Informing the Design and Refinement of Privacy and Security Controls

Daniel Smullen

CMU-ISR-21-111

September 2021

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Norman Sadeh (Chair)
Lorrie Faith Cranor
Alessandro Acquisti
Rebecca Weiss (Mozilla)
Yaxing Yao (University of Maryland, Baltimore County)

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Software Engineering.

This research was supported in part by grants from DARPA and AFRL under the Brandeis project on Personalized Privacy Assistants For the Internet of Things (FA8750-15-2-0277), by grants from the National Science Foundation Secure and Trustworthy Computing program (CNS-15-13957, CNS-1801316, CNS-1914486), and by an unrestricted grant from Mozilla.

Keywords: Privacy, security, usability, settings, awareness, control, mental models, machine learning

This work is dedicated to my father and mother, Merril and Norine, who always believed I would accomplish great things.

Abstract

Amid increasing privacy and security risks, managing one's privacy and security settings is becoming ever more important. Yet, the proliferation of security and privacy controls is making this task overwhelmingly complex. Are they the right controls? Are they effective? This dissertation's objective is to study how effective existing settings are, assess whether they give users the awareness and control they need, and to inform ways to improve them. We begin by examining how people interact with browsers' privacy and security settings. This is followed by a study designed to inform the development of more effective settings and defaults. Finally, we explore machine learning techniques with the aim of helping users configure their settings and further reduce user burden. Our results form the basis for our recommendations to improve privacy and security controls, the discussion of public policy implications, and generalizability to other domains.

Our first study explores people's ability to identify, understand, and control common data practices associated with privacy and security risks (e.g., fingerprinting, behavioral profiling, targeted ads) in their primary browser. Our results highlight some design choices in browsers which seem to work well for our participants, and some which need improvement. Though all of the browsers we studied offered unique settings, many were confusing and misaligned with the mental models of our participants. Specific and detailed descriptions of data practices seemed to help alleviate some confusion, but technical jargon and inconsistent terminology seemed to exacerbate it. Our findings suggest that the resulting lack of confidence may leave people vulnerable to risks that they are unable to effectively mitigate. Browsers should do more to educate their users, focusing on how they can ameliorate their privacy and security concerns using consistent language.

However, even if browsers offered clearer settings, ad hoc settings on websites can frustrate users. Many are redundant, and some offer no control at all. Our second study focuses on what might work better for people to manage a broader collection of online data practices more comprehensively. Our results suggest that the existing patchwork of settings may mislead people about the extent of their control; most users would prefer restrictive defaults within certain categories of websites, but have no way to express such preferences. Moreover, if all the required settings were available, accommodating people's diverse preferences would become an overwhelming and repetitive task on every website. Fortunately, we discovered commonalities in people's preferences among different contexts. These commonalities enable settings to be consolidated. Browsers which leverage this could permit data practices to be managed by users centrally in a single standardized interface. Browsers could then enforce users' preferences automatically as they browse. Beyond reducing user burden, this standards-based approach would offer a more consistent management

experience, building on our first study's findings. However, for this to work websites would also have to conform to standards requiring that they honor settings communicated by browsers – which has been resisted by industry so far.

Consolidation and reducing repetition can help reduce user burden, but this alone may not be enough to ensure users can effectively express their preferences. For our third study, the next logical step was to explore mobile app permissions, which incorporate standard app categories and settings to allow or deny access to sensitive data and APIs. Nevertheless, mobile app permission settings poorly align with people's mental models as they omit factors (such as purpose) that influence people's privacy and security decisions. The settings are already overwhelming, yet there is no distinction between permissions granted for different purposes such as advertising, versus core app functionality. Settings distinguishing among different purposes would increase the number of permissions, further increasing user burden. However, as seen in browser settings, we found correlations in people's mobile app permission settings. Despite being more complex, permissions which included purpose yielded additional predictive power and this can be leveraged with machine learning to make better recommendations. We show that this approach has the potential to overcome trade-offs between accuracy and user burden by effectively reducing the number of decisions users would need to make, despite also offering more complex settings.

This dissertation explores a broad cross section of privacy and security decisions, systematically exploring their effectiveness and manageability. We reveal that existing privacy and security controls may not be effectively addressing people's concerns or expectations. However, the problem is fundamentally about having appropriate settings, not necessarily the most options, as this fails to consider the limits of what people are realistically capable of configuring. To avoid redundancy and confusion, the settings also need to align with people's mental models. Moreover, people's diverse preferences and concerns can align across categories of apps and websites, data practices, purposes, and many other factors – these can form the basis for consolidation and standardization. Yet standardized settings, such as mobile app permissions, can still be misaligned with people's mental models. Simply adding more expressive settings is a tempting solution but improving control and effectiveness by proliferating settings can trade-off manageability and increase user burden. Machine learning can simplify the task of managing one's settings, which can help to overcome this trade-off. Privacy and security controls can be redesigned to be more effective – without exceeding users' ability to configure them.

Acknowledgments

Without the help and guidance of my advisor, Dr. Norman Sadeh, this work would never have been possible. I would also like to thank my other committee members, Dr. Lorrie Faith Cranor, Dr. Alessandro Acquisti, Dr. Rebecca Weiss, and Dr. Yaxing Yao, who contributed invaluable feedback that helped shape this work for the better.

Thanks also to the staff at the Institute for Software Research, particularly Linda Moreci, Nick Frollini, and Connie Herold, who kept everything organized and running smoothly.

My other collaborators deserve recognition for the part they played in helping me succeed, including: Abhilasha Ravichander, Dr. Arthur Edelstein, Dr. Florian Schaub, Dr. Hana Habib, Dr. Hanan Hibshi, Dr. Joel Reidenberg (rest in peace), Rex Chen, Dr. Sebastian Zimmeck, Shikun (Aerin) Zhang, Dr. Shomir Wilson, Siddhant Arora, and Dr. Yuanyuan Feng. In particular, I would like to acknowledge Dr. Peter Story, who was always there for me when times were tough.

Special thanks to my good friends Dr. Aaron Harlap, Dr. Cody Kinneer, Christopher Karpurk, Dr. Iain Cruickshank, Dr. Janos Szurdi, Dr. Joshua Tan, Dr. Mahmood Sharif, Dr. Michael Maass, Roger Iyengar, and Dr. Marat Valiev.

Thank you to my partner, Dr. Elizabeth Landzberg, who was always patient with me.

Finally, I would like to thank everyone not mentioned here who helped me along the way – you know who you are. Thank all you for your guidance, friendship, and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation is about improving the management of privacy and security controls. What do we mean by privacy and security, and why are controls needed? Data privacy has classically been characterized as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." [122] Contextual integrity theory provides an important nuance to this definition, adding that though people's individual preferences vary, they are also highly context-dependent [15]. Accommodating diverse personal choices is the most important guiding principle under this definition of data privacy. In contrast, information security focuses on defeating specific security threats. Usable Security experts in particular have recognized how difficult it is to choose the appropriate settings, advocating to not offer any choices when there is one clearly optimal or secure default [29]. Yet, privacy advocates promote myriad choices because "one size fits all" privacy settings rarely exist [124]. It would be a mistake to generalize from these statements that designing privacy and security settings is difficult because they conflict with one another. Privacy and security do not fundamentally conflict [106]. What makes designing privacy and security controls hard is not designing the most comprehensive settings, nor offering the simplest settings, but rather finding appropriate settings – that is, those which can actually address the concerns that users have.

Different people have diverse sets of concerns, varying degrees of tolerance for risk, and different levels of confidence in their ability to steer clear from threats or mitigate risks. Given that there is no universally optimal default that will satisfy every user in every possible situation, privacy and security controls should be both based around individual risk tolerances as well as preferences, which can in turn be

influenced by myriad human factors such as cost constraints, effort, time, expertise, and so on. It goes without saying that in cases where controls are needed, they should be offered. Naturally, most software offers ways to configure at least some form of privacy and security controls. Ideally, the way that these controls are offered, and the functionality they provide should be designed to ensure that the following statements are true:

- That people can understand the practices they may encounter.
- That people can understand the potential risks and implications involved.
- That people have the ability to restrict system behaviors to suit their preferences and risk tolerance.

Software developers can evaluate the design of their privacy and security controls by considering two simple questions, which are central to this work:

1. Does the system do a good enough job at informing users about the practices that they might potentially be concerned about?
2. Do users have the ability to restrict the practices that specifically concern them?

Unfortunately, privacy and security incidents are continuous reminders that what is offered to users has ample room for improvement. Many of today's security and privacy controls continue to fall short, exposing people to unnecessary risks. While many more controls have been made available to users in response to the increasingly diverse arrays of practices they may need to control, the controls have become unwieldy. Users are often unaware of their existence and have a poor understanding of what they can or cannot do. This leads to unrealistic expectations held by users, but users are also confronted with unrealistic expectations imposed on them by developers. This is especially evident when it comes to the unrealistic amount of time and effort users would have to devote to align their settings with their preferences and tolerance for risk. Therefore, in this work we ask the following overarching research question: how effective are today's privacy and security controls, and how can we improve on the current situation? We begin by evaluating the state of the art, starting with the approaches taken by popular web browsers. This evaluation sets the stage for a deeper exploration on how to improve what is offered online, on mobile, and elsewhere, informed by both our novel evaluation as well as the issues seen in prior literature.

1.1 Improving Privacy and Security Controls

Designing effective privacy and security controls is a hard problem, and our first technical chapter evaluates the state of the art. It is unclear how well browsers offer ways to recognize and control potentially intrusive (and insecure) data practices such as: fingerprinting, behavioral profiling, and crypto-mining. As an initial exploratory study, we examined the ways in which users interact with today's 5 most popular browsers. We sought to uncover shortcomings that need to be addressed, and identify features and design choices which are working well in certain browsers. However, throughout our work we aim to move beyond the state of the art, by informing best practices. Accordingly, we look for more effective ways to ensure people are being made aware of privacy and security risks, and their options for controlling them. Based on guidance from prior literature [96], we also aim to promote better standards to ensure consistency in awareness and control mechanisms across the web, mobile platforms, the Internet of Things, and beyond. Concretely, we identify ways in which many privacy and security controls can become overly burdensome or inaccurate as a result of sub-optimal design.

In our second study, we explore rich user perspectives on controlling an even broader variety of online practices, irrespective of individual browsers. Specifically, we survey users' preferences and expectations for control as well as their desire to be notified about online data practices. Using a mixed-methods approach, we analyze how people would prefer to be able to configure their browsers in an attempt to find ways to improve what is currently offered. Based on our findings, we suggest concrete ways to address some of the problems that are associated with the ad-hoc solutions offered on individual websites. We observed that users are confused about the extent of their ability to restrict certain practices. Improved settings which offer the control they expect to have on individual websites would be overly burdensome. Many websites offer settings which are too complex, while many others offer only trivial settings that are too inaccurate; these controls often fail to achieve their goals due to users' unwillingness to make the effort to use them, various usability issues, or misconceptions about how and when to use them. Our recommendations aim to standardize settings along factors which we show to better align with users' mental models and preferences, such as categories of websites and individual practices. This better-aligned model, combined with restricting intrusive practices by default, can decrease burden without sacrificing accuracy. However, this recommended approach would also require standardized settings and APIs which do not currently exist. In principle, our recommendations would move standard browser settings to more closely resemble mobile app permissions. In addition to these standards, we also

propose regulations which could mitigate the possibility of websites intentionally breaking when users express stricter privacy and security preferences that conflict with websites' business goals. We have identified this as a possible dark pattern, where websites may attempt to coerce users into relaxing their settings based on the prior knowledge that users tend to disregard their privacy and security preferences if they come in the way of completing their original task.

In our third study, we move beyond web browsers to study the potential for improving mobile app permissions. The design of privacy and security controls in web browsers are guided only by limited standards (many of which are voluntary or no longer supported [102]), resulting in settings which can vary from website to website. In contrast, within a given ecosystem, mobile app permissions are standardized and are uniformly enforced by mobile operating systems. Yet despite app permissions already conforming to well-defined standards, configuring permissions remains an overwhelming task given the explosive growth of apps and their increased use of sensitive permissions. Research has repeatedly shown that people express different preferences depending on the purpose for which a permission is being requested by an app [69, 72, 112]. Mobile app permissions currently do not capture this factor, but including settings subject to purpose would further increase burden. Mobile apps are in need of ways to provide more comprehensive settings without increasing the existing burden, which is already unrealistically high. Here, intelligent recommendations offer the promise of simplifying the management of complex and numerous settings. Using a combination of supervised and unsupervised machine learning techniques, we show that it is possible to shift the burden away from users. Decision support can provide recommendations based on more complex and nuanced factors. These factors would make permissions prohibitively burdensome if they were naively introduced into existing permissions. By sweeping the parameter space of this approach, we are able to demonstrate that different parameters are capable of increasing accuracy, reducing user burden, or balancing both objectives simultaneously. By quantifying the relationship between these parameters and the expected accuracy versus user burden, we provide concrete guidance that can enable developers to tailor their solutions according to their accuracy requirements or user burden limits.

1.2 Key Dimensions of Privacy and Security

In this work, we focused on promoting two key dimensions of privacy and security: *awareness* and *control*. For users to be able to mitigate both security and privacy threats, they need to be provided with information about the potential risks, and the

implications of mitigating or accepting them. To provide adequate control, systems must also provide users with the ability to meaningfully express their preferences to accept or mitigate these risks. This dual requirement applies both in a privacy context (where this concept is often referred to as "notice and choice" [30]) and also in a security context. Though privacy and security conceptually overlap to a large degree, we would like to avoid overextending notice and choice beyond its intended scope within privacy. Awareness and control are concepts which more coherently apply to both privacy and security.

Our definition of awareness subsumes notice. Beyond notice, awareness additionally refers to understanding the risks which may be present in a given scenario. In contrast, notice is principally about the act of facilitating perception, observation, and understanding of one's mitigation options. Notice and awareness both come with an inherent concept of usability. For example, an impenetrable wall of text is unlikely to effectively make users aware of anything [97]. However, the scope of this work makes it insufficient to refer to the principle of notice alone when we discuss privacy and security. Notice alone cannot sufficiently address the clandestine, obfuscated, or obscure nature of many contemporary privacy and security risks, nor can it address the disconnect between risks and their mitigation strategies. Similarly, our definition of control subsumes choice. Choice describes an option, a decision, or an opportunity to choose; to express some specific preference or preferences out of many. In contrast, control also refers to having some authority or ability to influence a given scenario – potentially beyond the clear-cut set of available options. This broader definition is important when we consider scenarios in which the choices offered by a system may be misaligned with the privacy and security risks they purport to address. The lens of control also provides a better way of evaluating situations where users' mental models poorly align with their immediately obvious options. These situations can result in awkward, constrained, or impossible to express choices (all of which our work aims to alleviate wherever possible).

Beyond providing a way to describe key dimensions of privacy and security, awareness and control are intended to provide a conceptual framework for addressing and mitigating privacy and security risks. Techniques such as machine learning, fingerprinting, profiling, and other forms of automated reasoning are becoming increasingly sophisticated, pervasive, and capable [64, 5, 121, 116]. People may now experience such practices nearly constantly as they engage with software [99]. Unfortunately, in spite of their potential benefits, these practices may also expose users to privacy and security risks; users may object to practices involving surveillance, may experience threats to data confidentiality, and may feel violated if subjected to these practices

without their consent. Such data collection may also result in insecurity through unanticipated dissemination, breach, or secondary usage [6, 127, 25, 105].

1.3 Human Limitations

In theory, awareness and control empowers users to make themselves aware of the practices that they are subject to, understand their options, and take action to restrict those that they deem unacceptably intrusive or risky [30, 94]. In a practical sense, control is confined within the design parameters of what systems are capable of providing to users. The basic assumption software developers make is that settings are provided to users with the expectation that users will understand and make use of all of them. This approach results in the configuration burden being placed solely on users. However, users have limited attention, and configuring their settings is a secondary task [3]. This trade-off between increased burden and more comprehensive settings makes designing usable settings even more difficult, and evaluating whether a particular design has incorporated the right trade-off is critical. This also does not discount the importance of default settings (which should be conservative) to ensure that users are initially protected until or unless they decide to opt for less risk-averse settings. Designs which incorporate more expressive, granular settings can potentially take into account different contexts. They can also enable settings to be more accurate, and can better account for users' various individual preferences – ideally, this is what developers should strive for in their designs. On the other hand, engaging with more complex settings comes at the cost of a higher cognitive and attentional burden [18]. If managing their settings becomes too burdensome, users will refuse to engage meaningfully with the settings altogether [5, 4, 87]. Users need to have settings which make sense to them, can address their concerns, and are not overly complex.

The trade-off between accuracy and user burden in privacy and security settings is an understudied phenomenon, despite being evident in the literature for many years [18, 71]. Each of these factors and more, in different contexts, is known to contribute to a measurable change in a particular individual's preferred settings [15]. Overly simplistic or one-size-fits-all solutions are therefore unlikely to be satisfying for many users [124]. Effective designs must strive to maximize accuracy (the ability to *correctly* capture users' preferences) and strive to minimize user burden (the amount of effort users must endure to *enact* their preferences). More importantly, choosing one particular approach to offering settings over another should be based on a principled approach that considers what best captures what users want.

1.4 Philosophical Underpinnings and Framing

Our work is primarily based on a standpoint of *libertarian paternalism* [113], which seeks to preserve freedom of choice, but also incorporates guidance which is intended to steer developers and users in a direction that will promote the users' welfare. In this way, we aim to advance the common good of society. We argue for a broader impact, beyond improving the existing libertarian choice architectures. The consequences of developers' choices have potential negative externalities, and when considering the implications of different choice architectures, improving controls (given the possible trade-off choices) creates the potential to reduce some of these negative externalities. Throughout this work, we have highlighted trade-offs (such as accuracy conflicting with user burden) which may have consequences for some abstract portion of users. While one goal of this work is to inform best practices for designers, we also acknowledge that designers are faced with the decision of what is optimal in their view, and their philosophy may or may not coincide with the choice architecture we envision as part of this work. In particular, we argue for policies which could potentially impose constraints on developers, but more importantly these constraints are intended to help realign business incentives and rightfully place control back in the hands of users. For example, in our second technical chapter we offer that public policy could protect users from externalities brought by systems whose misaligned incentives may compel users to abandon their privacy and security preferences in favor of accomplishing a particular task.

Libertarian paternalism [113] also leads to the justification of certain defaults over others. In this work, we argue for standards and defaults which maximally align available options with many users' preferences and mental models while promoting the design of systems which can still uphold individualistic choices. Acknowledging that users' expressed preferences can become meaningless and incoherent within bounded rationality [5], we also argue for defaults which reduce user burden wherever possible. Accordingly, we have identified that there are opportunities to overcome certain design trade-offs related to user burden in the face of overwhelming choices, thereby mitigating negative externalities which are faced by overwhelmed users. This is accomplished through systems which are capable of accurately providing paternalistic recommendations.

Principally, whether any individual chooses to follow our recommendations is ultimately their decision, but our goal is to meaningfully empower users without worsening the paradox of choice [21]. To achieve this, it is insufficient to be solely focused on the three hallmark considerations of a *traditional libertarian* [114] posi-

tioning: maximizing the available options, elevating the level of control users have, or enabling the broadest set of possible preferences. Our view is that the software industry appears to be adopting this philosophy, and it may be the case that this is simply because browser providers, app store operators, and others simply could not come up with a single set of settings. Many software developers are correctly recognizing that a "one size fits all" approach to privacy and security settings will not work [124]. However, in the process they devolve responsibility to users, requiring them to manage complex privacy and security decisions that they do not fully understand.

This is a trend which has been going on for a very long time in privacy and security – in particular with privacy, as regulations in the United States do not mandate privacy protective defaults. At the end of the day, most technologies adopt mixed approaches, where some practices are considered too egregious to be tolerated, and others are viewed as practices that users should be given the flexibility of controlling themselves. However, users are expected to use controls that they are often unaware of, and unable to properly use. Thus, the second order effects — on society — are potentially far larger than what negative effects individual users may experience. Accommodating broad preferences but with controls that fail to allow choices to be consistently expressed [96], or controls which are so overwhelming that they offer no meaningful choices at all [30, 29], may cause a collective loss of confidence in user agency. They may also cast doubt on users' ability to achieve privacy and security entirely.

In this dissertation, we study these problems, starting by observing how a particular set of users interacts with and understands the controls made available for managing data practices in their primary browser. Faced with confusion and lack of confidence, our participants demonstrated that they did not reflect the design assumptions inherent in their browsers. Next, we explore the management of a broader set of data practices holistically, with the aim of improving the settings which are offered by increasing their alignment with people's mental models. We offer that settings can be consolidated significantly, which would make them far more tractable to configure. This further challenges the purely libertarian philosophy. Finally, we explore the use of machine learning to further reduce user burden. Our findings show that machine learning has the capability of both reducing user burden and improving the alignment of settings with people's mental models. Using paternalistic recommendations, users can be empowered to configure a larger number of more complex settings while simultaneously making fewer decisions.

1.5 Thesis Summary

This dissertation is comprised of three technical chapters, each based on a separate research study. The first technical chapter explores the effectiveness of the privacy and security interfaces present in today's most popular browsers. It is unclear whether the privacy and security options in these browsers are offering awareness and control to users effectively. Our work answers this question by exploring how different alternatives offer different levels of effectiveness, both in terms of users' understanding, and their ability to make use of them. Our approach is centered around an interview study. We apply thematic analysis to arrive at our results, and discuss the implications of our findings for each of the five browsers we study.

The second technical chapter focuses on better understanding what users want and expect, beyond the constraints of any particular browser. In this chapter, we explore user-centered perspectives on the management of intrusive practices encountered during web browsing. This exploration is centered around a mixed-methods study. By improving the understanding of users' perspectives, we make the case for changes which can enable browsers to improve the alignment between what they offer and what users expect them to offer. We show evidence that this could be accomplished by using standardized models of context-specific settings.

The final technical chapter determines whether there is a way to help people get what they want by exploring the potential of using machine learning to help users with their privacy and security decisions. This chapter centers around mobile app permissions, which employ a greater degree of standardization than browsers.

1.5.1 Main Contributions

This dissertation identifies several common shortcomings in current browsers' and mobile app permission managers' abilities to make users aware of privacy and security risks, and provide effective and manageable controls. Known problems associated with configuring online privacy and security controls are plentiful. An even deeper exploration can also highlight areas where the controls are working well, and where they are still falling short. This is a key part of our contribution. As part of this evaluation, we determine and characterize the trade-offs that are inherent in today's popular browsers' various settings and interfaces – some offer granular information and controls, while others are minimalist. We offer a thorough study of the way in which users are made aware of, interpret, understand, and make use of these browsers' varied controls. The findings of this study lead naturally into

further questions about users' preferences, and what they want to control versus what they are actually able to do in reality. These further questions are addressed in the second technical chapter.

This dissertation identifies ways to improve software engineering practice. Studying the effectiveness and manageability of security and privacy controls should be viewed as part of the software engineering process. Today, many technologies (including browsers and mobile apps) feature controls that are often misaligned with people's mental models, and are not sufficiently prominent for users to be aware of their existence. Our findings point to the need for ways to better educate and inform users, offer more rational controls, and incorporate more sensible defaults. If the state of practice is to move forward, we show that the following issues must be addressed in software generally: First, interfaces for managing privacy and security settings are in need of improvement and require principled and systematic evaluation. Second, people's preferences are often not well aligned with what is offered, and more effective standardized controls would be simpler and easier to configure. In the case of browsers, standardizing controls around factors such as website categories and intrusive practices can enable browsers to be a neutral platform for expressing people's preferences more broadly. Third, we must address public policy issues associated with standardization, such as the need for APIs, and the need to restructure incentives that have historically resulted in website operators resisting voluntary standards [111, 102]. Finally, in controls that already employ standardization (such as mobile app permissions) machine learning shows the potential to take advantage of complex information to make configuration easier.

This dissertation provides design guidance for practitioners that will enable them to maximize the benefits, accuracy, and acceptance of the controls they provide while minimizing user burden. Good software engineering is about making appropriate trade-offs, and we show that it is possible to understand, mitigate, and even overcome the trade-offs inherent in designing privacy and security controls. The design guidance we offer in this work is intended to promote approaches which can increase the accuracy of privacy and security settings, while minimizing unnecessary burden. We show that there is the potential to use the data we derive about users' preferences to develop and evaluate models of alternative browser settings. These models enable us to make scientific claims about whether people will be more or less likely to accept these options, based on their accuracy and level of user burden. Combining qualitative perspectives and quantitative preferences, we inform a deeper understanding of the way users reflect on the practices they may potentially encounter while browsing. Our results argue for denying intrusive

practices by default, and the adoption of simpler, standardized controls and APIs. Accordingly, we propose such APIs and standardization for browsers, principally to address the trade-offs browser designers face with accuracy and user burden among alternative models of settings. However, the question remains whether we can design systems without having these trade-offs and provide more meaningful controls accordingly. Thus, answering this question is the primary focus of the third and final technical chapter. Here, our work using machine learning provides strong evidence that we can maximize the benefits of controls that are based on permissions to allow or deny, such as those we propose for browsers and are already present in mobile apps. Machine learning can also minimize the drawbacks in terms of user burden, by providing recommendations for settings which are well aligned with what people prefer. These recommendations make greater use of available information than existing models that rely solely on user input to manually configure permissions. Instead, we show that it is possible to use machine learning to accurately infer many of the preferred settings of individual users. By characterizing the parameter space for a specific machine learning approach, we show that it is possible to tailor these recommendations to increase accuracy and reduce user burden.

1.5.2 Outline of Remaining Chapters

In chapter 2, we summarize the existing literature and distinguish this work from the prior art. In chapter 3, we introduce the first technical chapter, which is an exploratory study on the privacy and security settings and awareness mechanisms found in today's five most popular browsers. In chapter 4, we present our second study, which is a deeper exploration of users' preferences, mental models, and understanding of a variety of different data practices among different categories of websites. In chapter 5, we move to explore mobile apps, and improving the accuracy, management, and associated user burden of smartphone app permissions. In chapter 6, we conclude by summarizing the findings shown in each of the technical chapters, and present several avenues of future work.

Chapter 2

Background and Related Work

In this chapter, we review prior literature and distinguish our work from prior studies. We begin by highlighting prior studies on design guidelines and evaluation principles. We note methodological differences of our studies in comparison to prior work. Moving on, we review several approaches which support awareness and control, which reveal problems and trade-offs. We note recurring themes of promoting standardization, and the importance of user-centric design. Finally, acknowledging that standardization has limits, we outline prior work aimed at reducing user burden.

