R. Binns, J. Zhao, A. Slack, S. Lee, D. Ottewell, and N. Shadbolt. X-ray refine: Supporting the exploration and refinement of information exposure resulting from smartphone apps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

- [339] T. Van Leeuwen. *Discourse and practice: New tools for critical discourse analysis*. Oxford university press, 2008.
- [340] K. Vaniea, E. Rader, and R. Wash. Mental models of software updates. *International Communication Association*, pages 1–39, 2014.
- [341] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau, and O. Goga. Privacy risks with Facebook’s PII-based targeting: Auditing a data broker’s advertising interface. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 89–107. IEEE, 2018.
- [342] N. Vincent and B. Hecht. Can “conscious data contribution” help users to exert “data leverage” against technology companies? *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–23, 2021.
- [343] N. Vincent, B. Hecht, and S. Sen. “Data strikes”: evaluating the effectiveness of a new form of collective action against technology companies. In *The World Wide Web Conference*, pages 1931–1943, 2019.
- [344] N. Vincent, H. Li, N. Tilly, S. Chancellor, and B. Hecht. Data leverage: A framework for empowering the public in its relationship with technology companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 215–227, 2021. URL <https://dl.acm.org/doi/abs/10.1145/3442188.3445885>.
- [345] P. S. Visser, J. A. Krosnick, J. Marquette, and M. Curtin. Improving election forecasting: Allocation of undecided respondents, identification of likely voters, and response order effects. *Election polls, the news media, and democracy*. New York: Chatham House, 2000.
- [346] J. Vitak, K. Shilton, and Z. Ashktorab. Beyond the belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pages 941–953, 2016.
- [347] E. A. Vogels, R. Gelles-Watnick, and N. Massarat. Teens, social media and technology 2022. 2022.
- [348] C. Vreese, M. de Bastos, F. Esser, F. Giglietto, S. Lecleher, B. Pfetsch, C. Puschmann, R. Tromble, G. King, and N. Persily. Public statement from the co-chairs and european advisory committee of social science one. *Social Science One*, December 2019.
- [349] S. Wachter and B. Mittelstadt. A right to reasonable inferences: re-thinking data protection law in the age of big data and ai. *Colum. Bus. L. Rev.*, page 494, 2019.
- [350] P. Walker, P. Crerar, and J. Elgot. Liz Truss resigns as PM and triggers fresh leadership

- election. October 2022. URL <https://www.theguardian.com/politics/2022/oct/20/liz-truss-to-quit-as-prime-minister>.
- [351] Z. Wang, D. Y. Huang, and Y. Yao. Exploring tenants’ preferences of privacy negotiation in airbnb. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 535–551, 2023.
- [352] S. L. Warner. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309):63–69, Mar. 1965. ISSN 0162-1459, 1537-274X. URL <http://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480775>.
- [353] S. D. Warren and L. D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5): 193–220, 1890. ISSN 0017811X. URL <http://www.jstor.org/stable/1321160>.
- [354] R. Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, pages 1–16, 2010.
- [355] H. Watson, E. Moju-Igbene, A. Kumari, and S. Das. ” we hold each other accountable”: Unpacking how social groups approach cybersecurity and privacy together. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.
- [356] J. Weidman, W. Aurite, and J. Grossklags. On sharing intentions, and personal and interdependent privacy considerations for genetic data: A vignette study. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 16(4):1349–1361, July 2019. ISSN 1545-5963. URL <https://doi.org/10.1109/TCBB.2018.2854785>.
- [357] G. Weinberg. What Is the Business Model for DuckDuckGo? 2017. URL <https://spreadprivacy.com/duckduckgo-revenue-model/>.
- [358] B. Weinshel, M. Wei, M. Mondal, E. Choi, S. Shan, C. Dolin, M. L. Mazurek, and B. Ur. Oh, the places you’ve been! user reactions to longitudinal transparency about third-party web tracking and inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 149–166, 2019.
- [359] A. Whitten and J. D. Tygar. Why johnny can’t encrypt: A usability evaluation of pgp 5.0. In *USENIX security symposium*, volume 348, pages 169–184, 1999.
- [360] P. Wisniewski, H. Xu, H. Lipford, and E. Bello-Ogunu. Facebook apps and tagging: The trade-off between personal privacy and engaging with friends. *Journal of the Association for Information Science and Technology*, 66(9):1883–1896, 2015.
- [361] Y. Wu, P. Gupta, M. Wei, Y. Acar, S. Fahl, and B. Ur. Your secrets are safe: How browsers’ explanations impact misconceptions about private browsing mode. In *Proceedings of the*