2.1 Design Guidelines and Evaluation Principles

Prior work has explored the parameters and considerations inherent in designing privacy and security interfaces which are intended to provide awareness and control. In particular, prior studies have sought to formalize and evaluate principles which make designs more effective, as we have also done in this work. Schaub et al. describe the challenges, requirements, and best practices for designing privacy notices [97]. Their work details a variety of design archetypes which provide a useful framework for describing the interfaces which are presented to users across a number of different domains. Cranor and Schaub also extensively detail a variety of approaches to creating usable and useful user interfaces for providing choices and obtaining consent [22]. These works serve as guidance for best practices, and also explore ways to mitigate risks as we have done in this work.

Prior work has also explored why many forms of awareness and control mechanisms have limited effectiveness, leading to unnecessary burden in various domains.

Schaub et al. explored ways to improve the design of awareness and control mechanisms encountered on websites [96]. Balebako et al. studied mobile app permissions [13], noting the importance of timing and salience on the effectiveness of what is offered. Kelley et al. and Lin et al. explored ways to evaluate mobile app permissions with respect to their manageability in the face of ever-increasing options [61, 71]. More recent studies have also explored ways to design mechanisms for awareness and control that can apply to the unique considerations surrounding the Internet of Things [26, 42].

What is apparent from the literature is that mechanisms encountered by everyday users "in the wild" stand to benefit from a better understanding of how the interfaces and settings such mechanisms offer align with people's mental models. Habib et al. showed that the ways that different websites may instantiate choices for opting out of data practices online (and deleting user data) vary considerably [50]. The literature also shows that many of the options presented by websites can be inconsistent, confusing, or difficult to use [49]. This is the same phenomenon we see again and again in our work, where we focus on how a particular set of users interacts with their browser to make use of the options they are provided. We also explore how people would ideally prefer to configure these settings across both browsers and websites.

Looking closer at the design patterns we see in our studies of browsers, the web, and mobile apps; we can recognize that many patterns have been detailed in prior literature. We recognize browsers employ only variations of the "on demand" and "decoupled" archetypes when providing awareness [97]. In other words, browsers typically notify users about privacy and security risks as they occur (rather than in advance), but only if the user actively seeks out this information by opening the privacy and security interface at that time. In the case of mobile app permissions, notifications are "just-in-time" and "blocking" which is part of what creates user burden. These notifications appear at the moment a permission is requested, and users cannot go back to their original task without making a decision to allow or deny at that moment [97, 13].

In this dissertation, we recognize the three key design characteristics of effective methods for providing awareness and control identified in prior work [97, 30, 29]: they should be relevant, actionable, and understandable. We refer to these guidelines for determining the effectiveness of methods to provide awareness and control in our research.

We are motivated to study how the design choices seen in browsers and mobile app permissions may be working to create unnecessary user burden, or otherwise limit the effectiveness of the choices they offer. We do this by employing contex-

tual interviews [56] and mixed-methods studies incorporating grounded analysis [44] which reveal qualitative insights. We also yield quantitative results from data mining large corpora of preferences, using machine learning and regression modelling techniques seen in prior literature [72]. Within the scope of these quantitative studies are four key methodological components:

1. Collecting people's privacy and security preferences [18, 69, 8].
2. Identifying the dimensions and contextual factors associated with people's privacy and security preferences that are the most expressive [73, 72].
3. Testing options which are limited to only a manageable number of security and privacy decisions [5, 3].
4. Using machine learning to further simplify decision making by offering recommendations users can review (and accept or reject) [71, 72].

2.2 Awareness and Control Mechanisms

Within the design space identified in the previous section, there have been a multitude of solutions which have sought to improve privacy and security awareness. There have also been a multitude of solutions to improve the effectiveness of controls. One common thread going back many years is work which exposes the disconnects between users' expectations and what is offered by different solutions [55, 28, 90]. Prior studies have evaluated a variety of solutions offered by browser add-ons and extensions [98, 115, 77], but we focus on what is offered in browsers' unmodified default configurations.

Prior studies chronicling historical changes to browser privacy and security awareness mechanisms show that changes over time have been subtle. Many browsers incorporate similar approaches and interfaces with common design themes [35, 66]. Many of these have questionable effectiveness [7, 66]. Some researchers have proposed specialized dashboards [19] and alternative browsers [33] which are intended to more comprehensively reveal and control privacy-relevant and security-relevant data flows. These examples, among others [18], illustrate design trade-offs favoring comprehensiveness or accuracy – but which can be too burdensome or technically involved for the average user [77, 57]. While the effectiveness of shorter and more targeted explanations of data practices can be further influenced by their framing [45], there is evidence that oversimplification worsens the likelihood of some

risks [23]. Prior work repeatedly highlights the overwhelming amount of information that needs to be processed by users to understand risks and filter out less essential information [118, 108, 58]. This leads to users facing difficulty identifying and managing risks independently of their modality [81]. However, we offer that our results speak for an entirely different problem; it is not solely the quantity or complexity of information and controls which users struggle with. This seen in prior work, such as those exploring understandable learning of privacy preferences [83], and social networking privacy preferences [40, 93]. The problem is with the inability to properly understand or relate to what is being offered. In our work, we show that the information and controls which are presented to users is most useful when it aligns with their mental models.

Many studies have shown that the ad hoc approaches seen online make restricting intrusive data practices exceptionally difficult, necessitating tool support [14, 67]. However, the way that settings, browser features, as well as privacy and security tools generally are portrayed has also been shown to be misleading. This may result in unmitigated risks, such as believing that security tools like anti-virus software also prevent online data collection, or that "private browsing" mode offers comprehensive privacy and security protection [109, 2, 1]. These are all clear examples of the misalignment between users' expectations and reality, as well the misalignment with their mental models.

2.3 Standards

Standardized awareness and control mechanisms such as nutrition labels [59, 62, 101] based on experts' advice [36] hold promise as a way of simplifying privacy and security awareness, as well as making risks easier to understand. Such work is part of a broader theme which upholds the importance of user-centered design [1, 52, 43] and accompanying standards [60]. Nudges [3] and personalized notifications [54] based around a standard set of data practices, disclosure decisions, and security choices have also been shown to be effective ways of drawing attention to, mitigating, and even preventing [34] risks.

Our work also shows that standardization has value for browsers, simplifying settings by emphasizing important factors that align with users' mental models. We explore the existing privacy and security settings offered by many browsers, qualitatively evaluating their effectiveness in the context of a particular set of users. Following this initial study, we survey what average users believe would be satisfac-

tory for them to feel that they are in control. We show that the constraints offered by appropriately standardized settings would have the effect of reducing the burden of configuring settings, without compromising their ability to express what people want. This leads naturally into a discussion of mechanisms for reducing user burden seen in prior literature, which we discuss in the following section.

2.4 Mechanisms Which Reduce User Burden

One of the biggest challenges in offering effective awareness and control mechanisms is to reduce the burden that they place on users. Strict standardization is already present in mobile app permissions, yet they face many of the same challenges of balancing comprehensiveness with user burden. This is due in part to the unique and emergent challenges associated with widespread data collection [100, 69], diverse preferences [70, 119], and the explosive growth of both apps [89, 68] and their associated permissions [71, 75, 9, 117]. Moreover, this problem is further exacerbated by the need for settings to be contextually relevant [15]. This is highlighted in studies such as the study by Zhang et al. which explored people's privacy expectations and preferences with respect to video analytics technologies [125], or Lin et al. on mobile app permissions settings [71]. These studies show that people's expectations and preferences are highly contextually dependent, but settings which are capable of accommodating both diverse preferences and diverse contexts come with an increase in user burden. Prior work proposes to reduce this burden through alternative interaction designs that involve users negotiating with systems to balance competing interests [12, 11], align semantics [95], or dynamically grant permissions as the circumstances evolve [123]. Benisch et al. studied how to balance trade-offs between control and user burden, including taking into account the level of complexity and effort users are willing to tolerate in practice [18].

The most promising approaches to mitigate user burden employ machine learning. There is a large body of existing literature which shows that people's settings, attitudes and preferences can be predictable. Studies have shown that user modeling [71, 72, 93, 73, 63, 83] and machine learning can be applied to make predictions about people's preferences about video analytics [126] and location tracking [8, 70]. They can also apply generally to mobile app permissions, as we have employed in our work. These approaches ease the burden on users by either performing configuration automatically [85], or are capable of offering recommendations based on profiles [73] which limit the number of manual decisions required to configure settings [71, 72]. In our work, we show that machine learning has the capability of not only simplifying

the configuration of existing mobile app permissions, but can further ease the burden of more complex permissions models that incorporate factors which existing settings do not support.

Chapter 3

Examining Browser Privacy and Security Settings

We start our exploration of security and privacy settings in the context of browsers. In particular, this study focuses on people's awareness and understanding of the settings available in the browsers they rely on regularly. In some ways, it doesn't really matter what settings people have available if they are not even aware that these settings exist, or do not know how to properly use them. Anecdotal evidence suggests that lack of awareness with respect to browsers' privacy and security settings is a major issue, but with the exception of studies focusing on very specific settings or tools, we lack an overall understanding of how much this affects today's browsers generally – and how much this affects users' ability to use the settings to get what they want out of their browser. Moreover, it is difficult to describe to what extent some browsers may be better than others at exposing their settings to users. In this study, we explore the extent to which people understand the settings that their browsers offer and how they interpret what they do. We also gauge people's ability to make use of the settings and other user interfaces in their preferred browser to identify and mitigate several common privacy and security risks.

Today, awareness of data practices and the settings to control them are offered in web browsers through a variety of streamlined privacy and security dashboards. Most browsers offer some combination of dashboards and additional options made available elsewhere in the browser. However, it is unclear whether these designs are effective in providing awareness and control to users. Users of different browsers have a variety of assumptions about what the default settings are. As a result, they also have a variety of assumptions about the types of protections that they are offered

by default. More generally, users have a variety of assumptions about what types of behaviors are allowed by their browsers. Others have varied expectations about what their browser's settings are capable of controlling, or what practices may be associated with these capabilities. Many of these assumptions and expectations may be misaligned with reality, or may be based on incorrect interpretations. Moreover, the way that browsers set users' expectations, convey information, and offer options may be contributing to this misalignment.

Our study is intended to provide a qualitative perspective on how users perceive the protections and controls offered by their browsers, particularly contrasting between design choices that differ substantially between different browsers. Further, we identify whether the controls provided by the browsers we studied enable users to take control of the practices that they are comfortable with, and restrict those that they are uncomfortable with. In particular, we explore whether people's expectations are aligned with what browsers can offer, subject to the limitations of their design choices. These design choices include: using different terminology, varying granularity of data practices and categories of practices, and approaches to making users aware of different practices. Where possible, we assess these design choices with respect to well-known principles seen in the literature [96].

As the first technical chapter, this study focuses on awareness and understanding. With respect to these concepts, we identify where today's browsers are working well, and find some ways in which they fall short. In addition, our results illustrate that there is more to learn in terms of individual perspectives and preferences. This sets the stage for the second technical chapter where we focus on configuration of settings and the associated user burden, exploring both individual and aggregate user perspectives in more detail.

3.1 Introduction

For several years, the privacy and security literature has shown that people's preferences and risk tolerances towards different data practices online vary substantially [6, 25]. In this study we look at a diverse set of practices, namely: fingerprinting, crypto-mining, tracking related to social media or sign-in services, targeted advertising, and behavioral profiling. We focus on these common practices in this study, though there are many more which are also prevalent. Many people are uncomfortable with these practices, misunderstand them, or at the very least want to be able to take control over them. Most importantly, people set their prefer-

ences according to their individual perceptions and understanding of the practices which are occurring as they browse, and their confidence in their ability to control them [90, 127]. To provide some level of insight into online data practices generally, different browsers have implemented mechanisms which offer distinct ways for users to be made aware of (and have control over) many data practices that are commonly encountered while browsing [82]. Each browser varies in terms of how (or whether) certain practices are described, and in terms of what settings are offered to control them. Many browsers have extensive privacy and security dashboards, while others offer more limited and simplistic settings. Some browsers offer no settings or awareness mechanisms at all [47].

Practical guidance for how to provide certain types of awareness and control to users has also been discussed in the literature [97]. For example, studies show that when users are made aware of potential threats, they are more likely to make protective decisions based on their individual privacy and security preferences [90, 3, 8]. This suggests that browser users should be made aware of privacy and security risks, and most browsers offer users a way to find out about data practices on a just-in-time [96] basis as they are encountered. However, is what is offered adequately ensuring that users can get the control that they want?

In other domains (e.g., mobile apps), notifications, privacy managers, and permissions deliver important privacy and security information to users and also offer relevant controls. Between various mobile platforms such as iOS and Android, there is evidence that the differences in design approaches to these interfaces and settings can make them more or less aligned with users' expectations. They can also make them more or less burdensome, and thus more or less satisfactory overall [103]. However, this work is centered around understanding the effectiveness of interfaces (such as dashboards and settings for privacy and security) built in to web browsers, and what users seem to associate with their needs. Our results show that there is a clear gap in this association. We suggest ways to fill this gap.

Since their invention, browsers have provided increasingly large arrays of privacy and security information and settings. Initially, these settings were introduced in a somewhat haphazard fashion. A classic example is the settings offered by Internet Explorer 6, which included control over cookies, supported machine-readable privacy policies, and offered multiple security zones [67]. These interfaces likely evolved in response to standards such as P3P [27] which were emerging at the time. As these early settings evolved, many of them began to be organized ad hoc into different control panels, menus, and categories. Eventually, developers realized it would likely make sense to regroup settings into dashboards. Ostensibly, the settings offered in

dashboards are designed to be streamlined and simple.

It is hard to know what specific traits offered by the different browsers work in a way that aligns with people's mental models without systematic assessment. Unfortunately, there is a gap in the literature – examples of systematic evaluations which report the effectiveness of different browsers' approaches are missing. Yet, anecdotal evidence strongly suggests that complex settings are often ignored [67, 30], which has led to some browsers offering much more simplistic settings as an alternative. Some browsers' privacy and security dashboards offer a smaller subset of options which are presumably thought to be the most relevant to users, with more granular or advanced settings offered elsewhere. Unlike more detailed settings, most dashboards are accessible with a single click during normal browsing activities. Reorganizing the settings or offering more simplistic options does not guarantee that they are more usable. Though they may be far simpler, it is still possible that such settings are too complicated to be useful for the average user, or provide interfaces that are difficult to understand.

There are clear trade-offs that are being made between the ability to provide more comprehensive settings that may potentially offer some which are in greater alignment with what certain users want, versus omitting or simplifying these settings in favor of simplicity (at the risk of ignoring users' need for awareness and control over specifics). Some of the most popular browsers do not have dashboards which offer any settings, though some offer some limited information [82, 80, 47] – is this information enough, and is it well understood? Are users sufficiently aware of the potential for unmitigated risks? In general, we observe that some browsers are trending towards less informative, less intrusive, less interactive interfaces with limited text and fewer options. Others offer more granular settings, with more complex explanations of data practices. These trade-offs illustrate the tension which exists between browsers' attempts to protect their users from some practices, and efforts by website operators to circumvent these protections [99]. This tension may in part be responsible for some browsers giving up on attempting to block (or even allow control over) some practices, retreating from the idea that perhaps with the right controls users can decide what they are comfortable or uncomfortable with. Effectively, some browsers are surrendering their "responsibility" and falling back on the purely libertarian approach that users must take full responsibility for the mitigation of privacy and security risks. Obviously, if users are unable to effectively use these settings (or if websites start breaking), this approach is not likely to be terribly successful.

It is worth noting that for many browsers, third party add-ons and tools exist which are designed to assist users in much the same way as dashboards, or make up

for the lack of privacy and security interfaces in some browsers. We have chosen not to include these add-ons in the scope of our study. We aim to limit the variability add-ons introduce, compared to the standard out of box experience offered by browsers (which is our main focus). Moreover, many browser add-ons are directed only at technically sophisticated users of specific browsers, or are narrowly designed to combat only certain specific privacy/security risks [57, 77]. Instead, we focus on the controls built in to browsers in their default configuration, and in particular we focus on users who claim to use one of the browsers we studied as their primary day-to-day browser.

3.1.1 Research Goal

This work provides an overview of the awareness and control mechanisms offered by browsers, in terms of how a specific set of users make use of them to complete tasks relevant to the management of online data practices. Notionally, our study is intended to address the following over-arching concerns: how good are these browsers in ensuring that their users are aware of relevant privacy and security risks? Do the users have the control necessary to mitigate these risks, and are they able to effectively take advantage of the controls to do so? Through a series of interviews, we gain qualitative insights into how a set of users interacts with and reacts to what is offered by their primary browser. We enumerate our specific research questions in more detail separately in § 3.1.3.

3.1.2 Main Contributions

In this work, we make the following main contributions:

1. We explore to what extent users can make use of their primary browser to monitor and control a representative set of data practices, focusing on users of five of today's most popular browsers (Chrome, Safari, Edge, Firefox, and Brave). These practices are associated with privacy and security risks. Though problems associated with configuring privacy and security settings are well known, our exploration uncovers some new problems, and highlights some areas where the state of the art is falling short.

2. We determine and characterizes some of the advantages, disadvantages, and trade-offs users experience when interacting with their privacy and security

settings in their primary browser. There are a variety of different approaches to settings and interfaces offered by different browsers, and we explore five popular examples.

3. We remark on whether less than ideal design choices that users encounter in certain browsers could potentially be improved, perhaps by incorporating features that work well in other browsers. We also identify some features which do not yet exist, which may be helpful (or even expected) by users.

3.1.3 Research Questions

This chapter is principally about assessing awareness, understanding. We also assess the privacy and security settings offered by five of today's most popular browsers' ability to restrict data practices. Crucially, we explore these browsers' settings through the lens of a sample of their users. Thus, in this study we are concerned with the following main research questions:

RQ1 What do people understand about their browser's ability to notify them about data practices, and their ability to take control of them?

RQ2 What browser features seem to be effective (or ineffective) at communicating information about data practices and providing control?

RQ3 Do people have the ability to monitor and control data practices with their browser? Specifically, how was this ability perceived and understood?

RQ4 What improvements can/should be made to make browsers capable of giving their users more awareness and control?

3.2 Methodology

Our study employed a *contextual interview* study methodology [53], with the goal of collecting qualitative information about users' experiences [56]. Interviews can provide more flexibility than other approaches which focus on quantitative data, enabling us to dig deeper – even if interviews preclude the collection of enough data to conduct a more quantitative analysis.

During our contextual interviews, the interviewer and interviewee worked together to complete a variety of tasks through a remote screen-sharing session. Participants were asked to guide the interviewer through the steps required to detect the presence of a variety of data practices, thinking aloud and explaining their reasoning as they explore. The participants were asked to determine the default settings, describe the implications of the different settings pertaining to the data practices, and find ways to adjust the settings to control the practices. Participants were encouraged to ask questions or ask the interviewer to interact (on their behalf) with any parts of the browser they saw fit, such as exploring buttons and menus in detail. All interviews were recorded, and their transcripts analyzed separately.

We limited the scope of our study to five of today's most popular browsers: Edge, Safari, Chrome, Firefox, and Brave. The contextual interview methodology is well suited to eliciting detailed accounts of how participants experience what is provided by their browser as they interact with it (assisted by their interviewer). Our approach was also suited to revealing insights about where participants' interactions succeeded, and where they resulted in struggles or frustration. It is worth noting that the research questions of interest to this study were less focused on desires, concerns and perceptions about intrusive practices that users may have – this is addressed separately in our next technical chapter. Instead, in this study we focused on participants' perceptions and understanding of the browsers they were presented with, allowing us to evaluate how well the status quo suits their needs.

Throughout our interviews, it was important that all participants have the same uniform experience of their browser, to minimize variability that was not the result of the participants themselves. To ensure everyone was shown the similar interfaces, participants relayed instructions to their interviewer and saw the results through remote screen-sharing. This approach ensured that the interviewer had full control over the browser and prevented the participant from attempting to perform an action which is outside the controlled parameters of the contextual interview. This also helped to ensure that the browser that was presented to the participant was always consistently configured; we presented browsers that were always using the 'factory default' configuration, further minimizing variations in experiences between participants. To further maximize the ecological validity of the study, the browsers presented through the screen-sharing session were frozen at a specific version, and displayed the same website created especially for the study. This website can be seen in Figure 3.1.

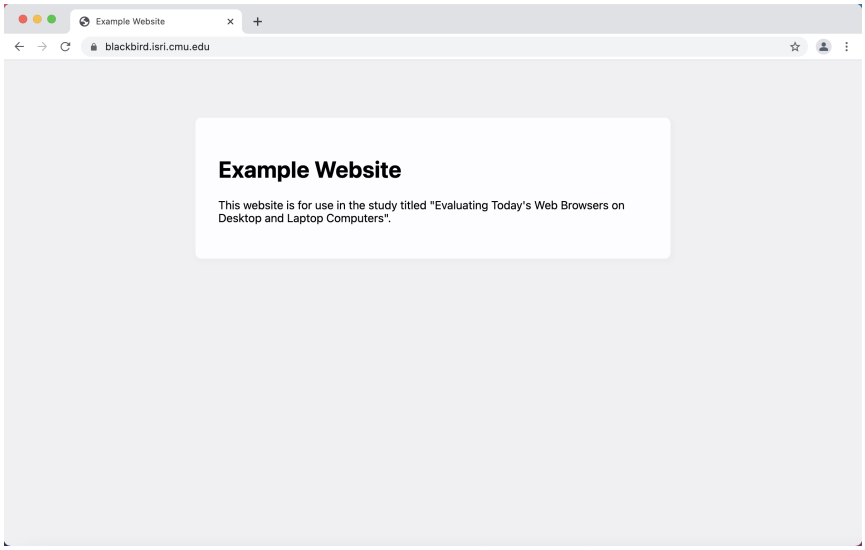


Figure 3.1: This screenshot shows the example website which was shown to participants. The browser pictured is Chrome.

3.2.1 Participant Sampling Strategy

One area where users may vary is their overall familiarity with their browser's features. Many users may be familiar with only a subset of the most popular browsers, or only one. Our approach was intended to elicit the experiences of people who were familiar with their browser, rather than people who were experiencing the browser for the first time. Accordingly, we used a *purposive sampling* [20] strategy. Rather than collecting data from a stratified sample of browser users in general, our approach was aimed at collecting data from users about their primary browser. Therefore, participants would be required to claim some level of familiarity with at least one of the browsers we studied, and we would interview them about the browser they claimed the highest level of familiarity with – their primary browser. To support our purposive sampling strategy, we performed pre-screening by distributing surveys to prospective participants who were compensated $0.20 for completing the 1-minute survey. We used pre-screening surveys to create a participant pool for each of the five browsers, ensuring that we had the same number of interviewees for each browser. During the pre-screening surveys, we asked questions which measured the participants' self-described familiarity with the five browsers, how often they claimed to use them, and how recently they examined their privacy and security settings.

We chose five of the most popular browsers in order to ensure that it would be

possible to find at least some users who claimed to have strong familiarity with at least one of the browsers. Though we focused on five popular browsers, Chrome in particular is an overwhelmingly popular browser [74] while Brave is far less popular. To collect a diverse set of experiences for each browser, we interviewed 5 different participants per browser, for a total of 25 interviews (excluding 2 pilot interviews, which were used to test our approach but did not undergo further analysis). Participants were paid $20 for their participation in both the interviews as well as brief demographic surveys. Interviews had an average duration of 50 minutes, though there was some variation as a result of different participants' exploratory activities, which we describe in our results. Demographic surveys were approximately 10 minutes in duration.

Participants who were unfamiliar with all five browsers were excluded, as were participants who did not state that they regularly browse the internet using any of the five browsers. We also excluded mobile browsers and their associated users, as we were only interested in desktop browsers. We wanted to focus our study on only one context at a time. The mobile context is substantially different to desktop browsing, and mobile browsing has limited resources which leads to different context-specific design decisions (such as accommodating touch-screens, lower resolution displays, limited battery life, and so on). Desktop browsers, in contrast, offer user interfaces which include dashboards and other more comprehensive privacy and security settings.

Following pre-screening, eligible participants were invited to sign up for interviews and asked to complete a brief demographic questionnaire prior to their interview. Each participant was assigned to be interviewed about only one browser, and their familiarity with the browser's features was later assessed by examining their interview transcripts during thematic analysis.

At a high level, each of the five browsers we studied offers a distinct approach to awareness and control. Brave (Figure 3.2) is the newest of the five browsers, initially released in 2019. It offers fine-grained information and extensive controls in both dashboards and additional settings (e.g., distinguishes fingerprinting from other practices, uniquely offers direct control over targeted advertising, uniquely offers controls specific to identity/sign-in services in the settings, but does not offer specific control over crypto-mining as Firefox uniquely does). Brave also has the smallest user base. It is advertised as a browser which offers additional privacy and security features, and settings which protect users' privacy by default. We studied Brave version 1.24 – in this version, we discovered that Brave exhibited a bug which resulted in the incorrect classification of certain data practices within the browser's

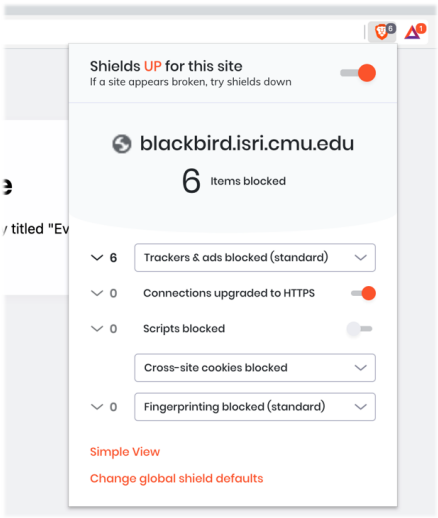


Figure 3.2: Brave has a large number of configurable options on it's privacy and security dashboard, which also incorporates numeric indicators. Brave may sometimes incorrectly classify fingerprinting under the general "trackers and ads" category, which also contains crypto-mining.

privacy and security dashboard. We detail these issues in our results.

Firefox (Figure 3.5) offers fine-grained information about data practices, and offers a dashboard but with simplified controls. Firefox has a moderately large user base, and since 2002 has generally been considered to be a popular browser. It has evolved considerably over time since the initial release – we studied version 91.

Edge (Figure 3.6) is a fairly recent browser, having been launched in 2015. At the time of our study, Edge had a relatively small user base, and had not yet incorporated many novel privacy and security features which were introduced in version 90. We studied Edge version 88, which in contrast to Brave and Firefox offers more simplistic information about privacy and security, and basic controls. However, unlike other browsers, Edge provides additional information about the origin of blocked trackers (Figure 3.7) and three separate "tracking prevention" levels (Figure 3.8).

Safari offers simplified information about blocked trackers (Figure 3.9) and very few, simplistic privacy (Figure 3.10) and security (Figure 3.11) controls. It also has a very large user base, as it is the default browser included with Apple computers. We studied Safari version 14, which includes a Privacy Report to show what Intelligent Tracking Protection has blocked, and enables full third-party cookie blocking by

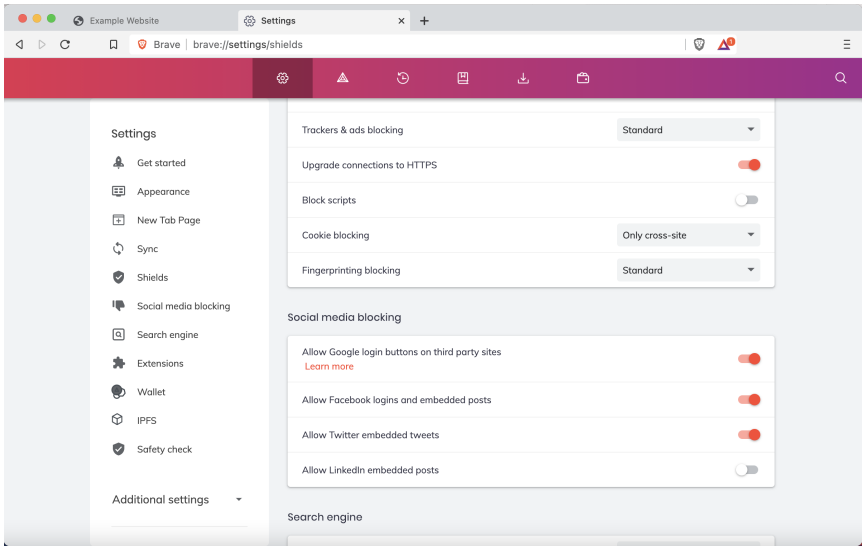


Figure 3.3: Brave also offers additional privacy and security settings within the main browser settings. Unlike all other browsers, Brave offers settings which are specific to identity/sign-in services provided by Google, Facebook, and other providers.

default.

Chrome (Figure 3.1) is by far the most popular browser we studied. However, it offers no information about data practices aside from those intended for developers (e.g., viewing the source code, web debugging tools) and no explicit ways to control the data practices we studied. Chrome does not incorporate a privacy and security dashboard which is accessible from the URL bar. However, Chrome does include general privacy and security settings (Figure 3.12) as well as interfaces for managing cookies (Figure 3.13) and HTML5 permissions (Figure 3.14).

3.2.2 Contextual Interview Scenarios

Contextual interviews are based around participants working with the interviewer to accomplish a series of tasks. First, identifying the presence of a list of practices, and then identifying ways to control them. The participant is always shown their preferred browser, displaying a minimalist example website with a small portion of text and no interactive content. Crucially, since there is a known relationship between people's perceptions of privacy and security with the websites they browse [104], the website we show is designed to minimally influence our participants. We focused the

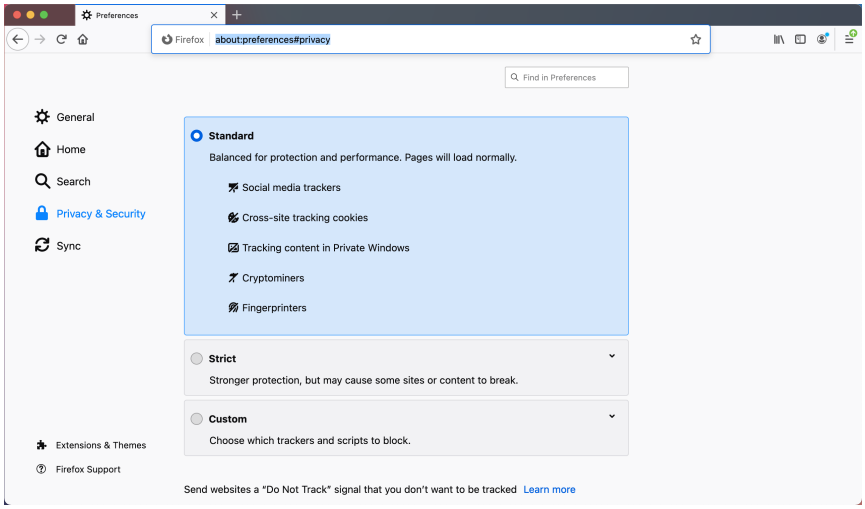


Figure 3.4: Firefox allows users to restrict many individual data practices, based on a fairly comprehensive list. These detailed settings are offered within the main browser settings, rather than on the browser's privacy and security dashboard.

interview questions on the browsers' privacy and security features, not the website itself.

We studied fingerprinting, crypto-mining, tracking related to social media or sign-in services, targeted advertising, and behavioral profiling, all of which were embedded in the example website. Fingerprinting refers to a broad set of practices which are used to uniquely identify browsers, preventing users from remaining anonymous to websites even if they have not signed in. Crypto-mining uses scripts which run in the background to perform energy-intensive calculations that earn crypto-currency. Tracking related to social media or identity/sign-in services can determine whether a particular user is logged in to a social media or identity platform (e.g. Facebook), and collect related data. Targeted advertising refers to the practice of collecting data for the purpose of personalizing ads (even if the ads are not present on the same website). Behavioral profiling is similar, involving more complex data collection and inference to build a profile about a given user's interests, habits, and more. The definitions used for each of these practices is seen in Table A.2 and Table A.1 in the appendix. Each of these practices were selected from a broader expert-validated taxonomy introduced in prior work [104]. The example website shown to participants incorporated one instance of each of these practices.

The specific implementations of the practices we studied (and the embedded

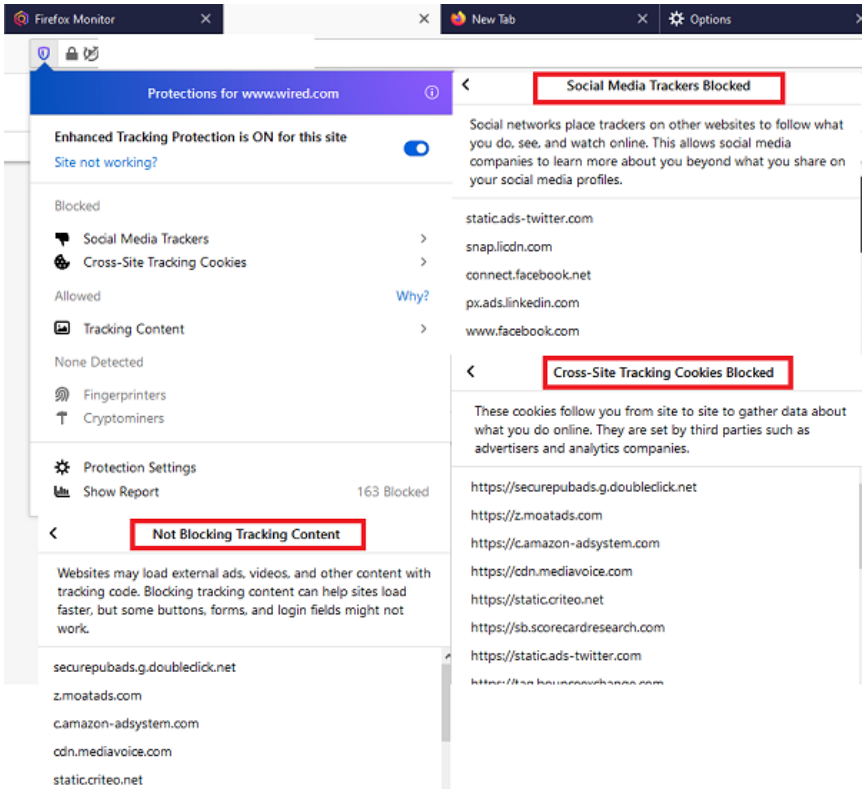


Figure 3.5: Firefox provides a privacy and security dashboard which breaks down data practices into several granular categories, each paired with additional information and descriptions (highlighted in red). The dashboard only allows control over enabling and disabling all restrictions.

source code associated with them) were chosen based on well-known examples recognized by authoritative sources [82] and used by at least one or more browser developers in their block lists or similar technologies. We chose Facebook's identity/sign-in services tracking script, which is capable of recognizing whether people are logged in to Facebook accounts and may also be associated with other data collection. For behavioral profiling, we used a script provided by Google and Doubleclick (which is a Google-owned entity) for targeted advertising. We also included a fingerprinting script used by PayPal called Simility, and the CryptoLoot crypto-mining script. Thus, our interview scenarios were created with full realism as these scripts were actually embedded in the website, incorporating the various practices as they would be seen in the wild. We also ensured that these practices were detected by browsers

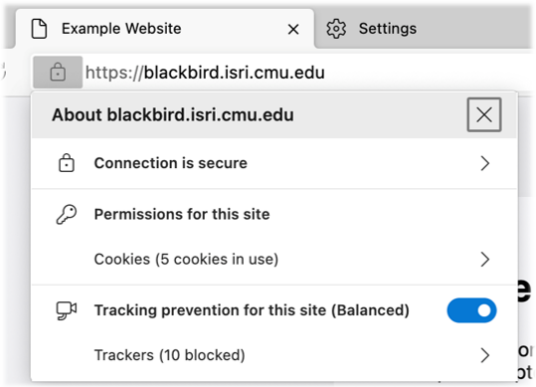


Figure 3.6: Edge offers a dashboard with information and settings about "trackers" and "cookies", hidden behind the HTTPS "lock" indicator.

which offer such capability accordingly – we note that some of these detection behaviors were not always consistent in certain browsers, however. Chrome offers no way to detect these practices and only indirect ways to control them, such as by deleting cookies (Figure 3.13), or by using developer tools to manually modify site contents. In contrast, Edge and Safari both organize practices into broad categories of "trackers" (Safari is seen in Figure 3.9) and "trackers and cookies" (Edge is seen in Figure 3.6).

Each browsers' definitions of practices and language used in settings differs, sometimes in subtle ways. For example, fingerprinting is specifically highlighted in Firefox (Figure 3.4) and Brave (Figure 3.2), but only implicitly referred to in Edge (seen in Figure 3.6, which also combines behavioral profiling, and targeted ads into one category of "trackers and cookies"). Safari uses a similar definition for "trackers" as Edge, with the difference that Safari blocks known "trackers" by default. However, the Safari documentation suggests that it may also block fingerprinting. The documentation does not refer to the precise circumstances in which this occurs and the interface does not distinguish fingerprinting from more generic "trackers". Though the definitions of data practices vary between browsers, we ensure technical consistency by using only the definitions taken from the taxonomy used in prior work [104]. The interview was structured such that each of these practices are introduced one by one, but only after the participant has had an opportunity to describe what they believe might exist (in terms of data practices). This approach ensured that it was possible to collect data about what the participant assumed absent any additional information (about the website, or about the practices in question). This also en-

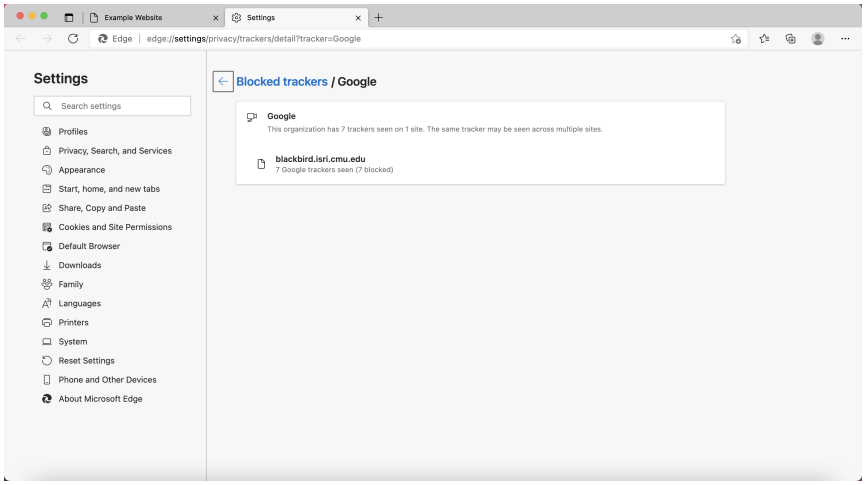


Figure 3.7: Edge offers additional information about blocked trackers when clicking on them from within the privacy and security dashboard. Unlike other browsers, Edge shows information about the organization responsible for the tracker, based on the URL.

sured that we minimally influenced the participants' a priori assumptions, such as what their browser does in terms of default restrictions of certain practices. The participant was encouraged to discover this on their own, in whatever way they saw fit, through their mediated interactions with the browser.

3.2.3 Interview Structure

In this section, we describe the overall structure of the contextual interviews. The complete script can be found in Appendix C.1.

Interview Introduction

The first section of the interview is introductory in nature. First, the participant is asked if they recognize the browser they are shown, in order to provide examples of their familiarity. Next, they are asked what browser they use most often, and whether they use any other browsers. This is intended to confirm their pre-screening and demographic survey responses. The interviewer then reveals that the browser shown is the one the participant declared that they use most often. The interviewer also mentions that the browser looks and behaves the same on all operating systems,

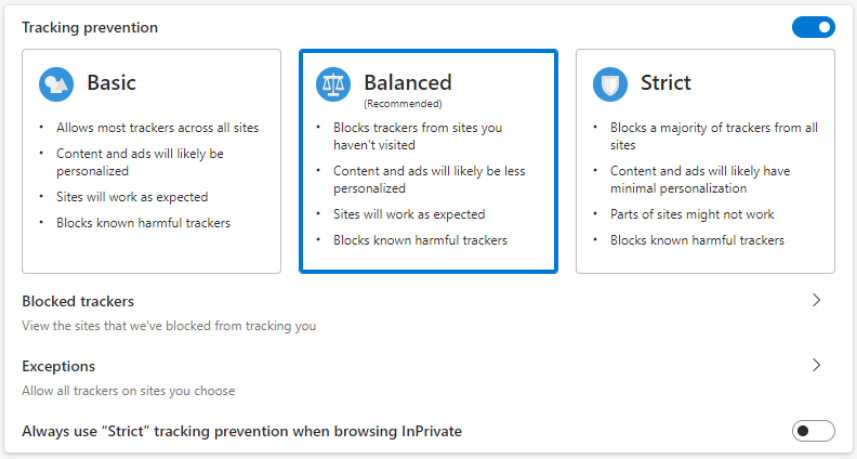


Figure 3.8: Microsoft Edge offers three levels of "tracking prevention", each controlling a variety of different data practices simultaneously. There are also additional settings which can be used to fine-tune certain browser features, such as allowing exceptions for websites, and deleting cookies.

ensuring that the participant is not confused or misled by the appearance of the title bar or menus surrounding the actual browser window. The interviewer then introduces the general concept of data practices, asks for examples of any data practices that the participant might be aware of, and what the participant looks for to find out about them. At this time, the example website is also seen in the screen-sharing session. An example of what this looks like can be seen in Figure 3.1, which features Chrome. Other browsers show the example website itself in precisely the same way, though the browser itself may differ in appearance or functionality.

Collaborative Tasks

The next section of the interview is the first collaborative task, where the interviewer defines each of the data practices, and asks the participant about their thoughts and experiences with the practice. Here, the interviewer and participant work together – the interviewer performs instructions that the participant provides. As they engage in conversation and exploration, the participant determines whether the data practices are present, and whether the practices are allowed or not by default. Typically, participants would explore the various user interfaces seen in their browser in order to determine this. For example, participants who used Brave, Firefox, or Edge might refer to their browsers' respective privacy and security dashboards (seen in

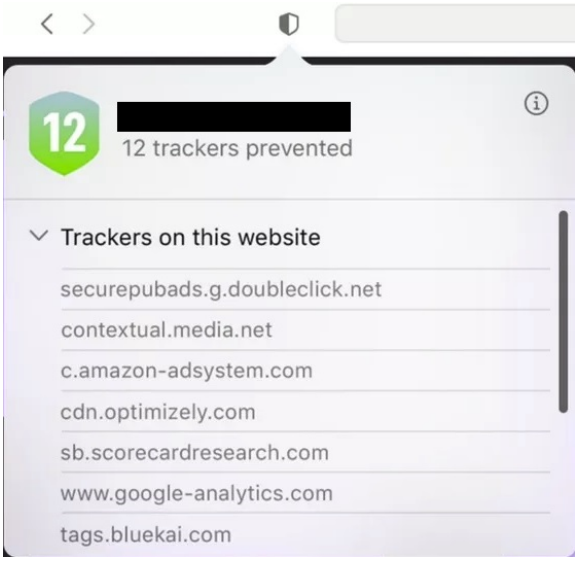


Figure 3.9: Safari offers a dashboard that displays information about blocked trackers, but offers no controls. Some limited controls are available elsewhere, in the main browser settings.

Figures 3.2, 3.5, and 3.6). The interviewer lists each of the data practices in random order, then asks the interviewee to show them how they might determine whether the data practice is present. Each practice is covered one by one, in serial. The interviewer does not move on to the next task until all practices have been covered. The second task is similar, following the interviewer revealing the presence of the practices (regardless of whether the participant was able to determine their presence) in the same randomized order as the first task. The interviewer and interviewee work together to find ways to control the practice. Again, participants would make use of their browsers' privacy and security dashboards, but they also made use of their browsers' additional settings (seen in Figures 3.12 and 3.14 for Chrome, Figures 3.7 and 3.8 for Edge, Figure 3.4 for Firefox, Figures 3.10 and 3.11 for Safari, and Figure 3.3 for Brave). After each practice is covered, the interviewer asks questions which encourage the participant to reflect on their experience, and describe whether they feel in control.

At the end of the interview, the participant is asked questions aimed at reflecting on their experience throughout the interview. In particular, they are asked questions about whether they felt their browser did a good enough job at informing them about the various practices, giving them control, and providing information in a way they

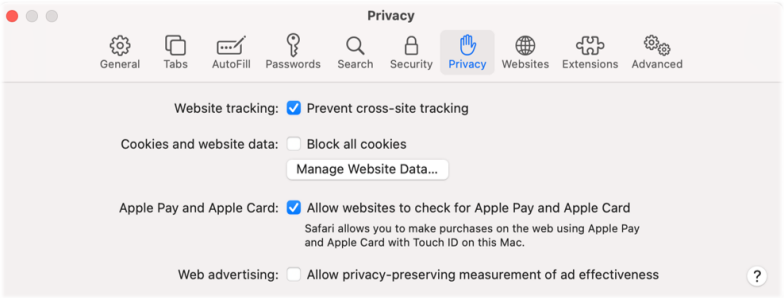


Figure 3.10: Safari offers a minimal set of additional settings related to privacy, located within the browser's main settings.

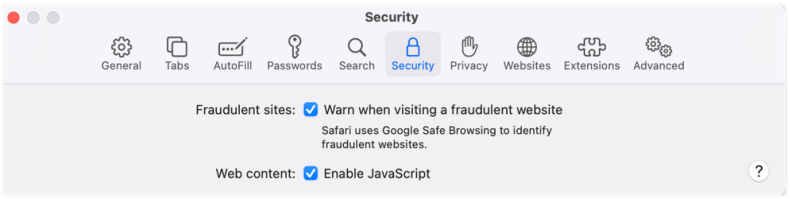


Figure 3.11: Safari also offers a minimal set of additional settings related to security, located within the browser's main settings.

could understand. Finally, the participant is asked if they have any suggestions for improving the browser (pertaining to the tasks they performed with the interviewer). Participants are given an opportunity to share any questions, comments, or concerns before the interview is terminated.

3.2.4 Analysis Approach

The initial output of our contextual interviews was a corpus of interview transcripts, which underwent thematic analysis as follows: Prior to analysis, each transcript was verified for accuracy with respect to the video recording by at least two annotators, resolving and correcting any transcription errors. To mitigate bias as much as possible, at least two annotators independently went through the transcript, annotating sections of the text by tagging them with codes. These codes were defined by an initial first-cycle coding frame which was created based on pilot interviews. Annotations were performed using an R package for computer assisted qualitative data analysis (referred to as *RQDA*) [24]. Thus, coding initially was performed with general cate-

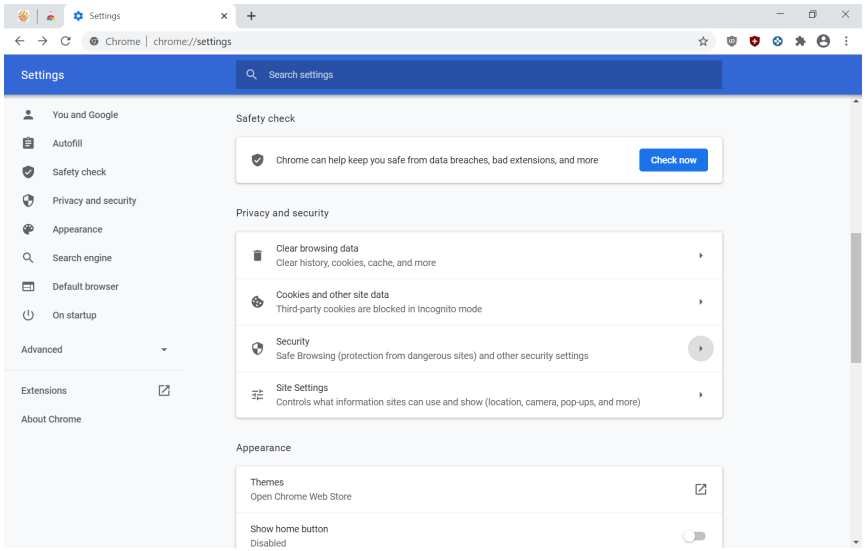


Figure 3.12: Chrome offers basic privacy and security settings in a single menu, accessible from within the main browser settings.

gories corresponding to the research questions: RQ1 focused on *understanding*, RQ2 focused on *awareness*, RQ3 focused on *control*, and RQ4 focused on *design improvements*. Later, we further refined these general categories by highlighting themes we saw upon collecting and reorganizing codes. The annotators compared and resolved any differences in annotations with one another before moving to the next transcript, and once all transcripts were annotated, the next cycle of coding began. Since our goal for this portion of our study was to gather qualitative data, we did not attempt to calculate measures of annotator reliability [78]. These additional cycles of coding were intended to capture more specific details and patterns seen among interviews. For example, we subdivided codes related to understanding by exploring themes of resignation, statements which suggest savviness (or unsavviness), highlighted examples of instances where the participant clearly was demonstrating that they understood or had made connections with the definition we provided, and found instances where the participant made reference to an add-on or a similar concept. The final list of codes and their definitions can be found in Appendix D.1.

In total, there were 4 rounds of coding. Once all interview transcripts were coded, a final analysis step was performed to summarize the annotations. Within each summary, we found examples of annotations that clearly demonstrated conclusions about the participant and their exploration of the browser. These summaries were

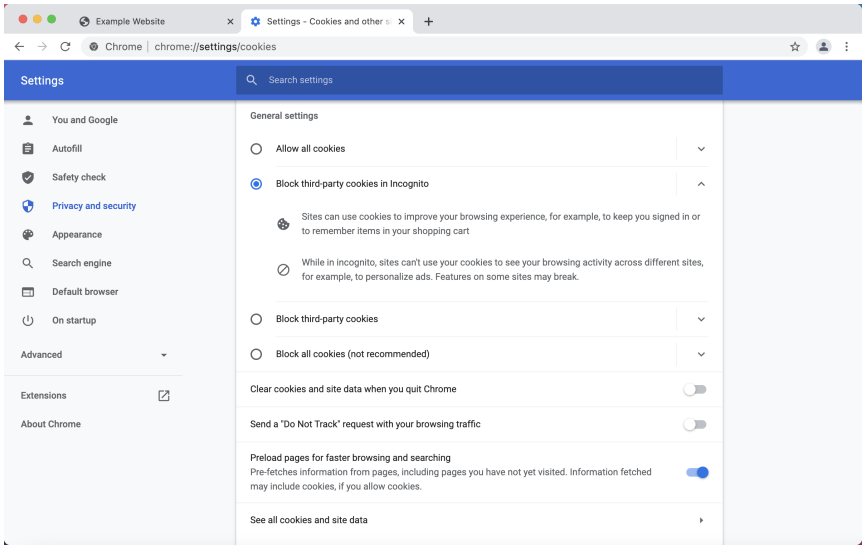


Figure 3.13: Chrome, like all other browsers we studied, offers additional settings for managing cookies.

organized as follows: First, the annotators summarized their general impressions about the participant, their perceived level of tech-savviness (on a scale of very unsavvy, unsavvy, neither savvy nor unsavvy, savvy, and very savvy), and their familiarity with their browser (on a scale of none or extremely limited, superficial, some, significant, and strong). Annotators also summarized the number of times a participant was coded as expressing frustration, or that they were struggling with the task, as well as the apparent reason for their frustration. For each practice, we also determined whether the participant successfully recognized the practice as present or not, highlighted whether the participant expressed a change in concern towards the practice as a result of the interview, determined whether their understanding of the default settings to allow or deny the practice were consistent with what their browser offers, determined whether they were able to find a way to control the practice, and concluded about whether the participant expressed that they felt in control over the practice.

3.3 Results

The results of our study are organized as follows. First, we present general findings about our participants, which includes the quantitative data we were able to

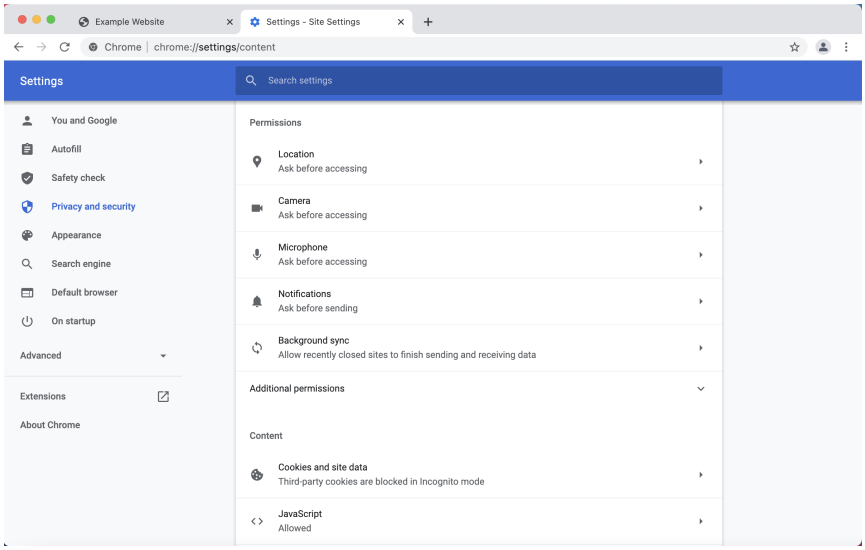


Figure 3.14: Chrome also offers additional "site settings" related to HTML5 permissions. Participants often confused these settings for privacy and security settings, even though they are organized separately.

derive from our annotations. This includes the level of tech-savviness and browser familiarity among participants, and summarizing whether they were able to detect, control, and understand the defaults for the different practices. Next, we describe the common themes we identified among our qualitative results, which are the most significant findings. These are organized into four main sections: missing or misleading information, unrealistic expectations, inaccurate mental models, and suggestions for improvement.