2018 World Wide Web Conference, pages 217–226, 2018.

- [362] Y. Wu, W. K. Edwards, and S. Das. “A Reasonable Thing to Ask For”: Towards a Unified Voice in Privacy Collective Action. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2022.
- [363] Y. Wu, W. K. Edwards, and S. Das. Sok: Social cybersecurity. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1863–1879. IEEE, 2022.
- [364] M. Wuerthele. ‘Privacy. That’s iPhone’ ad campaign launches, highlights Apple’s stance on user protection. March 2019. URL <https://appleinsider.com/articles/19/03/14/privacy-thats-iphone-ad-campaign-launches-highlights-apples-stance-on-user-protection>.
- [365] W. Xie, A. Fowler-Dawson, and A. Tvauri. Revealing the relationship between rational fatalism and the online privacy paradox. *Behaviour & Information Technology*, 38(7): 742–759, 2019. URL <https://doi.org/10.1080/0144929X.2018.1552717>.
- [366] A. Xiong. Effect of facts box on users’ comprehension of differential privacy: A preliminary study. In *Proceedings of the Human Factors and Ergonomics Society 2020 Annual Meeting*, 2020.
- [367] A. Xiong, T. Wang, N. Li, and S. Jha. Towards effective differential privacy communication for users’ data sharing decision and comprehension. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 392–410. IEEE, 2020.
- [368] A. Xiong, C. Wu, T. Wang, R. W. Proctor, J. Blocki, N. Li, and S. Jha. Using illustrations to communicate differential privacy trust models: An investigation of users’ comprehension, perception, and data sharing decision. *ArXiv*, abs/2202.10014, 2022.
- [369] K. Xu, Y. Guo, L. Guo, Y. Fang, and X. Li. My privacy my decision: Control of photo sharing on online social networks. *IEEE Transactions on Dependable and Secure Computing*, 14(2):199–210, 2015.
- [370] A. X. Zhang, M. S. Bernstein, D. R. Karger, and M. S. Ackerman. Form-from: A design space of social media systems. *Proceedings of the ACM on Human-Computer Interaction*, 2024.
- [371] D. Zhang and D. Kifer. Lightdp: Towards automating differential privacy proofs. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages*, POPL ’17, page 888–901. Association for Computing Machinery, 2017. ISBN 9781450346603. URL <https://doi.org/10.1145/3009837.3009884>.
- [372] L. Zhang-Kennedy, S. Chiasson, and R. Biddle. Password advice shouldn’t be boring: Visualizing password guessing attacks. In *2013 APWG eCrime Researchers Summit*,

- pages 1–11, 2013. doi: 10.1109/eCRS.2013.6805770.
- [373] L. Zhang-Kennedy, S. Chiasson, and R. Biddle. The role of instructional design in persuasion: A comics approach for improving cybersecurity. *International Journal of Human-Computer Interaction*, 32(3):215–257, 2016.
- [374] X. Zhao, C. Lampe, and N. B. Ellison. The social media ecology: User perceptions, strategies and challenges. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 89–100, 2016. URL <https://dl.acm.org/doi/abs/10.1145/2858036.2858333>.
- [375] M. Zimmer. “But the data is already public”: On the ethics of research in Facebook. In *The Ethics of Information Technologies*, pages 229–241. Routledge, 2020.
- [376] M. Zimmer and E. M. Chapman. Ethical review boards and pervasive data research: Gaps and opportunities. 2020. URL <https://doi.org/10.5210/spir.v2020i0.11369>.
- [377] J. Zong and J. N. Matias. Data refusal from below: A framework for understanding, evaluating, and envisioning refusal as design. *ACM Journal on Responsible Computing*, 1(1):1–23, 2024. URL <https://dl.acm.org/doi/full/10.1145/3630107>.
- [378] Y. Zou and F. Schaub. Concern but no action: Consumers’ reactions to the equifax data breach. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA ’18, page 1–6. Association for Computing Machinery, 2018. ISBN 9781450356213. URL <https://doi.org/10.1145/3170427.3188510>.
- [379] Y. Zou, K. Roundy, A. Tamersoy, S. Shintre, J. Roturier, and F. Schaub. Examining the adoption and abandonment of security, privacy, and identity theft protection practices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020. URL <https://dl.acm.org/doi/abs/10.1145/3313831.3376570>.
- [380] S. Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 2018. ISBN 1610395697.
- [381] M. Zuckerberg. Zuckerberg transcripts, 2011. URL https://epublications.marquette.edu/zuckerberg_files_transcripts/312.