While the number of participants that we interviewed poses a limitation on the statistical generalizability of our findings, we were able to identify some possible quantitative trends in our results. Overall, we found that a fair number of participants demonstrated a considerable level of savviness in their interview responses. This is summarized in Figure 3.15 – participants were assessed on a 5-point Likert scale based on at least two coders independently analyzing statements in interview transcripts annotated as *SAVVY* or *UNSAVVY*, then reconciling and determining the final categorization. Generally, participants who used Brave and Firefox were clearly more savvy than the other participants. Participants who used Safari appeared to be the least savvy. Similarly, almost all participants were able to demonstrate at least a superficial level of familiarity with their browsers' features (with the exception of one

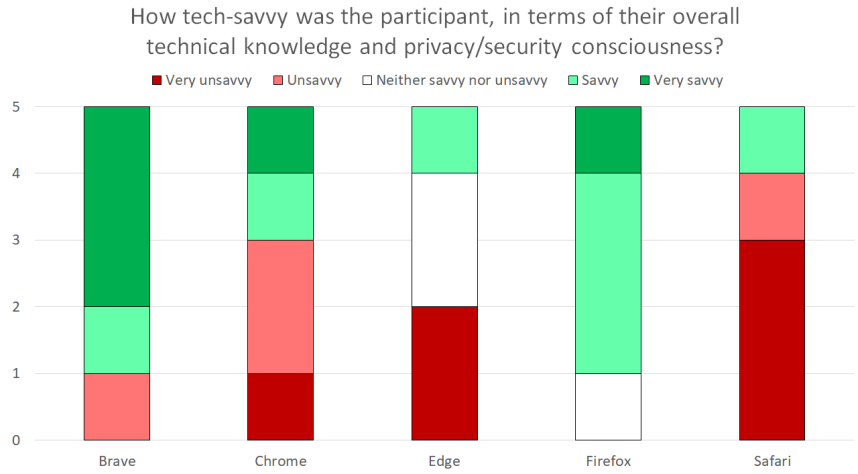


Figure 3.15: Measures of tech-savviness were determined based on agreement between at least two coders, analyzing both transcripts (citing specific codes, such as *SAVVY* or *UNSAVVY*) and summaries of annotations.

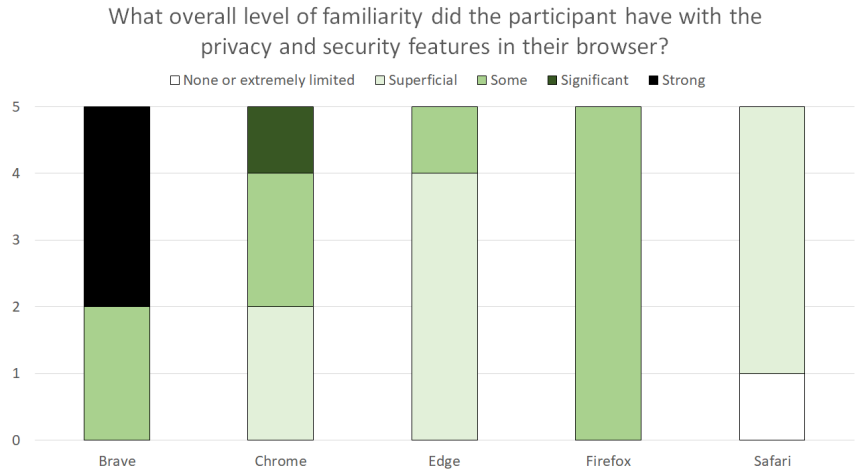


Figure 3.16: Measures of browser familiarity were determined based on agreement between at least two coders, analyzing both transcripts (citing specific codes, such as *FAMILIAR_BROWSER* or *UNFAMILIAR_BROWSER*) and summaries of annotations.

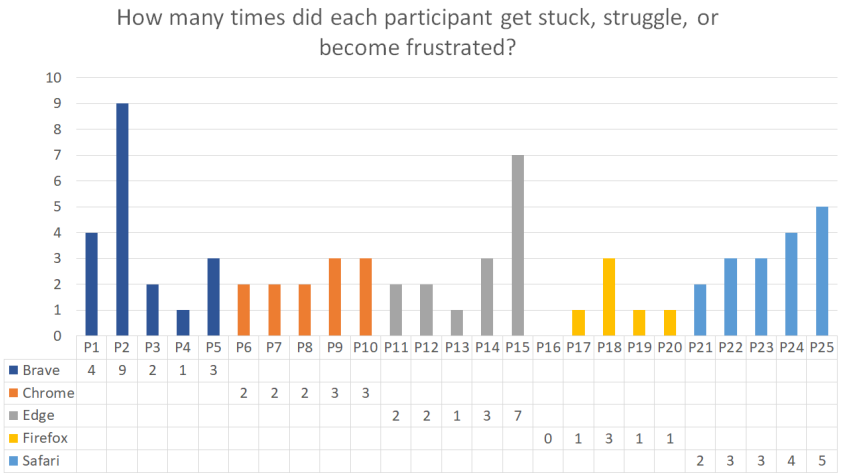


Figure 3.17: Each time it was apparent in interviews that a participant struggled, got stuck, or expressed frustration with the task, annotators tagged the section of the interview with *STRUGGLE*. These were further annotated with the apparent reasons, such as difficulty understanding the provided definitions, determining what the default settings meant, and so on (see Figure 3.18).

participant, who was a Safari user). This is summarized in Figure 3.16 – participants were ranked on a 5-point Likert scale based on at least two coders independently analyzing statements in interview transcripts annotated as *FAMILAR_BROWSER* or *UNFAMILIAR_BROWSER*, then reconciling as before with savviness. Here, we can see that similar to Figure 3.15, participants who used Brave showed the highest level of familiarity with their browser's features – perhaps due to the fact that they are highlighted with an eye-catching Brave logo that also incorporates indicators about blocked practices (similar to app notification badges on mobile). Participants who used other browsers demonstrated a fair level of familiarity – participants who used Firefox and Chrome in particular. We did not notice a trend associating savviness or familiarity with any of our other observations, but there was an apparent trend which seemed to suggest that the browser a participant used was associated with their ability to complete the tasks. This is likely explained by the interfaces offered by the browsers, rather than the participants' savviness or familiarity with the browser.

Most participants showed a similar number of instances where they struggled or became frustrated during the interview tasks. The overall number of instances per browser is shown in Figure 3.17. The trend seemed to suggest that each of the participants struggled approximately the same number of times, with the exception of

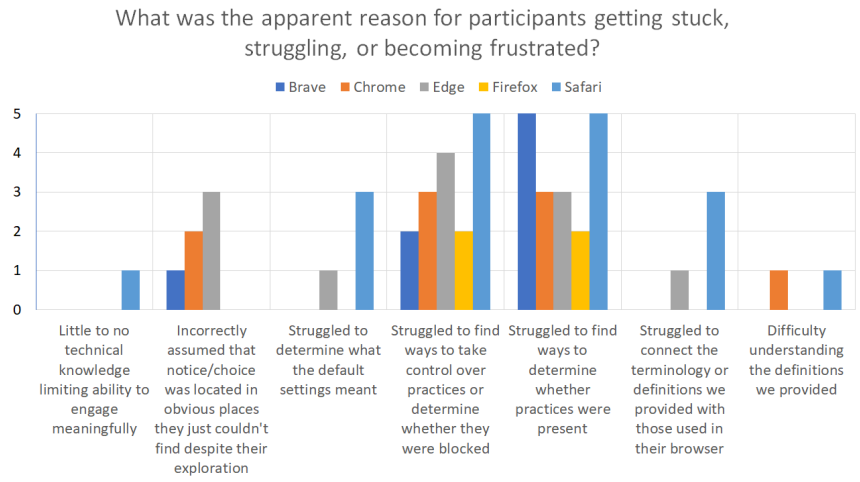


Figure 3.18: Multiple cycles of coding resulted in the *STRUGGLE* tag being subdivided into more specific categories, based on co-occurrence with other annotations related to the reasons why people struggled, got stuck, or became frustrated.

one Brave and one Edge participant with higher numbers of instances. For example, P2 struggled 9 times, becoming frustrated during their attempts to take control over a variety of practices:

> “I'm like, oh wow, do I really actually know what my settings actually are? –P2 (Brave)”

P15 struggled 7 times, repeatedly looking through the same menus for settings which they assumed were present, but that could not be found. Eventually they gave up, remarking:

> “I feel like we exhausted [everything] already, so I think I wouldn't gain anything from clicking through it [the settings] again. –P15 (Edge)”

When we examine Figure 3.18, we can see that the reasons for becoming frustrated largely center around the inability to complete the interview tasks (determining the presence of the practices, and finding ways to take control). However, we also see examples of frustration surrounding other problems, such as confusion about terminology, and assumptions about the available settings. If these are compared to the matrices summarizing participants' experiences with the individual data practices

(seen in Table E.3), it is clear that many participants became frustrated because they were unable to successfully complete interview tasks, or were not confident that they had completed the tasks successfully.

3.3.1 Missing or Misleading Information

As we examined the corpus of annotations from our interviews, we found that many participants appeared to have trouble distinguishing among different data practices seemingly because of terminology used in their browser which they found to be misleading. We also saw participants recognizing information which was misleading, or becoming frustrated as a result of missing information. In this section, we break down these examples of missing or misleading information raised by interviewees. Examples are organized by browser, beginning with trends we saw across several browsers.

Participants who used almost every browser, such as P4 (Brave), P6 (Chrome), P12 (Edge), and P25 (Safari) believed they could Google search technical terms, even though they had never heard of them before and/or their browser does not use these terms. For example, when P6 was asked about identity/sign-in services, they became confused about the definition and this lead to frustration. They were unable to determine how to proceed, with the task of determining whether the practice was present, because they couldn't understand how to apply the definition we gave them to what they were presented by their browser:

> “I usually would Google them or, you know, try to... like, my old job, you know, there were classes on this, to actually go through long explanations on them [data practices]. –P6 (Chrome)”

P4 (Brave) made similar remarks when they lost confidence about their approach to controlling fingerprinting, questioning whether they had indeed recognized the practice in the first place, based on seeing what they correctly recognized as a fingerprinting script in the page source and comparing it with the list of blocked trackers in the dashboard. When asked if there was anything they could do to increase their confidence that fingerprinting was blocked, the participant pointed to the list of URLs under the “blocked trackers” list in Brave and remarked:

> “I would Google it. That’s what I would do. I would Google the URL and see what people said about it, personally. –P4 (Brave)”

These remarks appear to suggest that there may not be enough information in many browsers for people to be able to find out what data practices are present, how to be made aware of them and their implications, what their options are for controlling them, and whether their approach was effective. This also begs the question, how would people know what to search for if they had never heard of the practice before? How would people know what to search for if their browser never mentions the practice or uses different terminology? In truth, it is unlikely that these study participants would have had any success if they were searching online for the terms used in their browsers, especially if these terms differed from the definitions we gave. Even in instances where this terminology is used, it should ideally be used consistently. For example, Firefox uses the term “profile” when defining fingerprinting, which is likely to cause confusion with behavioral profiling. Each individual browser showed examples of similar phenomena, which we detail below.

Missing/Misleading Information: Safari

Safari participants struggled with the fact that their browser offers limited information about data practices, and generically refers to “tracking” but with limited details of what this entails. P25 (Safari) had difficulty understanding fingerprinting, and believed they could Google search for information about it if they had been curious about it outside of the interview. However, they quickly recognized they would have no way of knowing what to search for since it was not a concept they were aware of prior to the interview:

> “I mean maybe if it said something about it [fingerprinting], then I would look into what it was, but, yeah, I didn’t know what it was and there’s nothing here about it. –P25 (Safari)”

In contrast, P22 (Safari) came to a different conclusion, and assumed their browser was not doing anything to block crypto-mining “because it wasn’t mentioned in the settings”. We saw other participants making similar assumptions, however, in the case of P22 this is indeed the correct conclusion (in Safari). P22 assumed this without any evidence provided by their browser.

Missing/Misleading Information: Firefox

Firefox uses the term "social media trackers" to refer to identity/sign-in services and uses the word "profile" in the definition of fingerprinting. Among our participants who used Firefox, this seemed to cause confusion. P20 (Firefox), and P16 (Firefox) were only able to connect "social media trackers" to our definition after being prompted and asking questions about alternative terminology. P19 (Firefox), and P18 (Firefox) kept searching specifically for "identity sign in services", but they could not make the connection despite being prompted repeatedly and gave up.

Firefox refers to "tracking content" as follows:

> ***Tracking Content:*** *Websites may load external ads, videos, and other content with tracking code. Blocking tracking content can help sites load faster, but some buttons, forms, and login fields might not work.*

P16 (Firefox) seemed to be especially confused by the term "tracking content" – this refers to a general category of practices in Firefox which could essentially describe any of the data practices we mentioned.

Missing/Misleading Information: Brave

Almost all of the Brave participants had some level of difficulty distinguishing among different practices, particularly the category which Brave refers to as "trackers and ads", and fingerprinting. P5 (Brave) thought that Brave lacked any useful descriptions and needed a tutorial of some kind so they could understand their options and the different categories of data practices:

> "I think I could understand it if I, like, you know, maybe looked up some YouTube videos and took some time to learn about it. But, just, like, as I am right now I think it's a little too hard [to understand the options]. –P5 (Brave)"

P3 (Brave) mentioned the need for an obvious "help section" because it was not obvious what was meant by the different blocked practices:

> “But I think, maybe they may offer like help or something like that, if there is a setting or a topic you’re not familiar with. Maybe you can go into their help section and read more about it, if you are unfamiliar with certain settings or information. –P3 (Brave)”

P1 (Brave) was confused about what could be controlled “at my end” (referring to in their browser), versus what control is offered by websites. They mentioned that Brave could do a better job of categorizing practices in order to assist with this, but that they also needed more explicit information about what was blocked and what was not:

> “I feel that Brave doesn’t explicitly say. For instance, crypto-mining or the logins, that it blocks them. This is sorted generically under, you know, the trackers that have been blocked. If you could go in, for instance [...] and let’s say that it had a subcategory, where it had, like, what crypto-mining was blocked, or login blocked, or what have you, I would feel like I would be more understanding of what is being controlled on my end. –P1 (Brave)”

P4 (Brave) mentioned the need for a “disclaimer” (i.e., a privacy policy, or similar form of disclosure) with a more overt notification that appears when practices are happening. It was unclear whether P4 was aware that many websites provide privacy disclosures already, or whether they were simply referring to the need for additional information.

Missing/Misleading Information: Edge

Edge participants seemed to have problems with the broad categories of practices in their browser, referred to in Edge as “trackers” and “cookies”. P13 (Edge) thought that cookies and trackers were confusing and overly general, mentioning that they should have clearer labels:

> “I will say, it doesn’t really seem to be doing anything to stop [behavioral profiling], just because I don’t see anything explicitly that it’s trying to do to stop it right now. But it’s not clear if it is doing it too. I would say honestly I’m not sure. It would be nice if their labels were a little bit clearer on what exactly the cookies are, and what the trackers are. –P13 (Edge)”

P11 (Edge) similarly thought it was unclear whether individual practices were allowed or blocked, because they had to make that determination themselves and could not do this based on them being all combined into one category of "trackers".

Issues with terminology in Edge also extended to the description of settings for the levels of protection the browser offers. Edge offers three levels of increasingly restrictive settings: no protections, "balanced" protections, and "strict" protections, each with a brief text description of what they offer to the user (see Figure 3.8). P15 (Edge) thought the implications of "strict" versus "balanced" modes were misleading because the descriptions were too complicated and jargon-filled, and needed to be more direct. P12 (Edge) believed that the descriptions of "strict" mode did not provide them with any way of confirming that the settings they had chosen would have the effect they wanted for any of the data practices.

Missing/Misleading Information: Chrome

Chrome participants seemed to find the collaborative tasks especially difficult, mainly because the browser offered them little actionable information. Every Chrome participant (P6, P7, P8, P9, and P10) became confused by the descriptions of the various HTML5 and similar permissions (e.g., location, background sync, seen in Figure 3.14), believing that they would need to be enabled or disabled to take control of various practices. To be clear, there is no relationship between these permissions and the data practices we covered.

Absent any clear indicators, Chrome participants struggled to find ways to reliably detect the presence of different practices (which can be seen in Table E.3). However, P9 (Chrome) and P10 (Chrome) were exceptional in that they went through the page source to determine whether practices were present or not, and were met with success.

Both P9 and P10 realized that viewing the source code was their only way to reliably determine if the practices were present, because other mechanisms were missing. It was unclear whether both participants recognized that viewing the source code would be unlikely to succeed in practice, because most websites would make it extremely difficult (if not impossible) to recognize lines of code which are specifically associated with the practices they were looking for. For example, P9 seemed to believe that it would be possible to detect targeted advertising in the source code on websites beyond the example we provided:

> *Interviewer: Do you typically scroll through the source code on web pages?*
>
> P9 (Chrome): I do not.
>
> *Interviewer: Okay, so what made you think that it might be a good idea to try it [looking at the page source] here?*
>
> P9 (Chrome): I mean, it's [targeted advertising] got to be somewhere, even if they're only being a little transparent about it.
>
> *Interviewer: Right, so do you think that by seeing it [targeted advertising] mentioned in the source code, it's probably there?*
>
> P9 (Chrome): Yeah.
>
> *Interviewer: Do you think that would also apply in cases where it's not explicitly mentioned?*
>
> P9 (Chrome): I feel like if it's there, it would be noticeable, and a targeted ad is pretty 'in your face'.

3.3.2 Unrealistic Expectations

One particular interesting category of findings was the different types of expectations different participants had about their browsers, or things that their browsers enabled them to do. Here, we focus on these expectations, and remark on the implications.

Some participants were unfamiliar with the indicators, dashboards, and settings in their browser and had preconceptions and expectations for them that did not align with reality. P24 (Safari) assumed the shield and lock represented "security features being enabled". P19 (Firefox) believed the shield indicated "whether the website is secure". P7 (Chrome), P13 (Edge) and P19 (Firefox) thought the HTTPS lock indicates whether a website is "secure" and/or "respects their privacy". This is a common misconception which is seen abundantly in the literature [2, 109]. However, this is also a misconception which has bigger implications for certain browsers. Edge, for example, combined the HTTPS lock indicator with the button that opens the privacy and security dashboard. The expectation that there is a separate way to

open the dashboard, or that this indicator is solely for HTTPS, may result in users being unaware of the existence of the dashboard. In contrast, Brave uses an icon which resembles the Brave logo, and clearly indicates the number of threats blocked which seems to potentially mitigate this problem.

Many participants were unable to find certain controls or notice mechanisms that they expected to find, because they did not exist in their browser. This is also a common, but unrealistic expectation which we see in the literature as well [104, 90, 84]. Other participants incorrectly believed the only way to find out whether a data practice was present was to block it – this was especially problematic because blocking practices does not necessarily yield information about whether it was present in the first place. P2 (Brave), P16 (Firefox), P21 (Safari), P23 (Safari), P7 (Chrome), P6 (Chrome) and P12 (Edge) all believed that the only way to find out whether a data practice was present was to block it. Some of these participants tried to do this even after they had already located the correct affordances for determining the presence of the practice, but had unrealistic expectations about what they truly offered. This mismatch in expectations may be related to the fact that these participants' explorations previously took them deep into the menus of settings, and they were unable to context-switch out of this mode of thinking. Moreover, perhaps the "allowed" or "blocked" framing seen in almost every browser contributes to this confusion. Given the expectations of our participants, it may be the case that users expect a permissions-oriented model of managing data practices, but that is not offered by any browser. Confusion with HTML5 permissions (e.g., location) seems to make this problem worse. Many participants incorrectly associated these with blocking fingerprinting, targeted ads, and behavioral profiling, in particular.

There were also several participants who incorrectly believed that data collection would not occur except in cases of self-disclosure because of their browser's built-in protections. We also found that many participants had expected that data collection associated with identity/sign-in services could not occur if the website did not offer a way of logging in, or if they did not sign in. P21 (Safari) expected that if they did not log in, that their browser would block all targeted ads and fingerprinting by default. P19 (Firefox) expected that if they did not log in, they would not be subject to any data practices we mentioned. Unfortunately, these expectations are incorrect, as the data practices can all continue unhindered without users logging in. Notably, none of the browsers' documentation seems to make reference to this fact, which may perhaps account for this expectation.

One unusual phenomenon we identified was specific to Brave participants. Some Brave participants had unrealistic expectations surrounding Basic Attention Tokens

(BAT), a feature in Brave which allows users to be paid in crypto-currency in exchange for purposefully viewing ads. They believed that BAT was a form of crypto-mining, which causes confusion, and led to the expectation that crypto-mining would always be present in Brave as long as they used this feature. This is incorrect, but understandable given our definition of crypto-mining, which could reasonably be interpreted to include BAT if users interpret that the process by which BAT is earned fits that definition. We did not attempt to correct this misconception during interviews. However, this expectation may also be attributed to the fact that Brave categorizes crypto-mining generically as a "tracker" and does not distinguish them from other data practices like Firefox does.

3.3.3 Inaccurate Mental Models

In this section, we describe our interpretation of our participants' mental models, based on their dialogue. In particular, we point to instances where the apparent beliefs of our participants might lead them to make poorly justified or irrational decisions. These mental models are predominantly associated with dialogue where we saw expressions of mistrust or lack of confidence. In particular, we saw mistrust and lack of confidence towards: the default settings offered by browsers, the effectiveness of controls, the (ostensibly good) intentions of browser developers, and the (perceived as bad) intentions of website operators.

Clear examples of mistrust were seen in many participants' interview responses. P23 (Safari) and P12 (Edge) believed that browser developers are "paid off" (i.e., suggesting that they might have incentives, financial or other) to make the settings more permissive or difficult to configure. This belief seemed to make these participants reluctant to explore the settings in any great detail, or make an effort to understand what options they had available. P6 (Chrome) believed that their browser would not block targeted advertising because it's how they (i.e., Google) make money. This insinuates that they did not have any way of controlling the practice, which is in fact true in Chrome, but may potentially have lead P6 to explore the settings less thoroughly than other participants. P24 (Safari) and P12 (Edge) believed that terms and definitions of data practices that we provided as part of the interview would purposefully not be used by their browser, in order to deliberately make the associated data practices harder to disable. These beliefs show mistrust that browser developers have good intentions. It is interesting that these participants did not express the desire to switch to another browser, given the distrust that they expressed. In contrast, P19 (Firefox) showed mistrust in website owners, suggesting that certain data practices

could never be fully blocked or detected, because they were deliberately designed by websites to evade detection and control.

In many circumstances where participants were demonstrably able to take control of certain practices, their lack of confidence in their ability to mitigate or restrict these practices caused them to become unsure that their choices would be effective. P3 (Brave) lost confidence when deciding between "strict" and "aggressive" options for a variety of practices. In particular, they seemed to be concerned about websites not working if they chose the "aggressive" options, which is understandable given the warning messages Brave provides about this potentially happening. P1 (Brave) seemed to have control over behavioral profiling, but suddenly lost confidence when contemplating the implications of their options. They worried about new trackers which might "fly underneath Brave's radar" making their choices meaningless. P15 (Edge) had the most extreme interpretation of their level of control, and believed that the settings were intended to offer only "an illusion of control", asserting that there was "no true way to leave no trace on the Internet". While there is some truth to this statement, this form of belief is also seen in the literature, resulting in inaccurate mental models [104, 109]. As people lose confidence in the effectiveness of proven privacy and security tools, though they are often effective in addressing a variety of concerns, people may choose not to use them.

3.3.4 Common Suggestions for Improvement

There were a variety of situational remarks made by participants, suggesting ways that their experience performing the collaborative tasks in the interview could be made easier. Here, we detail some of the trends in these suggestions which we saw among several participants.

Many participants expected or suggested that there should be a better way of determining the origin and implications of the different types of data practices they encountered. However, P11 (Edge) and P13 (Edge) seemed to benefit from the fact that the origin/organization associated with the domain for the trackers were displayed in the settings. While Edge only provided hints that helped them make educated guesses about their associated practices, rather than definitive information, this suggestion may generalize to other browsers which offer only the URL associated with the practice. In fact, many participants were annoyed that they only saw the URL of the blocked trackers and suggested they should be able to see what specific practices were associated with them. P1 (Brave) initially struggled to identify the presence of crypto-mining, but eventually noticed part of a URL in

the blocked trackers resembling "crypto", and correctly guessed the association with crypto-mining. P5 (Brave) and P11 (Edge) in particular struggled to determine what practice was associated with the Facebook URL (Facebook was the identity/sign-in services provider).

For Edge and Brave, many participants suggested there needs to be a more granular, more accurate classification for the different practices. P13 (Edge) was annoyed by the "tracker and cookie" catchall categories and did not find the explanation to be specific enough. P5 (Brave) realized that there was something wrong about the way that the browser was classifying different data practices, based on the observation that fingerprinting was incorrectly classified – this is an issue we discovered in Brave, where Brave often inconsistently categorizes fingerprinting more generally as "cross-site trackers", even though there is a specific indicator for fingerprinting. P5 (Brave) also found crypto-mining confusing because it is categorized generally under "trackers and ads". This strongly contrasts to Firefox which breaks out crypto-mining into a separate category of practices, which seems to alleviate confusion.

Several participants mentioned that there should be a guide or a tutorial which provides definitions, terminology, and examples of data practices. P3 (Brave) suggested a guide could be available to users when they first download Brave which would guide them through the settings. P19 (Firefox) suggested that examples of the data practices would make them more aware of what is going on (versus the general definitions provided already). P23 (Safari) had the mistaken impression that Safari already provides a guide on startup, which it does not. P12 (Edge) suggested an icon should be present on the browser which explains what is going on in terms of data practices and power usage (like Safari does with notifications about energy usage that appear during crypto-mining or on slow websites). This is worthy of additional consideration, as Edge offers a dashboard, but it may be difficult for users to recognize this as it is hidden behind the HTTPS lock icon, and it does not distinguish among data practices beyond "trackers" and "cookies". Safari, in contrast, gives indications about websites using excessive processing power or energy, which some participants were able to correctly associate with crypto-mining. Importantly, any such indicator should clearly explain *why* this might be of concern.

3.4 Summary and Key Takeaways

In this study we saw examples of terminological confusion, missing and/or misleading information provided by browsers, and unrealistic expectations. These unrealistic

expectations were expressed by participants, in terms of what they were actually capable of doing. However, the browsers we studied appeared to impose unrealistic expectations on users as well, namely about what their users understand about data practices. We also saw that some participants had developed inaccurate mental models, which seemed to hinder acceptance of (and confidence in) the mechanisms for awareness and control which were offered by their browsers. Our findings also suggest many possible improvements which seem to go beyond individual browsers, and may serve as general guidance for browser developers. Not all of the suggestions offered by users made complete sense, though most did; a few participants believed that their browsers needed to incorporate features that were ostensibly there. Even after exploring the full extent of the settings offered by their respective browsers, many participants were unable to make effective use of them, which can be clearly seen in Table E.3 in the appendix. These types of suggestions, though misguided, are understandable given the complexity of the settings which are offered. This implies that browsers ultimately rely on realistic expectations about what users understand, the amount of control that users have, and the types of the decisions they are able to make accordingly.

RQ1 focused on identifying what people understand about their browser's ability to notify them about data practices, and their ability to take control of them. In general, we found that people are vulnerable, easily misled, and even seemingly tech-savvy participants had problems recognizing and controlling data practices. Participants who used Chrome in particular struggled with the limitations of their browser. They were not able to take control of most of the practices we mentioned, and were left in a situation where if they wished to avoid or restrict those practices, they would not be able to. Participants who used Edge and Safari seemed to be confused due to imprecise and overly general terminology, and those who used Safari seemed to assume that they had far more control than they really had over the data practices we studied. This is potentially quite problematic, because this confusion may lead some people to neglect their settings altogether. Overconfidence is also likely to lead to unmitigated risks, which can cause harm. In stark contrast, participants who used Brave and Firefox seemed to have a much greater ability to understand what practices they were subjected to, but still suffered from imprecise terminology and generalized or unexpected categorizations of practices. People cannot make informed choices about how to configure their browser to allow or restrict data practices if they are unable to distinguish between the practices, or unable to take control over them to begin with.

RQ2 sought to identify features which seemed to be effective (or ineffective) at

communicating information about data practices and providing control. What we found here was that more granular categories of practices are apparently easier to understand, recognize, and take control of – we see clear examples of this with Firefox and Brave. However, vague or generalized descriptions of practices are confusing and difficult to understand, and these exist in Safari and Edge. However, we also see these vague or overly general descriptions in Brave (with "trackers and ads") and Firefox (with "tracking content") as well. From these issues, we can conclude that many browser users may need more information about practices they might encounter. In particular, this information may be necessary to prompt exploration, which may not be possible given the cognitive bias we saw where participants believed they would be able to search online for information about practices they were unaware of. These practices were not mentioned by their browser at all, or used terminology that would be unlikely to help them find information about the data practices they were concerned with. This makes users unnecessarily vulnerable to many different types of threats, as it is impossible for users to reason about taking protective measures against a risk which is entirely unknown. Our participants hadn't heard of these unknown data practices, so they had no vocabulary with which to seek new information. Combined with the fact that manipulating one's browser settings is a secondary task compared to actually browsing online, it is unreasonable for users to be expected to proactively inform themselves. Browsers need to provide them with a starting point, and ideally should offer conservative default settings. This could move browser settings toward a proactive informed consent type model for managing data practices, rather than a reactive model.

RQ3 was aimed at determining which browsers are perceived by their users to be effective at enabling them to take actions to protect themselves. Browsers with limited/simplified controls (such as Chrome, Safari, and Edge) seemed to show a pattern of our participants experiencing resignation and lack of confidence. However, sometimes simplified/limited controls (as seen in Safari) resulted in overconfidence. Some participants expressed "blind confidence" in Apple (the company which develops Safari). More granular controls, such as those seen in Brave and Firefox, seemed to improve confidence somewhat, but this confidence was easily lost as a result of confusing terminology or a lack of affirmative feedback about participants' actions. This lack of feedback, resulting in loss of confidence, was especially evident with participants who used Brave. Many participants needed reassurance that the actions they took were effective, but these expectations were unsatisfied, which resulted in resignation or confusion.

Lack of feedback by browsers about their privacy and security configuration leaves

users vulnerable, yet there are many existing approaches seen in practice and in the literature which serve as a reference to address this problem. For example, Facebook has introduced user interfaces where users can see what other people can see about them, as part of a "privacy and security check-up" interface [38]. Perhaps something similar could be developed for browser settings generally, where a user would be able to better see what practices are allowed by their browser, and what is blocked or restricted. This would provide a proactive approach, but would necessitate users to engage proactively. Just-in-time notifications, such as permissions, could be another viable alternative. If every data practice came with an associated setting of some kind, however, this approach would ensure comprehensive controls which are more in line with our participants' expectations. Most participants expected settings for controlling all of the data practices we mentioned. Such settings may benefit from some form of standardization, helping to maintain consistency between different websites and browsers.

Data practices continue to evolve, however, and our work emphasizes the need for browser providers to conduct more systematic evaluations with representative cross sections of users. This has the potential to help refine design decisions, particularly those related to taking different approaches to awareness and control. Safari takes a very simplistic approach. It takes very little time to explore the full extent of Safari's settings, which may satisfy some users, but may cause overconfidence in others (as we saw with our participants). Moreover, there was no obvious connection between our participants' expectations being satisfied and the control which is offered by more simplistic settings. Conversely, Brave and Firefox offer the most comprehensive and granular controls, but they still do not seem to offer the awareness and control our participants expected.

This study confirms that designing privacy and security controls is a challenging problem, and browsers are an especially challenging domain. The challenges highlighted by our study include: there are a variety of data practices, the underlying technologies are complex and continually evolving, and privacy and security are secondary tasks. Our study suggests that browsers struggle to adequately address these challenges. However, RQ4 in particular offers insight into the improvements that may make these browsers align better with users' expectations. What we can conclude from our findings is that browsers should provide clear and precise descriptions of data practices which are technically consistent and granular. Browsers should therefore eliminate broad categories of "trackers" and "cookies". Moreover, this general space could also benefit from some standardization, especially when it comes to terminology. Standardizing descriptions of data practices based on a stan-

dard taxonomy, such as the practices we studied in this work, could provide better consistency.

Browser developers could also provide tutorials or guides in an obvious location, which could make their users more aware of data practices and explain their options for mitigating them. Rather than leaving their users to their own devices, browsers should empower their users by helping users become more familiar with the features their browser offers. Clearer (and importantly, neutral [42]) explanations for the implications of data practices being allowed or blocked would ensure that users have an adequate understanding of the control their browser affords them. Similarly, we saw patterns in participants' responses which indicated that their browsers did not provide a clear enough association between the presence of a data practice, its origin, and the organization performing the practice. Mitigating this problem may also require browsers to have a clearer correspondence between mechanisms for being made aware of the presence of practices, and the ability to control them. Ideally, these two mechanisms should be separate, but they may still be offered in the same place if it is clear enough to users whether practices are allowed, versus simply being present (but blocked). More just-in-time notifications could be one approach to ensure that users can follow what is occurring in their browser. Another approach might be the use of nudges which educate users about the number of practices they are encountering within a given time frame. This approach has been tested in prior work, leading to changes in the mobile app permissions space which reflect these recommendations [8].

One limitation of this work was the fact that we did not incorporate add-ons into the scope of our study. This is an area which is best left to future work as we wanted to focus on browsers, rather than the variability that might have been introduced if we had included add-ons. It is possible that when people are more used to using third-party plugins, they might not intend to explore browser settings as thoroughly as if they are not. While there already exists examples of usability studies that focus on these add-ons in the literature [115, 98, 77], more work is needed to elicit people's understanding and awareness of what they can provide – particularly in contrast to what is available in browsers without these add-ons.

Another obvious limitation of our study is that we did not collect people's preferences to allow or deny the practices which we studied. In this chapter, we focused primarily on asking people how they would find out about certain data practices and how to restrict them. However, we did not probe them further about whether they would actually want to find out about them, or whether they would prefer to restrict them. Though we asked several questions intended to gauge participants'

level of concern about these practices, it was understood that when taking control of them during interviews it would be with the aim of restricting them. We also limited our exploration to the specific implementations of settings that were offered by today's browsers. In the next chapter, modelling people's preferences is in focus. We move beyond the constraints of what is offered by current browsers, and instead focus on idealized preferences. We ask people about how they feel about an even larger collection of data practices, whether people would like to be informed about them (including how often), and whether people would prefer to restrict them.

Chapter 4

Managing Online Data Practices: User Models and Perspectives

In the previous chapter, we studied how a group of users made use of the privacy and security settings seen in several of today's most popular browsers. We focused on their ability to monitor and control several data practices, given the design and constraints inherent in their primary browsers' user interfaces. What we were unable to focus on in that work was whether people would actually want to use these settings, assuming they were even able to provide them with the control they expected. Therefore, in this chapter [†] we move beyond what is offered by current browsers and explore how people would ideally prefer to control a broad set of data practices associated with privacy and security risks. By analyzing people's preferences and expectations, we aim to uncover and suggest ways to reduce mismatches between the level of control that is desired by users and what is actually provided to them. Unconstrained by the assumption that everyone has the same preference to universally restrict online data practices, there is much to uncover about people's individual preferences and how we can better accommodate them.

Reducing the mismatch between users' expectations and reality is important because of the potential for harm. The likelihood of harm is also increasing – as techniques including machine learning, fingerprinting, profiling, and other forms of automated reasoning become increasingly pervasive, users may experience them nearly constantly during everyday Internet browsing [99]. However, the application of these techniques can often provide users with improved, safer, and more relevant online

[†]This technical chapter is based on work which had been previously published in a peer-reviewed journal [104].

experiences [121, 116, 64, 5]. Therefore, users should be provided with controls that help them to restrict behaviors they are uncomfortable with in accordance with their preferences and tolerance for risk.

4.1 Introduction

In this chapter, we study a collection of controversial data practices seen online, which we refer to as "potentially intrusive practices" (PIPs). PIPs include common third-party tracking methods, as well as other types of malicious scripts that run in the browser: to collect data, monitor activity, redirect users' attention, or operate in the background to gather something of value. We focus specifically on eight categories of practices that fit this definition: identity/sign-in services, targeted advertising, behavioral profiling, fingerprinting, nag screens, session replay, crypto-mining as well as reporting and analytics. Each of these PIP can raise concerns associated with different dimensions captured by Solove's taxonomy [105], and have the potential to pose both privacy and security risks. However, whether any of these practices are viewed as overly intrusive or risky is determined by the individual's personal perspective – this user-centric aspect is the subject of interest in our work. In fact, these eight PIP may be seen as valuable by some users. Generally, websites increasingly employ profiling, reporting and analytics, and session replay to improve their products and services, increase business intelligence, and capitalize upon data brokerage [48]. Many websites use nag screens, crypto-mining, or targeted advertising to highlight new features, generate revenue from monetization, or make ads more relevant [64, 5]. Sign-in services and fingerprinting are used ostensibly for user convenience and to increase security. However, PIPs are both increasingly ubiquitous [99] and lack transparency – many users experience annoyance, frustration, fear, and feelings of insecurity or being spied upon when they find out that they had been subjected to them (especially without their consent [6, 127, 25]).

Our work focuses on the awareness and control made available by the browser itself rather than the ever-increasing array of third-party add-ons and tools. Often, add-ons require technical expertise to install and are not intended for use by the average individual [77, 57]. Outside of this tool-centric perspective, few settings are available in browsers or on websites for users to manage PIPs. Moreover, restricting PIPs using mechanisms that are not explicitly supported by websites can be fragile. Websites are constantly updated, and breakage can occur when their contents are manipulated. As a result, rather than risking breakage and losing users, many browsers' default settings are limited and there is little that can be done to restrict

or control PIP [37, 79]. Few controls are supported explicitly on websites, such as on Facebook [38]. Many others involve redirecting users through complex opt-out procedures, requiring interaction with third-parties through labyrinthine external links [14].

4.1.1 Research Goal

Managing PIPs online effectively is a significant problem that can be addressed by user-centered research; PIPs are complex, pervasive, and the extent to which users feel they have adequate awareness and control over them is unclear. However, a significant body of works has shown that users' privacy and security expectations are not currently fulfilled [90, 84, 76]. We intend to close this gap by modeling users' expectations, understanding, and preferences. Based on our findings, we suggest ways to improve the settings offered by browsers and shed light on some of the potential implications of alternative designs.

Main Contributions

In this work, we make the following main contributions:

1. We provide new insights into the understanding, preferences, and expectations of users toward PIPs beyond the tool-centric approach seen in the prior art. Our user-centered approach should enable us to expose a variety of misunderstandings, misconceptions, and assumptions about practices on different websites. For example, people believing there are no PIPs present if ads are not present.

2. We uncover ways to address participants' unfulfilled desire to be notified about PIPs, opt out of PIPs across different contexts, and determine the extent to which their preferences can extend across categories of websites.

3. We find opportunities to revisit the settings that browsers make available, and characterize their accuracy and user burden trade-offs. We also highlight new research challenges that would need to be addressed for these settings to be better aligned with users' expectations.

4.1.2 Research Questions

This chapter is intended to address the following research questions. Each question is focused on a particular aspect of designing interfaces for managing PIP.

RQ1 What are the signals that users rely on to determine whether they have encountered PIPs during browsing? (Signals)

RQ2 What interfaces or settings do users associate with allowing or restricting PIPs? (Interfaces)

RQ3 Are there PIPs that users want to control (e.g. opt-in, or opt-out), and subject to what factors? (Controls)

RQ4 What are users' preferences to be notified about PIPs on different types of websites? (Notifications)

RQ5 How well can the existing settings capture users' preferences, how often would they ideally need to be adjusted from the default, and what are the trade-offs associated with potential alternative settings? (Settings)

4.2 Methodology

Our study employed a mixed-methods approach, incorporating both qualitative ($n = 186$) and quantitative surveys ($n = 888$) which were administered to separate groups of participants. This way, we were able to gather qualitative perspectives and a large quantitative dataset of preferences from participants. Our surveys were contextualized to 8 different website categories: News and Information, Entertainment and Games, Shopping, Travel, Finance, Adult, Health and Well-being, and Social Media and Blogging. We used high level categories from Alexa [10] which we believed were broadly representative, and selected the 1st (popular) and 500th (esoteric) examples from within each category. Each survey presented one PIP to each participant.

To capture holistic categories of practices, we created a novel taxonomy elicited from experts at Mozilla. In total, we cover 8 potentially intrusive practices (PIPs): identity/sign-in services (e.g. "sign in with Google"), targeted advertising, behavioral profiling (including associated predictions and data collection about users), reporting and analytics (focusing on technical data collection), fingerprinting, nag screens (which forcefully redirect the user), session replay, and crypto-mining. Each practice

in the taxonomy is commonly encountered while browsing, and is seen by experts to have potential privacy and security problems based on Solove's taxonomy [105]. Each practice also met our overarching definition of PIPs. Crypto-mining and nag screens may involve overtly invasive acts, redirecting computing resources and attention respectively. All 8 PIPs may involve some form of surveillance, and collected data may be involved in aggregation. Data collected through these PIPs also have the potential for insecurity or harm related to dissemination. In particular, data collected during behavioral profiling, reporting and analytics, and session replay may be subject to secondary uses. Fingerprinting may be used for identification (or de-anonymization).

To maximize construct validity, we developed internal technical PIPs definitions and non-technical PIPs descriptions for surveys that were consistent and simple. We used abstract categories of practices instead of specific privacy and security threats, to avoid biases against potentially beneficial aspects of practices. We chose to create descriptions that were suitable for laypersons to easily understand so that we were not limited by how well the average user could understand the technical specifics. Using a top-down brainstorming exercise, we listed candidates for categories of practices, wrote technical descriptions, and summarized the associated risks and benefits neutrally. Our taxonomy intentionally included categories of PIPs that are not necessarily mutually exclusive, such as behavioral profiling and fingerprinting, which may often be closely associated with targeted advertising from a technical standpoint. We included these categories despite their potential overlap in order to tease out whether they were perceived differently by our participants. Each PIP is presented to separate participants, has separate descriptions, risks, and benefits, and is analyzed separately.

Neutral non-technical descriptions of PIPs intended for participants were iteratively refined. In the language used throughout the surveys, we always referred to PIPs as "web technologies" and avoided priming language, such as "intrusion", "threat" or "attack". Our descriptions were first piloted with two focus groups of non-technical employees at Mozilla. After each focus group, the text was modified based on the feedback. Clarifying details were added (e.g., fingerprinting is not referring to biometrics, giving specific examples of sign-in services) and priming language was eliminated wherever possible. Then, experts from our research team and external experts at Mozilla judged whether the corresponding PIPs opt-out scenarios were realistic and non-speculative.

Our study used both qualitative and quantitative surveys to gather data. Qualitative surveys were intended to answer descriptive questions RQ1 and RQ2. These surveys underwent grounded analysis [120] to collect and categorize general themes,

and use these findings to inform the design of a quantitative survey which could address statistical hypotheses. The grounded analysis results were used to discover trends in responses, find evidence of participants' assumptions, and determine their overall level of awareness and understanding of the surveyed concepts. Quantitative surveys measured preferences to opt out of and be notified about PIP, intended to address RQ3 and RQ4. In order to determine which surveyed factors impacted participants' expressed likelihood to opt out of PIP, our quantitative survey results underwent regression analysis. Opt-out preferences from the quantitative survey were used to create simulations that tested alternative settings models, intended to address RQ5. These simulations characterized how accurately the settings could match with individuals' expressed preferences, and how many changes to the settings (within the constraints of the alternative models) would be required to bring the settings in alignment with individual preferences.

4.2.1 Qualitative Survey

Our first survey focused on eliciting perceptions of PIPs that participants believed they had encountered, their attempts to control them, and associated experiences. Recruitment was performed on Amazon Mechanical Turk, implemented and hosted using Qualtrics, with a combined consent form and pre-screening survey. 186 participants were recruited and compensated $6 for the 20-minute average duration. Pre-screening required participants to affirm that they were over 18 years of age, resided in the US, regularly browsed the Internet, understood the consent form, and wished to participate voluntarily in research. Participants were each randomly assigned to a single PIP only.

The qualitative survey incorporated a pre-survey, consisting of a free-listing exercise about the website categories we surveyed in random order. Participants were instructed to look at their browsing history to find examples of websites that they would routinely visit if they did not immediately come to mind. This exercise focused on popular websites to evoke examples that were representative of their categories and properly contextualize their responses. All participants were required to provide two examples from at least four out of the eight website categories.

The main survey was a qualitative survey with free-text responses. Participants were asked to describe the personal risks and benefits of their assigned "web technology", and how they believed it might benefit companies who employ it. We then asked participants if and where they believed that they had encountered this "web technology" before. We also asked questions about how to protect themselves from

the potential risks; whether they had attempted to opt out, how they approached this, and whether they had succeeded.

Post-survey demographic questions asked participants about basic demographics; age range, gender, education, employment status, and city size. In addition, we administered the SA-6 questionnaire, a standard measure of security and privacy awareness [39]. Up until this point we had avoided using value-laden terms, such as "privacy" and "security", but an exception was made in our post-survey because such terms are required as part of SA-6. SA-6 was used as a proxy for measures of technical aptitude in our analysis, and participants with higher values were considered more tech-savvy.

Analysis began with removing responses where participants did not pass attention checks. Next, Glaser's grounded analysis was chosen to mitigate interpretation bias to systematically search for common themes [44, 120]. First-cycle coding identified general themes and trends. Several follow-up coding iterations were performed until saturation. Annotators were all usability, privacy, and security experts. Analysis occurred in unison, and our approach intentionally did not include measures of inter-rater reliability due to the qualitative data being collected [78]. The qualitative results were used to design a follow-up, large-scale, quantitative survey, as described below.

4.2.2 Quantitative Survey

The quantitative survey aimed to elicit the opt-out and notification preferences of browser users towards PIPs. Recruitment was performed using the same method and criteria as the initial qualitative survey, permitting only individuals who had not already participated in the previous survey. Participants were compensated $3 for the 10-minute average duration.

Each participant was randomly assigned one PIP only. The second survey began with the neutral PIP description with associated risks, and benefits. This ensured that all participants would have at least the same level of basic knowledge about their assigned PIP. Throughout the survey, we provided a link for participants to review the description, risks, and benefits. Next, each participant was presented with an example of a popular and unpopular website in each contextual category in random order. This was intended to help contextualize their responses to the category of websites. For each individual website within the category, participants were required to read about the category, when the website was established, the country it was

based in, and a detailed screenshot of the website itself. Adult websites were censored to remove explicit content.

Next, participants were required to read hypothetical scenarios describing a novel mechanism for opting out of the PIPs, and respond to questions about their preferences to use the mechanism. Scenarios were each contextualized separately to individual websites, then to whole website categories, and then to all websites broadly. We used bold fonts to emphasize important details in questions, and made it clear that participants could reverse their choice to opt out if they desired.

Our qualitative survey showed that participants expressed difficulty in identifying the presence or absence of PIPs. Therefore, at the end of our quantitative survey, we asked participants whether they would prefer to be notified about the presence (and/or absence) of the "web technology" as they browsed the various categories of websites. These questions specifically did not allude to any implementation details of the notifications – we were concerned with capturing participants' general perspectives, rather than testing a particular notification style. For each website category, participants could choose between one of "Notify me every time I visit", "Notify me only once per week", "Notify me only once per month", "Notify me only the first time I visit", or "Never notify me", corresponding to ordinal levels of notification desire.

Post-surveys evaluated the participants' SA-6 score and collected more detailed demographics; age, gender, marital status, education, employment, whether they worked and/or were educated in a STEM field, city size, when they last looked at and modified their browser privacy and security settings, their browser preference, and prior experience filling out similar surveys online. We chose to examine these demographic factors as they had been previously shown to correlate with some privacy and security preferences, particularly opt-out choices, in prior work [103]. A final question was asked about whether the participant believed they belong to a category or group of individuals who are especially at risk, due to surveillance or some form of systematic oppression.

4.2.3 Regression Analysis

In order to answer RQ3, we needed to determine which demographic factors (e.g., SA-6 score, age, gender, etc.) and/or vignette factors (e.g., website category, individual websites, popularity) impacted participants' expressed likelihood to opt out of PIPs. We performed regression analysis on our quantitative survey results to

determine this. We used regression models as a way to systematically test which factors may have influenced participants' likelihood to opt out; those which show statistically significant association with changes in opt-out likelihood across all PIPs would be suitable candidates for further testing in alternative settings, in part to answer RQ5. Likert-scale opt-out preferences were collapsed into binary categories (opt-out or opt-in) which served as the outcome variable for binomial generalized linear mixed-effects regression models. One regression was fitted for each PIP, so that they could be analyzed separately. Models were fit by maximum likelihood (Laplace Approximation) [17]. Demographic and vignette factors were modeled as fixed effects, and survey participants were each given a randomized unique identifier modeled as a random effect to account for individual variance between subjects. The 1st level of each fixed effect for ordinal factors was chosen as the intercept.

Each of our regression models underwent model selection. One by one, each factor was added to the model and the resulting candidate model was tested against a null model (with only random effects) using likelihood ratio tests. If the likelihood ratio test showed with $p < 0.05$ that the model including the added factor was statistically significant versus the null model, the added factor was included. In cases where the added factor was not significant ($p > 0.05$), the factor was excluded. As a final sanity test, we also tested for multicollinearity by reintroducing all factors (except gender and SA-6, which had already been universally excluded) into each regression and calculated the variance inflation factor. We did not find any evidence of moderate or high levels of correlation ($VIF > 5$) between any factors which had been previously included based on our likelihood ratio tests, and factors which showed high levels of correlation ($VIF > 10$) had already been excluded. We also explored whether interactions were present among factors included in the model, but no significant interactions were found.

4.2.4 Testing Alternative Settings

The corpus of preferences collected in our quantitative survey was used to perform a series of simulation experiments which test alternative models of settings for managing PIPs in the browser. These experiments characterize accuracy (i.e., how many instances in which participants' preferences coincided with what is offered by the settings), as well as user burden (i.e., the number of actions participants would need to take to adjust individual settings in order to make what is offered coincide with their preferences). The parameters of the experiment were bounded by the 16 websites collected, spanning the 8 categories of websites, across all 8 PIPs. As such, we are

simulating the effect of implementing the hypothetical settings which were introduced in our surveys in a highly constrained setting under conservative assumptions.

The experiments tested 6 different models: (1) No Toggle (closest to the current default in most browsers such as Chrome [47]) where all PIPs are allowed with no additional settings offered, then (2) No Toggle where all PIPs are denied with no additional settings. Next, (3) Default Allow and (4) Default Deny Category-level Toggles, where users can change their category-level preferences when they do not prefer the default to increase accuracy at the cost of additional burden but only based on website categories. Finally, (5) Default Allow and (6) Default Deny Individual Website Toggles, where users can change individual website preferences when they do not prefer the default to increase accuracy at the cost of additional burden across all individual websites. The experiments all assume that changing to different defaults requires one action to change each setting. One decision or changed setting amounts to one unit of user burden, accrued each time the user-preferred setting doesn't match the current default. Changing individual website/category settings requires one decision per individual website/category. Instances where users do not have consensus among categories (e.g., a user has equal numbers of allow and deny preferences within a single website category) do not result in a changed setting. Finally, we assume that we are only changing settings for one PIP at a time – a limitation imposed by our corpus being comprised of data for only one PIP per participant.

4.3 Results

Prior research focusing on specific practices and mitigation tools is consistent with our observations that most people are unaware of how to effectively identify or restrict the practices we surveyed [109, 2]. We show that people are generally unaware of the presence of intrusive practices, and don't seem to know how to detect these practices (independently of the browsers they use). In contrast to our study on today's 5 most popular browsers, in this work we studied preferences which were browser-agnostic. This included identifying controls and interfaces which people believed were associated with restricting PIP; many participants had unrealistic expectations about how their browsers and different websites gave them control over these practices.

The aim of our qualitative analysis was to categorize and organize themes in responses. We used these categories to isolate examples of signals that users rely on to determine the presence or absence of PIPs while browsing (RQ1), and the affordances

that users associate with controlling PIPs (RQ2). We received 186 responses. The demographics can be seen in Table 4.5, under Survey 1.

4.3.1 Unreliable Signals

We identified different signals that participants rely on to determine if there are being subjected to PIPs, including the presence of ads, changes in functionality (breakage), recognizing explicit notifications, and recognizing implicit notifications. Participants often told us about heuristics that they had developed to determine whether they are being subjected to (or protected from) PIPs.

Ten separate participants recognized that the presence of advertisements likely implied the presence of targeted ads and behavioral profiling, some participants referred to a lack of advertisements as a signal of not being subject to advertising-related PIPs. Another 21 instances among 19 participants showed that the presence of ads was also used as the signal for PIPs not explicitly related to ads, such as fingerprinting. One participant took note of ads that conflicted with their interests:

> “I know that I’ve succeeded [in opting out of fingerprinting] when I see ads for things that I would never eat such as meat or burgers. That’s a very simple example but it tells me some of these sites have no idea what my preferences are because I would never eat animal products of any sort. (Pt. 6f1d71b7)”

Many participants expressed confidence that they had successfully opted out because they did not see any ads:

> “Well with the ad blocker or script blocker program I know [I successfully opted out] because I don’t receive any ads at all. And with the script blocker, I’m pretty sure the website isn’t receiving any information from me based off my limited knowledge of the program. Same goes for private browsing I guess. (Pt. 6d4a0d2e)”

7 participants recognized connections between breakage and the effectiveness of their opt-out approach. We see evidence of participants using both ad-related signals and breakage in the following example:

> “I turn [targeted advertising] off or avoid it on websites that I feel are sketchy. [...] The advertisements that I was shown were different and parts of the website stopped functioning properly. (Pt. 1a64f804)”

These participants saw breakage as a signal of an effective approach. However, breakage can occur when PIPs are present. It is not definitive evidence of opting out.

99 participants speculated about the technical specifics of PIPs. We observed confusion about PIPs descriptions in focus groups and modified them accordingly. Our survey text specifically pointed out that fingerprinting did not refer to physical fingerprints or biometrics. However, 8 participants were especially confused by fingerprinting, suggesting that the data collected could include authentication tokens, security keys, biometrics, or encourage identity theft.

44 participants purported personal benefits of PIPs, expressing approval and did not mention any risks. 6 participants mentioned that with nag screens, “attention can be drawn to important things”. Though the rest found them annoying, one remarked that their interest “often outweighs or sufficiently overshadows any nag screens. (Pt. 789bf237)” We observe that many of the signals participants rely on may be unreliable.

4.3.2 Incorrect or Missing Affordances

Regarding the affordances users associate with enabling or disabling PIPs, we noted browser settings, third-party tools, extensions, and settings on websites as the most prominent examples which were mentioned in responses. 54 participants in total saw security tools as the most appropriate way to avoid PIPs as well as unrelated threats. This phenomenon was seen among technically sophisticated participants in particular. However, a significant portion of participants (15 instances in 13 responses) asserted that using their browser “safely” alone ensured their safety. These participants were unable to articulate what their approach entailed in terms of specific actions, interfaces, or behaviors. In contrast, more technically sophisticated participants would often recommend specific products. One participant detailed their usage of virtual machines to avoid data collection associated with sign-in services and malware:

> “[...] I browse using a VM (virtual machine, a cloned and contained version of my browser) when casually surfing the net or shopping. I use Shadow Defender. I can pick up all the trojans and malware I like, then with a click of a button, that “machine” is destroyed and my real computer is back in play. It’s great. (Pt. 4a171bc9)”

While using VMs or specific anti-malware security tools is useful and can potentially help mitigate malware risks, they are not effective ways to opt out of PIPs. VMs can be effective at mitigating some forms of fingerprinting, but in general, they do little to mitigate data collection associated with other PIPs. Most notably, the approach is ineffective when users sign in.

There were 12 participants who believed that to opt out of PIPs related to targeted advertising (but not specifically targeted ads), such as behavioral profiling and fingerprinting, they must block ads. These participants mentioned ad-blockers and anti-virus tools as the best way to ensure total protection from the risks associated with advertising-related PIPs. “Malvertising” (ads with embedded links to malware) [107] was also mentioned as a specific concern by 12 participants. Of these, 5 confused tools such as anti-virus with ad-blocking, perhaps due to the potential to protect from malware.

Some participants suggested that PIPs were intended to improve security. 8 participants mentioned security benefits with fingerprinting, session replay, and identity/sign-in services. We noticed that these participants seemed to emphasize security over privacy concerns, and were primarily concerned about their online accounts being hacked. One participant mentioned that they saw fingerprinting as a way to provide “greater protections against fraud (Pt. 4dc574f7)”. 3 participants perceived session replay as a beneficial feature, confusing it with history or session cookies.

62 participants in total had assumptions about having control over a setting that does not exist. Of these, 53 participants’ responses alluded to settings on websites as ways they had previously used to opt out. Settings to control PIPs are provided on websites in myriad forms, such as cookie-blocking banners and privacy dashboards, though many websites (including the examples we surveyed) do not provide any meaningful settings. 20 participants believed they had control over their surveyed PIPs when in reality there were no settings built into their browser or on the websites they mentioned. Their accounts revealed that they did not take any action to opt out because they were confused about where to find the settings – they had assumed that the settings existed. These participants expressed that it was too

difficult to configure these settings because they were unable to find them. The 33 other participants assumed that opt-out settings were available but never attempted to use them. One participant clearly knew where to look for privacy and security settings in their browser, but misinterpreted what was offered:

> “I assume the [fingerprinting] features would have their own browser setting location, just like other browser features, where you can disable and enable them. For example, I use Chrome most often, so I would expect to find these features listed under Settings > Advanced > Privacy and Security. (Pt. 2e78b57d)”

Of these 33 participants, 16 assumed their browsers would offer nonexistent settings and insisted that they should be available without specifying where.

In summary, our qualitative results offer insights into users' perceptions of PIPs. Many of the perceived affordances mentioned by participants are inadequate or nonexistent, showing a relatively high level of misconception about PIPs. We saw evidence of differences in participants' technical sophistication, confusion about the risks and benefits of PIPs, and reliance on unreliable signals. These findings informed the development of our quantitative survey.

4.3.3 Quantitative Preferences to Allow and Deny

We answered two research questions regarding participants' preferences to opt out of PIPs (RQ3) and their preferences to be notified (RQ4). 888 responses to the quantitative survey were collected in total, and we observed 92% power in post-hoc power analyses. On average, 111 responses were collected for each practice. The demographics can be seen in Table 4.5, under Survey 2.

In general, participants were opposed to most practices, wanting to opt out with little variance. Overall, participants preferred to opt out in 81% of instances on average. Table 4.1 shows participants' opt-out preferences across all surveyed practices. The outlook is somewhat different with preferences to opt out on a per website category basis as in Figure 4.1. While the majority of participants prefer to opt out in all website categories, there is some variability in preferences. For example, the preferences we collected about PIPs on finance websites is somewhat skewed towards preferring to allow. This is likely due to fingerprinting and sign-in services being seen as beneficial here. This echoes our qualitative results that some participants saw security benefits associated with these practices.

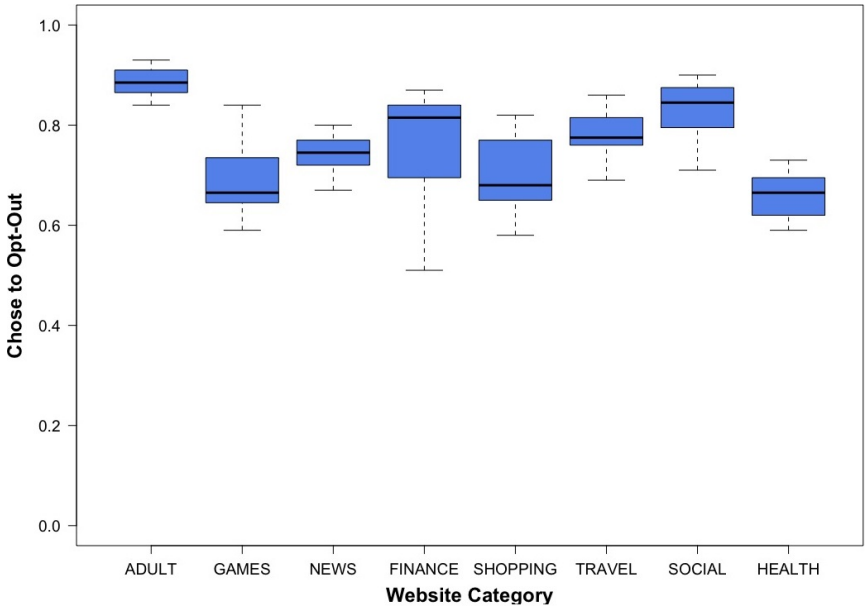


Figure 4.1: These box-plots show the aggregate opt-out preferences for participants, per website category.

There was a clear mismatch between people's expectations and reality, particularly evident because people expressed the desire to opt out of PIP but could not. However, our user-centric approach also revealed diverse views about the perceived risks and benefits of PIP. Though a majority wished to opt out in general, there were those that had slightly different preferences depending on the category of website. Still, as can be seen in Table 4.1 and Figure 4.2, the majority of participants wished to restrict and be explicitly notified about the surveyed practices.

As shown in Table 4.2, many different factors seem to play a statistically significant role in opt-out likelihood for specific PIPs. Age range, education level, STEM education, city size, marital status, employment status, STEM employment, how recently the participant looked at and changed their settings, their browser of choice, whether they had recently participated in online surveys about privacy, and whether they self-identified as being at risk of privacy or security breach were all shown to be associated with changes in opt-out preferences for at least one PIP. The detailed odds ratios and p-values from our regression models (Z-test versus the intercept) for each PIP can be found in Table E.2 in the appendix. We found that only website category was associated with changes in likelihood to opt out in all PIPs, and while other PIPs had factors which may be associated with opt-out likelihood, these

Table 4.1: This table shows the mean opt-in (allow) and opt-out (deny) preferences for all participants, averaged across each category of practice.

	Prefer Allow	**Prefer Deny**
Behavioral Profiling	20%	80%
Reporting and Analytics	20%	80%
Session Replay	18%	82%
Targeted Ads	19%	81%
Crypto-Mining	14%	86%
Identity and Sign-In Services	18%	82%
Fingerprinting	23%	77%
"Nag" Screens	17%	83%

Table 4.2: This table shows the results of our model selection, showing factors which significantly influence PIP opt-out likelihood. Included factors are labeled ●. Excluded factors are labeled ○.

	Age Range	Education Level	STEM Education	City Size	Marital Status	Employment Status	STEM Employment	Looked Settings	Changed Settings	Browser Used	Recent Surveys	At Risk	Gender	SA-6	Website Category
Behavioral Profiling	○	○	●	○	○	○	○	●	○	○	●	○	○	○	●
Reporting and Analytics	○	○	●	○	○	●	●	○	○	●	○	●	○	○	●
Session Replay	○	○	●	○	○	○	○	○	○	○	○	●	○	○	●
Targeted Ads	●	○	●	○	●	○	○	●	●	●	●	○	○	○	●
Crypto-Mining	○	●	○	●	○	○	●	○	○	○	○	●	○	○	●
Identity/Sign-in	●	○	○	○	○	○	○	○	●	○	○	○	○	○	●
Fingerprinting	○	●	○	○	○	○	○	○	○	○	○	○	○	○	●
"Nag" Screens	○	○	○	○	○	○	○	●	○	○	○	○	○	○	●

findings were inconsistent between PIPs. For this reason, website category was the only factor that we found was appropriate to use in alternative settings models. We also note that there was no evidence of interactions among factors, nor was there evidence of the participants' browser having any significant association with changes in preferences.

4.3.4 Notification Preferences

RQ4 questions which practices users prefer to be notified about, how often, and in what contexts. We observed clear preferences to be notified about the presence and absence of PIPs on most websites. We expected that participants would prefer not to be notified in most circumstances, reflecting most contemporary browsers' designs which do not notify without direct user engagement and notifications are very subtle

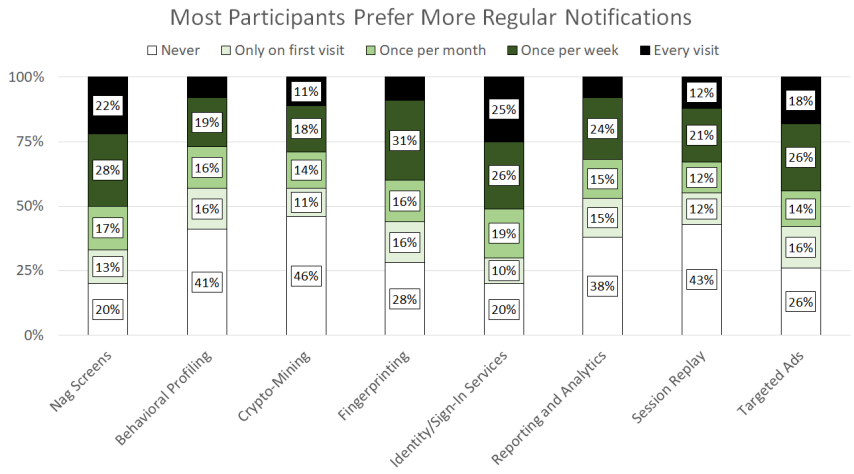


Figure 4.2: Aggregate notification preferences for the surveyed practices. In a majority of instances, participants preferred to be notified at least on the first visit, if not more frequently.

if present at all.

However, we see a clear trend towards the preference for notifications, summarized in Figure 4.2. We saw a similar trend in the responses broken down by website category, with the exception of Adult websites which were preferred to be notified about significantly more frequently. We view the desire to be notified more frequently as evidence of concern. Moreover, our qualitative results show that users are facing difficulty identifying where PIPs are present, and these results seem to provide further evidence.

Note that we deliberately chose not to study the implementation of a specific notification mechanism. Instead, we asked questions about notifications in the abstract. Some users may prefer some types of notifications over others or may find some so annoying that they would prefer to forgo them completely – our results show that there is simply an expectation that may not currently be adequately met.

4.3.5 Alternative Settings

Generally, more accurate settings are more desirable, as they fulfill the expressed preferences of users. Less burdensome settings are also more desirable, as they limit the distraction and annoyance associated with configuring these settings. Accuracy

Table 4.3: Accuracy of the various alternative setting models. No Toggle (Allow by default) reflects Chrome settings.

	No Toggle	No Toggle	Category Toggles	Category Toggles	Website Toggles	Website Toggles
Default Setting	**Allow**	**Deny**	**Allow**	**Deny**	**Allow**	**Deny**
Profiling	25.8%	74.2%	92.3%	92.3%	100.0%	100.0%
Reporting	27.9%	72.1%	91.9%	91.9%	100.0%	100.0%
Session Replay	24.5%	75.5%	92.7%	92.7%	100.0%	100.0%
Targeted Ads	24.6%	75.4%	90.0%	90.0%	100.0%	100.0%
Crypto-Mining	19.6%	80.4%	95.7%	95.7%	100.0%	100.0%
Identity Services	25.6%	74.4%	90.7%	90.7%	100.0%	100.0%
Fingerprinting	33.6%	66.4%	89.9%	89.9%	100.0%	100.0%
"Nag" Screens	24.5%	75.5%	92.1%	92.1%	100.0%	100.0%
Mean	25.8%	74.2%	91.9%	91.9%	100.0%	100.0%

Table 4.4: User burden is the average number of settings changed per user, per PIP. No Toggle is considered the current setting, with zero burden. The maximum possible burden per practice is 16 (given 2 websites in 8 categories).

	No Toggle	No Toggle	Category Toggles	Category Toggles	Website Toggles	Website Toggles
Default Setting	**Allow**	**Deny**	**Allow**	**Deny**	**Allow**	**Deny**
Profiling	0.00	0.00	5.32	1.45	11.87	4.13
Reporting	0.00	0.00	5.12	1.58	11.54	4.46
Session Replay	0.00	0.00	5.45	1.37	12.08	3.92
Targeted Ads	0.00	0.00	5.23	1.17	12.06	3.94
Crypto-Mining	0.00	0.00	6.09	1.23	12.86	3.14
Identity Services	0.00	0.00	5.21	1.30	11.91	4.09
Fingerprinting	0.00	0.00	4.50	1.88	10.63	5.37
"Nag" Screens	0.00	0.00	5.40	1.33	12.08	3.92
Total	0.00	0.00	42.32	11.31	95.03	32.97

is calculated based on the percentage of individual websites correctly aligned between the settings and expressed preferences of each individual user collected in our surveys. User burden is calculated based on the number of instances where settings must be realigned, within the constraints of the model. Both accuracy and user burden are subject to the constraints of the models we tested. Note that any time the settings are changed from the default, accuracy can increase but user burden is incurred. No Toggle has no settings and only the default applies. With Category Toggles, the default is applied but the settings can be changed per website category at the cost of additional burden. Website Toggles can be adjusted for each individual website. While the language used in our surveys referred to "opting out" of a practice, for clarity we refer to these settings as "allow" (i.e., to opt in) or "deny" (i.e., to opt out).

Table 4.3 describes the accuracy of the different settings models. Here we see that the settings which can offer the greatest accuracy are Category Toggles and Website Toggles, while No Toggle are far less accurate. Our results reveal the trade-off between accuracy and user burden inherent in the number of settings that are offered.

This relationship is evident when comparing Table 4.4 with Table 4.3. However, deny by default settings based on website categories or individual websites require fewer changes for users to achieve their preferred settings compared to allow by default – between 75% and 65% fewer actions required on average. This is consistent with our findings that users broadly prefer to opt out (Table 4.1), and website categories are a significant factor in opt-out likelihood (shown in Table 4.2).

In our simulation, the upper bound of user burden is limited by the number of surveyed website categories (8 total) and websites (16 total, 2 per category). Therefore, more choices are possible when settings are offered based on individual websites rather than categories. We found that both Category and Website Toggles spanned the entire range of possible choices for individual users, requiring zero changes in the best case and 16 in the worst case (Table 4.4). Website Toggles offer the best accuracy but are more burdensome even when the setting is allowed. In contrast, Category Toggles can provide very high accuracy with minimal user burden. The optimal trade-off will depend on the specific user, but we speculate that the middle ground would be appropriate for most, suggesting that Category Toggles would be best.

Overall, our results argue for contextually-aware settings which can distinguish among categories where certain practices are seen as permissible, proactively notify users about their presence, and otherwise restrict intrusive practices by default. Standardizing these settings in the browser rather than on websites would have the advantage of providing a uniform interface to support awareness and control, which would be easier to configure in one place. This would have the effect of eliminating the need for users to reconfigure their settings across individual websites as they browse.

4.4 Summary and Key Takeaways

This technical chapter sheds light on the signals users rely on when determining whether they are encountering or avoiding intrusive practices while browsing. We also uncover perspectives on the existing settings that people believe are associated with managing these practices and survey how they would prefer to configure their browser. This research was performed using a mixed-methods study consisting of qualitative and quantitative surveys. Confirming prior work, we find that people have very little insight into the practices they encounter during browsing and have unrealistic expectations about controlling the types of intrusive practices they do en-

Table 4.5: Breakdown of self-reported demographics from our surveys. Note that the quantitative Survey 2 had additional demographics collected, which were not collected during the qualitative Survey 1.

		Survey 1	Survey 2
Total Responses		223	1069
Rejections	Survey Abuse	24	48
	Poor Quality	13	0
	Rejection Rate	17%	4%
Gender	Male	65%	57%
	Female	35%	42%
	Other	0%	1%
Age Range	18 to 24	9%	6%
	25 to 44	69%	69%
	45 to 64	19%	21%
	≥ 65	3%	4%
Education Level	< High School	<1%	<1%
	High School	14%	11%
	Some College	14%	18%
	2-year Associates	11%	13%
	4-year Bachelor's	48%	45%
	Advanced Degree	12%	13%
City Size	Rural Area	10%	12%
	Town or Suburb	41%	37%
	City	31%	32%
	Large City	18%	19%
Marital Status	Never Married	51%	42%
	Married	37%	47%
	Divorced	8%	6%
	Other	4%	5%
SA-6 Score [39]	Mean	3.7	3.8
	Median	3.8	3.8
	Std.	0.86	0.83
Education Field	STEM	-	41%
	Non-STEM	-	59%
Employment Field	STEM	-	50%
	Non-STEM	-	50%
Preferred Browser	Chrome	-	80%
	Firefox	-	13%
	Safari	-	3%
	Edge	-	1%
	IE	-	1%
	Other	-	2%
Looked at Settings	This year	-	24%
	This month	-	42%
	This week	-	31%
	Never	-	3%
Changed Settings	This year	-	33%
	This month	-	43%
	This week	-	18%
	Never	-	6%
At-Risk Group	Yes	-	19%
	No	-	81%

counter. We also find that people are unable to identify reliable ways of determining their exposure to PIPs that they would prefer to restrict, resulting in unmitigated risks. Many users struggle to confirm the effectiveness of their attempted interventions. Often, they end up making inconsistent inferences based on the presence of ads. The lack of insight associated with these inferences may also further increase the potential for unawareness and unmitigated risks when ad-blockers are used. This leaves users frustrated, confused, and unsatisfied. Over-reliance on security tools that are not typically intended to address users' perceived risks and concerns exacerbates this problem further.

Unfortunately, users also expect settings to be available on websites to control most practices. Even though they wish to express the desire to opt out, they are unable to take any appropriate actions because such settings often do not exist. Users need better ways of objectively determining their exposure to PIPs and need to have the ability to take control. It is apparent that the ad-hoc patchwork of settings provided across browsers and websites is falling short.

This chapter also identifies alternative controls which potentially capture people's preferences more accurately, given that most people want to opt out of most practices. Alternative settings that offer opt-out by default are more likely to be in line with people's expectations. We show that alternative settings would potentially be easier to configure, particularly when they incorporate factors which we show would better fit with people's mental models (e.g., website categories and intrusive practices). Importantly, this chapter establishes that people's preferences can be accurately captured by standardizing controls based on these factors. Such standards introduce the possibility to eliminate the need for users to reconfigure their settings on every website separately. These standards would need to be uniformly implemented through the use of APIs which do not currently exist. Further improvement of controls may be hampered by website operators who may be incentivized not to accommodate people's preferences. Such website operators may even attempt to coerce users into relaxing their preferences by breaking the website intentionally. However, combined with regulation, new APIs could provide the standardization that is needed to enable browsers to act as a neutral platform for centrally managing users' preferences. This approach would share similarities with the way that mobile app permissions operate. We turn our attention to mobile app permissions in the final technical chapter.

Chapter 5

Mitigating Accuracy and User Burden Trade-offs

In the previous chapter, we showed how one can consolidate many settings based around factors which align better with people's mental models. Standardization can further reduce the scores of heterogeneous settings that would be required to be reconfigured as people browse from website to website. This would have the effect of reducing the number of times people have to effectively make the same decisions again and again within similar contexts. Within these contexts, our results showed that people's preferences to manage online data practices are correlated, especially across factors which may serve as good avenues for standardization (such as data practices and website categories). However, even if browsers were to effectively consolidate and standardize broad arrays of heterogeneous settings based on these factors, there still remains the possibility that the number of settings which users are expected to configure becomes unmanageable. A clear example of this problem is also seen in mobile app permissions, which are heavily standardized yet offer overwhelming numbers of settings. The contextual factors in web browsers also bear some resemblance to contextual factors seen in mobile app permissions. Namely, mobile app permissions are based around access to sensitive data or APIs (similar to categories of data practices), and app categories (similar to website categories). Therefore, rather than attempting to study possible standards for browsers which do not yet exist, in this chapter[†] we turn our attention to mobile app permissions. Given that mobile app permissions are already widespread, this domain offers an avenue to

[†]This technical chapter is based on work which had been previously published in a peer-reviewed journal [103].

study approaches that further reduce user burden while maintaining a high level of ecological validity.

On top of the large number of settings mobile app permissions have, another problem is that permissions also appear to be insufficiently expressive. Mobile app permissions fail to capture many important factors shown to influence people's privacy and security decisions, which are highly contextual [15]. One example of missing context is the purpose for granting permissions. In this chapter, we explore whether using machine learning we could effectively take advantage of correlations between the way people feel about many of these different privacy decisions and effectively reduce the number of decisions they have to make.

The reality is, not everyone feels the same way about the collection and use of their data, hence the need to provide users with privacy and security options that enable them to control data flows. These controls should also ensure that these data flows are aligned with people's individual preferences. Regulations such as the EU General Data Protection Regulation (GDPR) mandate that users be given proper control over the collection and use of their data, such as securing informed consent [88]. People's expectations for having control combined with mandates for control and consent create a situation where simply eliminating settings is infeasible. However, the alternative of simply offering more settings is unlikely to be effective. As data continues to be collected and used in ever more diverse ways, users are also expected to make an increasingly unrealistic number of decisions about whether to allow or restrict these practices.

In general, the alignment of options to control data flows with users' expectations for control can be characterized as *accuracy*. Conversely, the amount of effort required to perform this alignment is referred to as *user burden*. A rather prominent example of the trade-off between accuracy and user burden is found in the context of mobile app privacy and security settings (referred to as *app permissions*), which allow users to control the sensitive APIs an app can access. Prompts appear when apps first request access to sensitive data categories (e.g., contacts, location, audio, calendar, camera, etc.). This is referred to as "ask on first use" permissions. Permissions determine the ability for an app to access one or more specific categories of data on the smartphone. Many apps ask users to grant access to multiple permissions. On average, Android users would have to make over a hundred decisions to configure the permission settings associated with their apps [8, 72]. It is no surprise that the vast majority of users do not take the time to configure many of these settings, even though research shows that they truly care about many of them. Indeed, many users express both surprise and discomfort when asked to take a look at what their

permission settings actually allow [69, 8, 72].

5.1 Introduction

Recent research has shown that, using machine learning techniques, it is often possible to predict many of people's preferences based on a relatively small number of factors, such as: prior permissions, decisions, or answers to permissions-related questions [83, 71, 73, 72]. This approach offers the promise of helping to reduce the number of decisions that users have to make, by possibly giving users individual recommendations on how they might want to configure their permission settings, or by possibly combining multiple closely correlated decisions for individual users. While research on how to best take advantage of these findings is still ongoing, early results involving the deployment of personalized assistants that use these models, to recommend settings to users, suggest that such an approach can make a big difference [72]. The question that no one has attempted to answer yet is: to what extent could more expressive mobile app permissions might lend themselves to the construction of preference models with greater predictive power? Furthermore, to what extent might these stronger predictive models help mitigate the greater user burden that would otherwise be associated with the configuration of more expressive privacy and security settings? Specifically, we focus on answering these questions in the context of mobile app permissions, comparing models with permission settings that take the purpose of permissions into account versus models that do not. We present quantitative results aimed at evaluating this trade-off between accuracy and user burden across a number of parameter configurations.

5.1.1 Research Goal

Our first goal was to create a large corpus of user preferences about a variety of app permissions, for a variety of Android apps. The purpose of this corpus was to perform data mining which could potentially reveal insights into common factors along which user preferences would align. These patterns are indicative of the potential for improving predictive power. Our study sampled 5964 observations of preferences toward three sensitive Android app permissions (calendar, location and contacts), with user preferences across 108 apps, from a large sample of Android users ($n = 994$) in the United States. Having analyzed this corpus to find statistically significant factors, the next goal was to determine how to use this predictive

power to improve recommendation models for preferences. Thus, we created profiles that incorporated a combination of supervised and unsupervised machine learning (agglomerative hierarchical clusters and conditional inference trees).

Main Contributions

We make the following main contributions:

1. We offer empirically derived guidance for the design of systems which employ profiling-based machine learning techniques to improve permissions management systems.

2. We empirically determined the number of questions required to successfully profile users and count the instances where additional user input is required to make strong predictions. We measure the differences in user burden and accuracy between the models which consider purpose and those which do not.

3. We demonstrate that it is possible to improve the expressiveness of mobile app permissions models, without trading off accuracy for user burden and vice versa. Models which incorporate purpose make more accurate predictions and can also reduce the overall user burden, even when compared to other similar state of the art approaches.

5.1.2 Research Questions

This chapter is intended to address the following research questions. Each of these questions surrounds an aspect of improving Android permissions through the use of machine learning.

RQ1 What is the impact of purpose (and other contextual factors) on the predictive power of machine learning models for Android permission preferences?

RQ2 What effect does this predictive power have on the accuracy of recommendations made by profiles?

RQ3 Can we make better predictions without increasing user burden?

5.2 Methodology

In this work, we administered a survey which collected participants' Android permissions preferences for a variety of apps under two conditions: one with purpose-specific permissions and another with permissions that extend across all possible purposes. We analyzed responses using logistic regression and a combination of other machine learning techniques. Our aim was to discover whether machine learning could help mitigate the trade-off between accuracy and user burden when it comes to configuring Android app permissions. We elicited the app permission preferences of Android users using a large-scale IRB approved survey with 994 participants. Participants were recruited and compensated via the Amazon Mechanical Turk platform. A consent form with screening questions was presented prior to participation and data collection. All participants were required to be Android smartphone users located in the United States, and at least 18 years old. Participants were required to affirm that they met all required criteria when signing the consent form, otherwise, they were ineligible to participate and were removed from the participant pool. Additionally, we designed attention check questions to reconfirm the answers to the screening questions elsewhere in the survey.

First, participants were asked about their preferences independent of purpose, where no purpose was expressible. We generally refer to these preferences and their associated analyses as *purpose-independent*. Next, participants were asked to reconsider their preferences under three expressed purposes: internal, advertisement, and unspecified/other purposes (detailed below). These are referred to as *purpose-specific*.

5.2.1 Survey Design

We designed our survey to collect data about participants' app permission preferences through a large number of realistic vignette scenarios. It consists of a main survey and a post-survey demographic questionnaire. The first subsection is a primer on Android permissions in layperson's terms, which we revised from the Android developers' documentation [46]. The primer gave participants of varying technical fluency basic knowledge about Android permissions. This knowledge was necessary to complete the survey. The primer also explained the three sensitive permissions we asked about in the survey (i.e. contacts, location, and calendar), and the three general categories of purposes we considered: (1) internal, which is required for the app to deliver its basic functionality; (2) advertising, including personalized advertisement, generally

collecting and analyzing data about the user; and (3) unspecified, including any other unspecified or unknown purpose.

The second subsection elicits app permission preferences. Participants answered questions about their app permission preferences towards six different Android apps randomly selected from a pool of 108 Android apps. We curated the pool by randomly selecting 54 popular apps ($> 5M$ downloads) and 54 less popular apps (between $50K$ and $5M$) across all app categories in the Play Store. Three popular apps and three less popular apps were shown to each participant in randomized order. The distribution of apps from each category roughly approximated the frequency of app categorization in the Play Store at the time of surveying. Apps were revealed along with questions about the participants' preferences to allow or deny each permission.

First, we showed a screenshot of an app from the Google Play Store, in the identical format seen on a typical Android device. To simulate a realistic app download scenario, we instructed participants to examine the screenshot as they would normally do when making the decision to download and install an app on their phone. Following the app screenshot, we asked questions about their familiarity with the app, their frequency of use, and their preferences to allow or deny the app access to the three permissions. Throughout the survey, participants could hover over information icons to see the definition of each permission as introduced in the primer. These questions served as the baseline of participants' purpose-independent preferences – no mention of purposes was made, and the specific purpose for the permission was not expressible. Last, we asked participants about their preferences to allow or deny the app access to the same permissions in three scenarios, where the three purposes described in the primers are expressed. For each app, we collected 12 binary preferences (to allow or deny) in total: 3 purpose-independent, and 9 purpose-specific.

The post-survey questionnaire asked about participants' demographics and smartphone usage behavior, including: frequency of use, number of apps installed, and number of apps used. These questions helped to determine the likely number of permission decisions a typical participant would encounter. The number of privacy-related surveys participants had previously completed was also measured. All responses were mutually exclusive categorical factors. In total, the instrument sampled 16 control factors, including: app familiarity, app usage, demographics, and smartphone usage. Traditional rankings of privacy awareness such as IUIPC were omitted from our survey, due to lack of statistical significance in prior work [72]. Additionally, we embedded attention check questions throughout the survey. We withdrew participants who failed to correctly answer 2 or more attention check questions, and their responses were automatically discarded. Participants who completed the sur-

vey were each compensated $3.00 for the 15-minute nominal survey duration. The comprehensive list of factors and their statistical significance in regression models can be seen in Table E.1 in the appendix.

5.2.2 Analysis Using Machine Learning

Logistic regression allows for a systematic and principled approach to feature selection for machine learning models. Thus, we built profiles that can make recommendations for individuals based on the features included in the model following regression analysis. Profiles can further tailor predictions about app permission preferences to representative segments of Android users, mitigating the need to ask additional questions to get personalized recommendations. Participants' responses were clustered and aggregated using individual feature vectors comprised of the fixed effects found to be significant in logistic regression across each permission and app category. The survey dataset was divided into a validation set and a training set for machine learning with a 90/10 split, using 10-crossfold validation. To test the statistical significance of all 16 control factors included in the survey, we used a matrix of binomial mixed-effects multiple regression models. Each model was fit by maximum likelihood (Laplace Approximation) [17]. We modeled the random identifier assigned to each participant and the names of the apps they were shown as random effects. The 12 outcome variables and 16 factors were modeled as fixed effects. *A priori* power calculations were performed using G*Power [41] to determine the required number of participants and error rates to achieve a statistical power of 95%. We assumed a small effect size ($f2 = 0.03$) with an error probability of $\alpha = 0.05$ ($Power = 1 - \beta = 0.95$), which required $n = 873$ to achieve noncentrality of $\lambda = 26.19$, a critical F score of $F = 1.76$, and an expected actual power of $Power = 0.950$.

Bonferroni-corrected hypothesis tests were used to determine whether any of the tested predictors were influenced by the control factors. Each regression model in the matrix was tested using χ^2 analysis of variance (ANOVA) against a random-effects-only null model. The design matrix consisted of each permission on one axis, with all fixed effects on the opposing axis. Twelve independent hypothesis tests were performed on each fixed effect, one for each of the tested predictors (permissions). Fixed effects which were shown to be statistically significant ($pr\chi^2 \leq 0.05$) were kept as features for further analysis with machine learning. Fixed effects with weak or no significance ($pr\chi^2 > 0.05$) are reported in our results, but were not included in later models – these may have some limited predictive power, but are inconclusive.

Once the design matrix was tested, the purpose-specific models were subjected

to ANOVA against the purpose-independent models. This tested the hypothesis that, given the same effects, the outcomes (expressed preferences) differ, depending on the purpose. The purpose-independent model was used as the null model. By determining that there was a statistically significant difference between the models in the design matrix, hypothesis testing confirmed if the affordances related to purpose influenced the participants' expressed preferences. If a null hypothesis was rejected, there are measurable differences in responses when that purpose is expressible. Based on the rejection of null hypotheses, we show that purpose is a significant factor in the regression model. By examining the fixed effects coefficients in each regression model, we can quantify the impact of each factor on the likelihood to allow or deny, based on the levels of each factor. These manifest as changes in regression $\boldsymbol{\beta}$-coefficients per differences in age, app category, and so on.

Next, we applied machine learning techniques to evaluate if profile-based models could improve app permission management in terms of accuracy and user burden. We generated agglomerative hierarchical clusters for similar individuals in our data set, aggregating their preferences into profiles. A *profile* is a model of either (*app category* $\times$ *permission*) recommendations or, (*app category* $\times$ *permission* $\times$ *purpose*) recommendations. All machine learning models we tested contain either: purpose-specific, or purpose-independent permissions, but not both. Once an unknown individual has been matched to a profile (referred to as *profiling*), the profile can be queried for recommendations across all permissions and app categories. Conditional inference decision trees are used to perform profiling and to evaluate the number of questions needed to profile. Profiles and the decision trees used in profiling are static models. Once trained, they do not continue to learn from profiling or queries.

One way that profiles differ from traditional classifiers and recommendation systems is that in some cases profiles cannot make a recommendation for a particular permission. This can be due to sparse data or lack of consensus. Where the clusters of individuals that make up a profile have greater than a specified threshold for consensus about a preference, the profile makes a recommendation. In our study, we tested multiple thresholds between 70% and 90% consensus. Where recommendations cannot be made (known as *null recommendations*), we default to the original prompt where the user is directly asked whether to allow or deny a permission instead. Traditional measures of classifier performance (such as precision and recall) are too limited to evaluate our techniques, since they cannot account for null recommendations. We employ two alternative measures of performance: Our measure of accuracy is the number of cases where recommendations are made that coincide with

participants' surveyed preferences divided by the total number of recommendations made. User burden, in contrast, is the measure of individual user interactions required to both perform profiling plus the number of traditional preference elicitation prompts users that would encounter in cases of null recommendations.

Profiling uses conditional inference decision trees to re-estimate the regression relationship between clusters and individual preferences. Trees are composed of unidirectionally connected decision nodes based on the most statistically significant model features. These model factors are: app categories, demographic factors, and permissions from the design matrix used in the logistic regression analysis. The permutation tests used in the tree generation are based on Strasser and Weber's method, using Bonferroni-corrected p-values [110]. Significance is the same as the original logistic regression models ($\alpha = 0.05$). The length of the tree path traversals from root to leaf nodes are used to characterize the number of questions required to profile an unknown individual from the test data set. The decision nodes in the trees are questions that must be answered by the participant which determine which profile they should be assigned to. The answers are known *a priori* from their survey responses. The leaf nodes represent a probabilistic conclusion for which profile that the individual ought to be assigned to. By counting the number of decision nodes required to arrive at a leaf node, we can directly observe the number of user interactions required to profile an individual.

Regardless of the number of recommendations a profile is queried about, profiling need only occur once per individual user. For any given individual's profile, the ability to make a recommendation does not change based on the number of queries it undergoes or the number of recommendations it makes over time. Profiles and the decision trees used in profiling are static. Therefore, with respect to user burden, no additional assumptions are required for our analysis or evaluation. Profiles can be queried for recommendations *ad infinitum*, and can be asked to make recommendations for an unlimited number of new apps without the need to profile individuals more than once. As such, the number of interactions required for profiling is always constant for any given individual, and user burden can only increase proportionally to the number of instances where no recommendation is made. Querying a profile about additional apps for a particular individual introduces opportunities to make more recommendations and possibly null recommendations, worsening burden.

5.2.3 Measuring Accuracy

Measuring accuracy is based on the proportion of correct recommendations, and is not sensitive to the number of user interactions. Profile accuracy (A), is given as $A = (C + null)/Q$ where C is the number of correct recommendations, $null$ is the number of instances where recommendations were not made, and Q is the number of queries for recommendations. Based on this formula, in an instance where no recommendations can be made, the accuracy is assumed to be 100%, as we must assume that the interactions that would take place in lieu of a recommendation always elicit user preferences accurately.

To simulate the accuracy of a profile when queried about an arbitrary number of apps, we must make an additional conservative assumption; that the expected accuracy of the profiles' recommendations for an arbitrary number of apps lies within the Bootstrap distribution of accuracy for our 6-app data set. We use the mean of this distribution for 6 apps when simulating querying profiles for 36 apps, for all values of k in our hyperparameter sweep. This is a reasonable assumption given that the profiles that are being queried in our simulation are the same static profiles that were trained and evaluated with 6 apps, subjected to additional queries.

Because of our limited assumptions our analysis, simulation, and evaluation are conservative. Our results show that profiles can help mitigate the need for additional interactions by users to elicit their preferences as more apps are installed, in many circumstances. In contrast, the current permissions model in Android always requires the maximum number of additional interactions to elicit preferences for new apps, in all circumstances.

5.3 Results

Android privacy and security permissions have advantages over browser settings because they are standard. Permissions incorporate factors which allow preferences to be expressed across categories of APIs and their associated data practices, such as location access. However, adding additional factors to the permissions model to increase accuracy also increases the amount of user burden. For example, mobile app permissions could offer the ability to moderate permissions subject to purpose, but this would multiply the configuration burden by the number of specified purposes. Our results suggest that machine learning can indeed help mitigate trade-offs between accuracy and user burden. In the context of models that take the purpose

of permissions into account, our study suggests that it is possible to get the "best of both worlds", namely doing a better job at accurately capturing people's preferences while simultaneously reducing the number of decisions they have to make. This is accomplished using machine learning to assign users to profiles and using these profiles to infer many permissions for each user.

In total, our survey gathered 1092 responses. 98 participants' responses were removed. Of those 98, 38 were removed due to withdrawal or incomplete surveys (3% withdrawal rate). 60 responses were rejected due to failure to correctly answer several attention checks (6% overall rejection rate). Rejected responses were analyzed for evidence of systematic survey abuse; the mean time for responses was approximately 13 minutes, similar to the overall expected duration of the survey based on pilots. However, the median time for rejected responses was only 8 minutes, with a standard deviation of 12.2 minutes. Among all respondents, there was a mean of 0.24 erroneous responses to attention check questions, with median 0 errors, and standard deviation of 0.94 errors per survey. When respondents did fail attention checks, they failed most of them. Based on this data, we observe that overall most participants did not fail any attention checks, and approximately 95% of respondents made no mistakes on attention checks. This seems to suggest that most participants filled out the survey in earnest, and were paying close attention.

5.3.1 Relationships Between Expressiveness and Burden

For each of the 3 purpose-independent permissions, and the 9 purpose-specific permissions, logistic regression models identified clear patterns of significance in the fixed effects factors. The final design matrix contained only the factors which were shown to be strong predictors based on strong statistical significance ($pr\chi^2 < 0.05$). These included: Familiarity with App, App Category, App Usage Frequency, Age, Education Level, Participant City Size, Marital Status, and Number of Apps Used.

Factors with marginal or weak significance were discarded. These included: Gender, Employment Status, Smartphone Usage Frequency, Smartphone Usage Duration, Android Version, and Participant's Number of Recently Completed Privacy Surveys. Gender showed weak significance across all permissions, regardless of purpose. Surprisingly, Marital Status was a very strong predictor across all permissions preferences. In particular, participants who were divorced, widowed, or never married were most similar and were more likely to deny permissions broadly. Participants who were married or separated were more likely to allow permissions, in comparison to those who were divorced. While Employment Status was generally a very

weak predictor, it was observed to have strong significance for Calendar permissions. Participants who were not working because they were retired or disabled were more likely to deny, while those who were students, paid employees, laid off, or otherwise looking for work were more likely to allow.

The variance in Participant Smartphone Usage Frequency and Participant Smartphone Usage Duration was the likely explanation for the observation that these two factors had very weak significance, with only marginal significance (if any) in many cases. This suggests that this aspect of smartphone usage behavior is not a useful predictor. However, Number of Apps Installed and Number of Apps Used appeared to be very strong predictors in almost all cases; there is a clear trend where participants were more likely to allow access to permissions if they had many apps installed, and if they reported that they used many of them. Participants with small numbers of apps installed and used were more likely to deny permissions in many cases, perhaps because participants who are more privacy conscious downloaded fewer apps.

Android Version and Participant's Number of Recently Completed Privacy Surveys had too little significance to observe any clear response trend. Participants' number of recently completed privacy-related surveys did not appear to correlate with any particular characteristic of responses, nor did Android version. Many participants self-reported outdated Android versions, including some which do not support Ask On First Use style permissions (the current app permissions model on all smartphone platforms), which suggested that they may not have had the technical knowledge to determine what version they had on their device.

The results of the final ANOVA are summarized in Table 5.1. The null hypothesis is that there are no differences in responses between the purpose-specific and purpose-independent models. The purpose-independent model is the null model, which is subjected to ANOVA versus the purpose-specific models across the three permissions. The alternative hypothesis is that the purpose-specific information has measurably different responses. It is clear based on the rejection of the null hypotheses that there are measurable differences when comparing purpose-independent to purpose-specific regression models, except in the case of Internal. One possible explanation is that participants already assumed that the app permissions which were purpose independent are already declared because they are ostensibly for Internal purposes. Regardless, rejecting the null hypothesis provides strong evidence that the participants' purpose-independent expressed preferences do not intersect with their purpose-specific expressed preferences in a significant number of instances. In other words, there is a significant difference in expressed preferences between the two types of affordances. This implies that users would clearly benefit from the ability

Table 5.1: ANOVA of regression models which cannot specify any purpose (Null) versus purpose-specific models.

	Contacts			
	Df	χ^2	χDf	$pr(>\chi^2)$
Null vs. Internal	57	≈ 0	0	≈ 1
Null vs. Advertisement*	57	1039.1	0	$\leq 2.2 \times 10^{-16}$
Null vs. Other/Unspecified*	57	1577.6	0	$\leq 2.2 \times 10^{-16}$
	Calendar			
Null vs. Internal	57	≈ 0	0	≈ 1
Null vs. Advertisement*	57	1292.1	0	$\leq 2.2 \times 10^{-16}$
Null vs. Other/Unspecified*	57	2025	0	$\leq 2.2 \times 10^{-16}$
	Location			
Null vs. Internal	57	≈ 0	0	≈ 1
Null vs. Advertisement*	57	1180.7	0	$\leq 2.2 \times 10^{-16}$
Null vs. Other/Unspecified*	57	1952.1	0	$\leq 2.2 \times 10^{-16}$

to express purpose-specific preferences.

Our results show that greater expressiveness in the settings does not have to necessarily translate into greater user burden and that machine learning can help mitigate trade-offs between user burden and accuracy. This is evident in Figure 5.3(a) and 5.3(b), which show that permissions subject to purpose are consistently less burdensome, yet are able to achieve higher accuracy.

There are outliers from the highlighted areas in Figure 5.2(b). In particular, it is worth noting $k = 2$ and $k = 4$ are outliers in both models, appearing to suggest that the best accuracy/user burden trade-off might occur with very small numbers of profiles. However, using such a small number of profiles is impractical for the same reasons identified in prior work, which found that a single set of defaults or a very small number of profiles are too internally heterogeneous to generalize well [71, 72]. With 36 apps, small values of k prove far worse than they appeared in Figure 5.2(b) with only six apps, because they cannot make recommendations in a much higher percentage of instances. With low values of k, it takes a very small number of questions to profile individuals; the highest numbers of additional interactions are required here, as these profiles seldom make any recommendations. As such, they trade higher accuracy for many more additional user interactions.

Profiles with small values of k are far more timid about making recommendations due to lack of consensus, but make accurate recommendations in limited cases when they can. In contrast, in terms of user interactions, small values of k will always

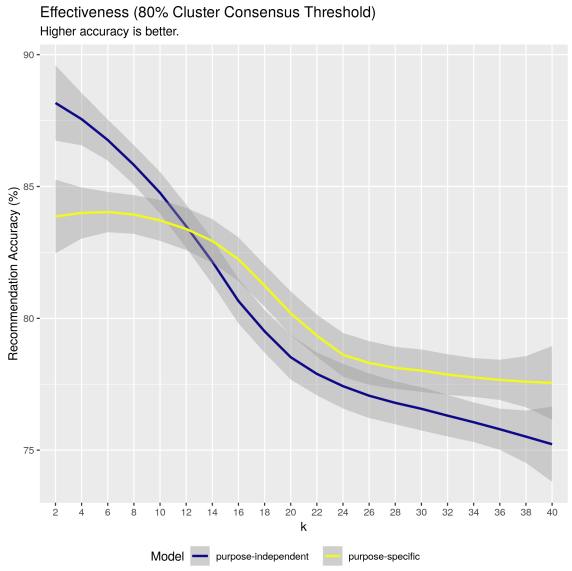
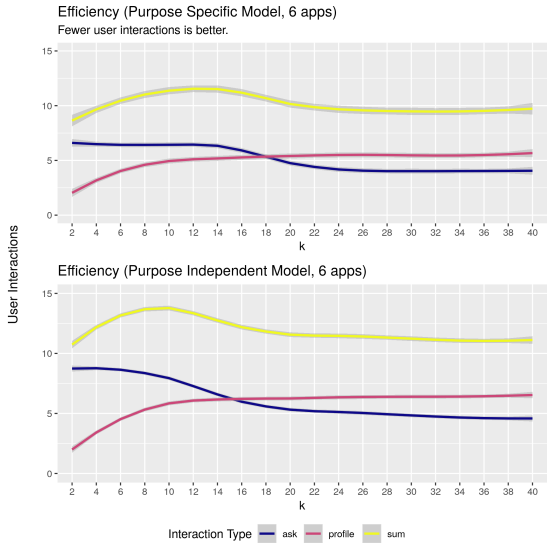


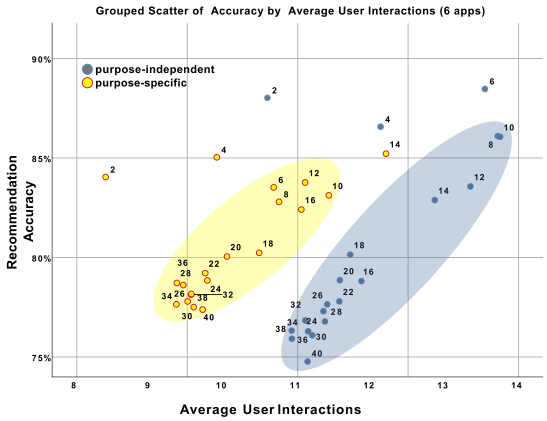
Figure 5.1: There is a minimal difference in accuracy between the purpose-specific and purpose-independent models. The accuracy is higher for most values of k in the purpose-specific model.

be worse than more personalized models (with more clusters) at higher values of k. Recall that once trained, profiles do not change, so the number of questions required to profile an individual is always the same regardless of the number of queries they are subjected to. As one would expect, the overall number of average user interactions increases in the 36-app simulation (Figure 5.3(b)), but the points under both groups show a much more consistent manifold and the outliers now fall within the central tendency.

We observe that while it is easy to profile individuals with a very small number of profiles, the true cost in additional user interactions comes from the profiles' inability to make recommendations afterward. Recall that user burden is the measure of user interactions required to both profile users, and additionally to ask their preferences when a recommendation cannot be made for a particular app. As the number of profiles increases, the number of questions required to profile individuals increases, but this increase flattens out substantially after $k > 10$. This can be seen in Figure 5.2(a) and Figure 5.3(a) as well. Note that there is a similar inflection point in the decreasing trend for additional interactions where recommendations could not be made.



(a) This graph breaks apart the user burden measurement into the number of interactions required to profile a user (profile), and the number of additional interactions required when recommendations cannot be made (ask). The sum of the two is also shown.



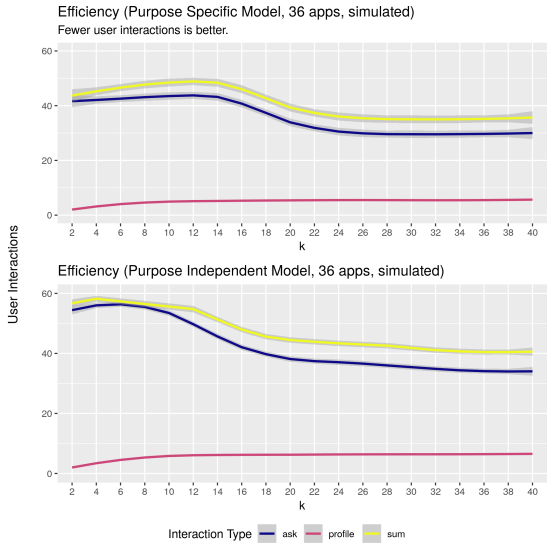
(b) This plot shows the overall relationship between accuracy and user burden at different values of k, using the data for 6 apps. Higher accuracy and fewer user interactions are more desirable.

Figure 5.2: Hyperparameter sweep for 6 apps.

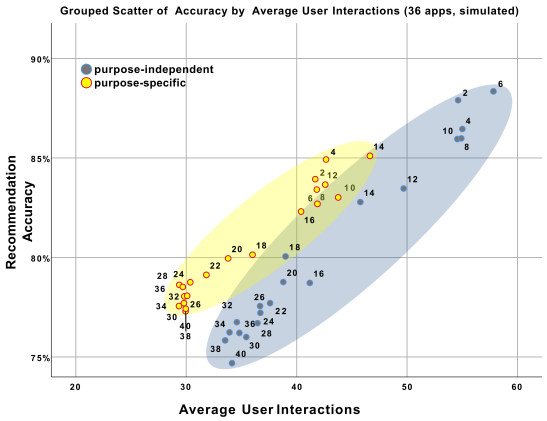
5.3.2 Choosing k for a Given User Interaction Budget

Instantiating a profile-based recommendation system requires that one model (either purpose-specific or purpose-independent) be chosen, using a single value of k. Ideally, one would choose a value which is best suited to a desired level of user burden and accuracy trade-off; either by choosing an upper limit for the number of user interactions, or achieving a particular minimum accuracy.

Choosing one of the values seen in the purpose-specific model (yellow-highlighted point cloud) in Figure 5.2(b) or Figure 5.3(b) would be most ideal overall, as they lie within a Pareto-optimal grouping with higher accuracy and fewer user interactions. A helpful way to frame this is to describe the graph in terms of the maximum accuracy that can be achieved for a given limit on interactions for 36 apps. Note that there is no single optimal k value overall; the choice must be made based on either a maximum budget of user interactions, or a target accuracy. Thus, for a budget of around 30 user interactions, we can see that the purpose-specific model is



(a) The number of interactions required to profile a user is static. The model which does not include purpose makes fewer recommendations, and must ask additional questions more often.



(b) The model incorporating purpose is less burdensome while providing higher accuracy.

Figure 5.3: Hyperparameter sweep for 36 apps, simulated using the Bootstrap distribution from 6 apps.

optimal at $k = 28$ or $k = 24$, achieving an accuracy of around 78%. In contrast, the purpose-independent model has no value of k which can work within this budget. For a budget of around 40 user interactions, the purpose-specific model is optimal at $k = 16$, achieving an accuracy of around 83%. In contrast, the purpose-independent model is optimal at $k = 18$, achieving an accuracy of only 80%. For a budget of around 50 user interactions, the purpose-specific model is optimal at $k = 14$, achieving an accuracy of 85%, well under budget (with only around 45 interactions). In contrast, the purpose-independent model is optimal at $k = 12$, achieving an accuracy of around 83%. It is worth noting that the purpose-independent model is able to achieve the highest accuracy at $k = 6$. In this case, the accuracy is misleading because the model makes recommendations in the fewest circumstances and requires the highest budget (60 user interactions).

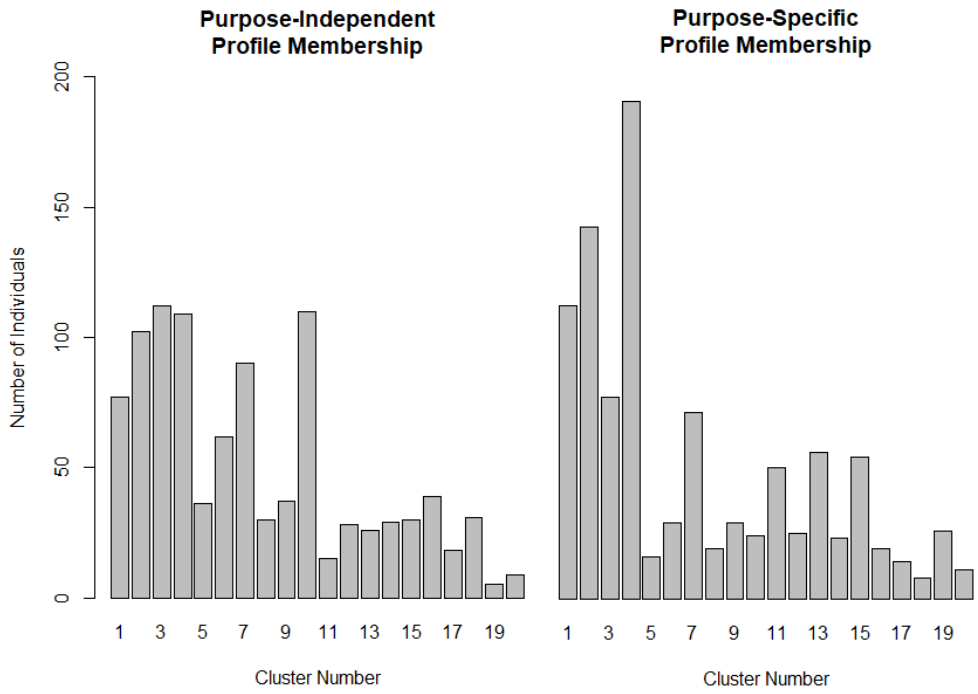
Our assumptions allow the possibility to achieve maximal 100% accuracy by abandoning profiles altogether and resorting to Ask On First Use style permissions. In the case of 36 apps, 3 permissions, and 3 purposes, $36 \cdot 3 \cdot 3 = 324$ user interactions would be required. In contrast, profiles would require new interactions in only 17%

of instances in the worst case (28 interactions at $k = 6$ in the purpose-independent model), and only 8% of instances in the best case ($k = 34$ in the purpose-specific model).

5.3.3 Example at *k=20*

To highlight differences in the characteristics of the purpose-specific and purpose-independent models, we show an illustrative example at one value of k. We can see in Figure 5.4(a) that at $k = 20$, the purpose-independent model shows several dominant clusters. This suggests that a large proportion of users fit into a small number of dominant categories, however, the remaining clusters still account for the majority of users overall. In the purpose-specific model (Figure 5.4(b)) we observe that there is a greater tendency towards a single dominant cluster, but there is greater variability in the proportions of the other clusters. A similar trend in cluster membership was observed across other values of k, particularly those where $k > 8$. This supports the idea that while many individuals generally trend towards similar preferences, there is broad variability that can be better expressed along the extra dimension of purpose. Variability makes preferences more heterogeneous when individuals are clustered among a small number of profiles. This observation is further supported by measures of intra-cluster similarity. Larger numbers of profiles are more internally homogeneous in the purpose-independent model. Comparing silhouette coefficients, we see the average silhouette coefficient for the purpose-independent model (in the $k = 20$ example) is 0.03, and for the purpose-specific model is -0.07. Both coefficients suggest overlapping clusters, which is unremarkable considering the nuances of individual preferences, but the purpose-specific model has slightly less internal homogeneity. However, balanced against the user burden associated with higher values of k, it is clear that small numbers of profiles, while able to achieve accurate recommendations, fail to make recommendations in a large number of instances. Models with low values of k serve to group everyone into a small number of clusters that are largely heterogeneous.

The $k = 20$ example is again illustrative of the kind of predictive power that purpose-specific profiles are capable of. Given 3 permissions, and 25 app categories, there are a total of 1350 recommendations made across $k = 20$ profiles (75 per profile), with 326 null recommendations (where profiles achieved less than 80% consensus on allowing or denying), and 1024 actual recommendations. Once an individual is profiled, recommendations can be made in approximately 76% of circumstances overall, with only very small differences among app categories of varying popularity.



(a) Purpose-independent model. (b) Purpose-specific model.

Figure 5.4: Cluster membership histogram for $k = 20$ profiles. Note the the overall flatter characteristic for 5.4(a), and the tendency towards more dominant clusters for 5.4(b).

There are on average 18 null recommendations per profile, and 57 recommendations per profile across all app categories. The purpose-independent model can profile an individual within an approximate range of 3 to 7 questions, across all values of k. The purpose-specific model can profile an individual within a smaller range of 2 to 6 questions on average.

Our hyperparameter sweep of k values showed two dominant tendencies of values for k, for both the purpose-independent and purpose-specific models. There appeared to be a relationship between accuracy and user burden as the agreement threshold changes; as the threshold is raised, the profiles are able to make slightly more accurate recommendations, at the cost of an increased number of user interactions. In our results, we report on the hyperparameter sweep with an agreement threshold of 80%, as it was found to be the best trade-off between accuracy and user burden. We found that with an agreement threshold of 70%, the mean accuracy of our recommendations

decreased by approximately 5%, and the average number of additional interactions decreased by 3 nearly uniformly for all values of k. At a threshold of 90%, the mean accuracy increased by 5%, and the average number of additional interactions increased by 3 for all values of k.

As can be seen in Figure 5.1, the accuracy overall appears to be within the range of approximately 75% to 90% across all values of k in both models. The accuracy of the purpose-specific model is a few percent higher on average, particularly at $k > 14$, even though it incorporates additional context.

While the difference in accuracy between the two models is not particularly large, there are larger differences in user burden. This can be seen at a glance in Figure 5.2(a) and Figure 5.3(a). Questions about the overall "best" models and k are best answered using the scatterplots in Figure 5.2(b) and Figure 5.3(b). In these graphs, the x-axis is the overall user burden measure, showing the number of user interactions (in the expected case) to perform profiling and recommendations for 6 apps (the sum of the two lines in Figure 5.2(a) and Figure 5.3(a) respectively). The y-axis represents the overall recommendation accuracy (seen in Figure 5.1). The individual points are labeled with the value of k, colored by model type. What is evident is the relationship between accuracy and the number of user interactions. Where the highlighted regions show the central tendency of the two models, one can observe that the purpose-specific model consistently shows fewer user interactions for proportionally higher accuracy overall.

5.3.4 Contextual Factors' Impact on Preferences

In general, we would have expected to see similar results with other machine learning techniques (e.g. collaborative filtering techniques or techniques such as those discussed in prior work [73]). However, in examining the studied contextual factors, we found that preferences change significantly subject to the more expressive permissions which incorporate purpose. There is also evidence that participants cannot distinguish between cases where purpose is unspecified and cases where the purpose is "internal" (i.e., for the app to provide basic functionality). This is shown in Table 5.1, where factors which significantly differ from the null hypothesis are marked with an asterisk. These findings are consistent with prior research which shows the purpose for granting a permission has an effect on the likelihood to allow or deny it [112].

5.3.5 Purpose-Specific Preferences

In addition, our results also strongly argue for the introduction of purpose-specific permissions in mobile operating systems such as Android and iOS. As our results show, people's app permission preferences are strongly influenced by the purpose for which permissions are requested (see Table 5.1). Regulations such as the EU GDPR further mandate obtaining consent from users for the collection of their data for specific purposes [88]. Our results further suggest that, using machine learning, interfaces could be built to mitigate the increase in user burden that would otherwise result from the introduction of purpose-specific mobile app permissions. Such permissions can also be used as a standard way of obtaining consent.

5.4 Summary and Key Takeaways

Mobile apps, unlike web browsers, employ well-defined standardized permissions that are configured centrally and enforced by mobile operating systems. In principle, similar standardization may help improve browser controls as well. However, standards are not a panacea. Mobile app permissions are simpler and more uniform, but also do not account for many important factors that are known to affect users' preferences to allow or deny them. One example is the purpose for an app requesting access to sensitive APIs. Unfortunately, the explosive growth of mobile apps already makes configuring permissions too burdensome without adding additional factors such as purpose. While the introduction of new factors such as purpose has the potential to make the controls more accurately express what people want, this would further exacerbate the burden. In this work, we began by measuring the impact of various factors on the ability to predict people's permission settings, and determined which factors would have an impact on people's likelihood to allow or deny. We approached this by collecting a large corpus of user preferences, and performed regression analysis to determine the impact of surveyed factors on likelihood to allow or deny app permissions across a broad range of different types of apps.

People's preferences are complex – how can we reconcile capturing accurate settings with the corresponding user burden? Unsurprisingly, we found that people's preferences can be influenced by a variety of factors. However, we found evidence that the addition of factors such as purpose, which would ordinarily increase user burden, had the potential to improve the predictive power of our models. These models could be used to build machine learning based recommendation systems, and in turn potentially alleviate user burden by accurately inferring people's preferences.

Then, is it possible to make better predictions without increasing user burden? By experimenting with a combination of supervised and unsupervised machine learning approaches, we found that it was possible to leverage this predictive power and generate recommendations which can make configuration easier. When incorporating additional factors which significantly differentiate users' preferences (such as the purpose), the machine learning models allow more comprehensive settings to be offered. As the accuracy of recommendations improves with this added complexity, configuration requires less manual decision-making. By sweeping a large portion of the parameter space for our approach, we found that it is possible to optimize for accuracy or user burden depending on the chosen parameters. Our results provide guidance for developers to tailor their implementation depending on their specifications for the minimum required accuracy, and the maximum tolerable burden. We also show that by optimizing for both objectives, it is possible to mitigate the trade-off between accuracy and user burden.

Chapter 6

Conclusions

This chapter breaks down the conclusions of the studies seen in the three technical chapters of this dissertation, highlighting the key points. Though these conclusions are most strongly evident within each of the three studies, these points can also be generalized to extract the common thread that ties our findings together. These three key points are enumerated and summarized as follows:

1. **Though today's browsers offer different approaches to awareness and control, none of these approaches adequately address users' needs to be aware of privacy and security risks, nor do they offer the controls that users expect.**

In our first study, we used contextual interviews to assess how a particular set of users interacted with their primary browser. Their goal was to make themselves aware of and control a representative set of online data practices. Our observations were aimed at answering questions about how good these browsers were at ensuring that the users were aware of relevant privacy and security risks. Did the users have the control necessary to mitigate these risks, and are they able to effectively take advantage of the controls to do so? What we found was that the browsers we studied were leaving their users vulnerable to unmitigated risks, which would be far more adequately addressed by using clearer language and more consistent settings. Importantly, these settings need to employ terminology and descriptions which are technically consistent. However, what we were unable to study through these interviews was how users would ideally prefer to manage online data practices. A different approach would be needed to study this aspect of improving the settings. It was possible to explore this in much more detail through the mixed-methods study

which followed. This study revealed the potential impact of standardization.

2. Standardization would help people control their privacy and security settings in a more consistent way, while supporting diverse preferences.

Our first study focused on users' understanding, awareness, and control over their browser settings. In contrast, our second study was unconstrained by the limitations inherent in the design of today's browser settings. In order to understand how people would prefer to configure their settings (assuming it was possible for them to do so everywhere), we collected a large corpus of qualitative and quantitative preferences. Primarily, we were concerned with determining: what level of control do users expect to have over the practices which they are uncomfortable with, and in what ways do similarities manifest between users? Is it possible to take advantage of these similarities to offer better (less redundant, less repetitive) settings, which would reduce user burden? We found that most people wanted to to restrict and be notified about an even broader set of controversial online data practices than our first study explored.

Expecting users to configure these settings across countless individual websites is completely unrealistic. It would be far more effective to offer settings which can be configured in one place, without repetition. This would require standardization, but these standards do not yet exist for browsers. Even if they did, browser settings may be subject to the same unmanageable proliferation as is already seen in other domains. Unfortunately, there is no way to know for sure what this proliferation would look like in browsers until novel standards emerge. Therefore, answering questions about how to further reduce user burden necessitated a different kind of study. If one is to explore ways to address the burden associated with this proliferation, it is necessary to study settings which are principally similar to those we advocate for in browsers. Such an opportunity arose with mobile app permissions, which were the focus of our third study. By studying mobile app permissions, we were able to move beyond the apparent limits of standardization and explore alternative approaches to reduce user burden.

3. **Offering finer-grained and more accurate settings is possible without making configuration more burdensome, but there are requirements that must be fulfilled for this to be achieved:**

    - **If the settings are well-aligned with people's mental models, then it is possible to build machine learning models with strong predictive power. These models can help alleviate user burden.**
    - **If the settings are poorly aligned with people's mental models, then finer granularity will not buy anything; it will cost users more in the form of even greater burden.**

Recall that our first study showed that many users can face difficulty in configuring their browser settings, largely because the settings were poorly aligned with their mental models. In fact, many of our participants developed inaccurate mental models because of how their options were presented, and because of how the risks their options purported to address were described. Both our first and second studies also showed that there are many ways to improve this alignment. These improvements can be based on correlations we observe in people's preferences across contextual factors (which lend themselves to standards), and also based on the use of more technically consistent descriptions of data practices and settings (which may also benefit from standardization). In contrast, though mobile app permissions are already standardized, they are still overly burdensome. Standardization can only so so far, it would seem.

Mobile app permissions also fail to incorporate contextual factors which are already known to have an impact on people's decisions. If these factors are incorporated into mobile app permissions, they become even more burdensome. Is it possible to apply these additional factors to mobile app permissions, to make them more expressive, without making them unmanageable? It seems clear based on our findings of this third study that the answer is yes. However, we have also learned from our prior studies that settings which are poorly aligned with people's mental models will gain nothing by simply offering more options. Would machine learning potentially be able to mitigate this increase by finding the most important correlations among different factors, in order to offer better options? As we have seen in browsers, the proliferation of settings can quickly become redundant or confusing. Standards can only go so far to eliminate this redundancy, at which point it is necessary to consider other approaches to reducing user burden.

Luckily, we can conclude that machine learning offers a promising approach to mitigating this increase in burden. By nature, machine learning takes advantage of the predictive power inherent in correlations seen between users, their associated preferences, and contextual factors. However, machine learning is not a panacea, and cannot be applied without consideration for the conclusions we have derived from our two initial studies. Namely, it would be impossible to apply machine learning to endless arrays of ad hoc settings as seen in browsers and on websites presently. The sparsity of data inherent in defining endless categories of practices and websites would limit the accuracy of recommendations – mobile app permissions have a far more limited taxonomy. What is clear is that extending this taxonomy with a limited number of additional factors, specifically those which matter most (such as purpose), enables machine learning to offer more effective recommendations. It may be possible to apply these recommendations to browsers, but first there are several issues which must be addressed. We discuss these issues in the sections which follow.

6.1 Unmitigated Risks and Unmet Expectations in Browsers

Our first study was intended to identify some areas in which the privacy and security affordances built into five of today's popular browsers are working well, and where they are falling short. Through semi-structured contextual interviews, we worked with participants using their primary browser to determine whether different data practices were present on an example website. We also examined how these participants arrived at their conclusions, and explored their approaches to taking control over the data practices. Here, we focused on determining how well people seemed to understand what they were being exposed to, what options they had to mitigate the associated risks, and evaluated the specific features that people associated (or fail to associate) with awareness and control. In answering these questions, we discovered that many of the issues seen in prior literature continue to be present in today's browsers; people are vulnerable, and easily misled. Even people who appeared to have a high level of technical knowledge still faced challenges in recognizing and controlling data practices.

Many of the problems users faced when taking control of data practices seemed to be rooted in the way in which information was communicated by their browsers. We recognize that our studies were specifically geared towards desktop web browsers, as opposed to mobile browsers, because the mobile context is different, and has limited

resources. The mobile context presents many different challenges when compared to desktop browsers, especially while offering mechanisms for awareness and control. Exploring the extent to which our findings generalize to the mobile browsing context would be a good avenue for further study, as the methods used in this dissertation could be trivially modified to use mobile browsers instead.

However, within the desktop browsing space, some browsers take an approach that describes more general categories of practices, while others present more granular terminology, categories, and descriptions. We did not see an obvious relationship between the granularity presented and our participants' confusion – the problem goes beyond the level of detail or number of categories data practices are divided into. More broadly, even the browsers offering detailed descriptions of data practices confused our participants with imprecise terminology and overly general (or unexpected) categorizations of data practices. The fact is, the browsers we studied were not presenting information or options to our participants in a way which was relatable, and their mental models differed substantially from what appears to be intended by browsers. Our results suggest that people need information which is more precisely aligned with their concerns, in order to prompt them to explore and understand the parameters and limits of their control. More precise information does not necessarily imply the need for more information, rather the right information: users need information that they can relate to, namely information that is aligned with their mental models, and in particular their privacy and security concerns. Not only does this information need to be understandable, it also need to be presented at the right time. For example, browsers could use context-aware/just-in-time notifications, rather than those which require the user to interact with them after the fact, or settings which require a significant level of proactive effort to be configured. Within the spectrum of what is currently offered, we saw that both browsers with simpler controls and those with more expansive controls all seem to cause resignation and lack of confidence at different points. In some cases, simplistic controls, combined with a lack of understanding and high levels of trust in the browser vendor, resulted in overconfidence – for our participants who used Safari in particular, some claimed to "blindly trust Apple (P21)". In all of these cases, it is clear that the issue is not that the controls are too complex or overly simple – they just fail to address what matters to their users.

Among the five browsers studied, there was no single approach that appeared to be ideal, and even those which offered the most comprehensive controls (such as Brave and Firefox) did not necessarily offer the awareness and control people expected. Brave may have been the only browser to directly offer control over data

collection associated with sign-in services like Google and Facebook, but including crypto-mining in a broadly general category called "trackers and ads". Firefox takes a different approach, offering specific settings for crypto-mining, but combining many different categories into "tracking content" which makes it difficult to tell the rest apart. Chrome, Edge, and Safari take this to an extreme, lumping everything together into one general category of "trackers". Our study culminates in the recommendation that browsers should provide clearer and more precise descriptions of data practices which are technically consistent. Browsers should eliminate commonly used but vague terms such as "trackers" and "cookies."

Clearer explanations for the implications of blocking or allowing data practices are also needed. Browsers need to move beyond describing only the potential consequence of websites breaking, due to functional elements being blocked by their decision to restrict certain practices. Every browser we studied offered at least one example of a warning when participants chose to block a particular category of data practices or cookies, which created uncertainty around the consequences. This uncertainty cast doubt on the effectiveness of our participants' approaches to taking control.

Moreover, clearer correspondence between the mechanisms for being made aware of the presence of data practices and the ability to control them is also needed. Unlike Edge, Safari and Chrome, only Firefox and Brave offered settings which both categorized practices granularly and provided ways to allow or block some data practices in one place. None of the browsers offered visibility into whether a particular practice was present on a website independently from whether it had been allowed or blocked. Adding further confusion, browsers used a variety of different terminology to refer to different practices, including some instances where the terminology was technically inconsistent. For example, Firefox used the term "profile" when referring to fingerprinting. Browsers would generally benefit from using technically consistent terminology to avoid this kind of confusion. Standardization also has the potential to assist with this. We discuss standards in the next section in more detail.

6.2 Consistent Settings Require Standards

Ultimately, our findings show that the mechanisms for awareness and control which browsers offer would be improved by standardization. Standards (such as those discussed in § 3.4, § 4.3.5, and § 5.3.5) are needed to ensure there is a uniform way for browsers to describe data practices and the options for control over them. As our

first study revealed, different browsers take different approaches to awareness and control, and there are some benefits and drawbacks amid the studied alternatives. Exploring a broader list of potentially intrusive practices, our second study moved beyond the scope of any one particular browser and focused on people's understanding, awareness, and misconceptions about data practices across multiple contexts. Our second study employed a mixed-methods approach, beginning with collecting qualitative data about about how people generally expected to be made aware of and able to control these practices. These findings corroborated with many of the findings from our first study, and also guided the development of a large-scale quantitative survey focused on how preferences to control data practices varied by practice, website category, and other contextual factors (e.g., website popularity). By collecting and analyzing the resulting large corpus of people's quantitative preferences to opt out, we identified potential ways to address gaps in current browser settings. We identified gaps in capabilities offered to users, including settings which simply didn't exist, yet many people expected comprehensive settings to be available either on websites or in their browser. However, we also found that naively solving the problem by comprehensively offering these settings would be unlikely to improve the situation. What we found was needed was more sensible defaults, which could also serve to reduce user burden. If we were to recommend a default, "deny by default" would be the better choice, even though browsers such as Chrome adopt exactly the opposite default settings.

With this in mind, even though a majority of participants expressed the desire to opt out of most practices in most situations, simply denying them all by default offered only a marginal improvement in accuracy. By contrast, offering comprehensive settings on every website for every practice resulted in a massive increase in user burden. The settings and defaults that we find users would need to have the necessary awareness and control over intrusive practices are missing, but one-size-fits-all settings are inaccurate, and the patchwork of settings offered by individual websites is equally unsatisfactory.

The problem of unsatisfactory settings goes deeper, as we can see that there is something fundamentally lacking between all browsers, namely adequate control, which should be contextually-aware. More specifically, our findings suggest that browsers should provide deny by default settings, but also allow control over more granular categories of practices that are further differentiated by website categories. Such settings would offer the prospect of significantly enhancing user control without imposing undue burden. These settings would require standards to be accepted by website owners as well as browsers in order to be effective. DNT [111] is a clear

example of where voluntary standards have fallen short. As was also seen in our first study, there is a significant and pervasive concern about breakage associated with opting out of practices, and our findings suggest that regulators may need to enact steps to ensure that websites do not intentionally break in an attempt to coerce users into relaxing their settings. Empowered by such regulation, browsers could — in principle — act as neutral agents that empower users to effectively restrict practices that they are not comfortable with, in contrast to websites which stand to benefit from the practices they implement. We acknowledge that it may be difficult for some to envision a future where this is possible amidst the conglomeration of companies who are both performing intrusive data collection online and who are also browser developers – this discussion goes far beyond the scope of this work. Regardless, from a technical perspective, enabling browsers to act as neutral agents on the behalf of users would also require websites to offer standardized APIs. These APIs would enable browsers to communicate users' preferences as they browse, provided that websites (bound by regulation or otherwise) honor the control settings communicated by the browser. If such standard APIs were to be introduced, there would be the potential for machine learning to be applied, which can potentially improve the refinement of settings even further as we detail in the next section.

6.3 Machine Learning Can Help Alleviate User Burden

Our third study is on smartphone permissions, which to some degree employ the greater levels of standardization we see as necessary to control online data practices. As an example, Android incorporates a rich taxonomy of 20-some different categories of apps, and offers at least a dozen different permissions modulating access to sensitive APIs and data. Yet, both our findings and the literature show that these extensive standardized settings still fall short in providing users the control they expect as they are poorly aligned with their mental models. The literature repeatedly identifies the stated purpose for requesting a permission as an especially strong influence on people's decision-making. Our findings further reinforce this notion, as our participants' preferences changed significantly when subject to permissions that incorporate purpose versus those which did not. We also observed that users' preferences did not change significantly between regimes where permissions could not specify the purpose at all, and when the permissions were specified for purposes internal to the core functionality of the app. This phenomenon implies that users may

assume that granting permissions that do not specify a purpose is always necessary for apps to perform their function. Consequently, our observation has implications suggesting permissions as they are currently offered may be easily abused. Thus, it is clear that enabling mobile app permissions to be allowed or denied subject to their purpose would alleviate this problem, and also result in more expressive permissions which better align with people's preferences. Yet, counter-intuitively, to naively introduce these more complex and expressive settings would result in a drastic increase in user burden. Rather than simply increasing the number of settings, we show that it is possible to make smart recommendations to help users configure *better* settings using machine learning.

Crucially, we found that the additional predictive power associated with these more complex models can be leveraged by machine learning in order to make more accurate predictions and recommendations for permission settings. By building models that incorporate additional factors, such as purpose, it is possible to offload the excess burden from the user. We show that machine learning can make recommendations based on the factors and permissions that most strongly influence people's preferences, which can mitigate the increase in user burden that would otherwise result from the introduction of purpose-specific permissions. This has implications that go far beyond smartphone app permissions. If browsers or websites were subject to a similarly standardized taxonomy of website categories, potentially sensitive data practices, and controls based on a model similar to smartphone permissions, machine learning could alleviate the need for large numbers of settings to be manually configured in this domain as well. What's more, by leveraging the awareness and control mechanisms which are emerging for smart devices connected to the Internet of Things [32], a similar approach could be extended to this domain as well.

Fundamentally, we recognize that one of the basic requirements for the machine learning approach detailed in this work is that it must defined based on an enumerable set of attributes that are strongly associated with people's mental models concerning their privacy and security settings. If the chosen attributes are well aligned with people's mental models, the trade-off between accuracy and user burden resulting from finer-grained settings can be mitigated. Practitioners should consider applying the attributes that we show empirically to be associated with users' mental models; such as categories of apps or websites, purposes, or data types and practices. If the chosen attributes do not align with people's mental models, more complex settings with finer granularity may be received unpredictably, and our approach will be unlikely to help.

6.4 Future Work

Though this dissertation addresses several research questions which inform the design and refinement of more effective privacy and security controls, many of our answers give way to more questions which could be answered with future studies. As a discipline, software engineering has only begun to scratch the surface of how best to approach privacy and security generally, and our work provides a small step forward into a vast frontier. There is still far more work to be done, with far more areas left to explore. In this section, we identify some of the more promising questions, and provide guidance for methods that may serve to answer them.

6.4.1 Organizing Privacy and Security Concepts

In this work, we used a taxonomy of data practices that was tested and refined based on focus groups and pilot studies. The goal was to create categories of data practices that were technically consistent, commonly encountered, and broadly representative. We also used other existing taxonomies, such as categories of websites based on Amazon's Alexa web rankings [10], and categories of smartphone apps seen in the Google Play Store, to establish context. All of these taxonomies represent ways of classifying and organizing privacy and security concepts, whether they are data practices, or contextual factors.

The issue with these taxonomies is that they are necessary to contextualize study participants' responses, which is crucial to eliciting meaningful preferences and perspectives, but also they present a limitation in terms of scope and generalizability. What's more, testing these various organizations can only test one set of assumptions about the true underlying mental models of privacy and security preferences which people have. Future work should attempt to address this by studying different categories of practices, contexts, and alternative interpretations of the same. Do people feel differently about these concepts, or do their preferences change, when they are organized or framed differently? There is abundant evidence in the literature which suggests that people's expressed preferences will indeed change depending on the framing and interpretation of these concepts [3, 5, 109, 90, 86, 71], but is there a way that this phenomenon can be used to encourage people to engage with their settings more thoughtfully, or to encourage people to express their true preferences? Is it truly the best approach for researchers to remain neutral in their framing, in terms of the outcomes for users and influencing their protective behaviors? These questions are far beyond the scope of this work, but they are important questions

none the less. It may be possible to answer these questions using variations of some of the approaches used in this work, such as a variation of our survey methodology, where neutral framing was used.

In addition to alternative framing, new types of data practices, websites, apps, and other factors will need to be considered. Crucially, researchers should employ methods such as those we have incorporated into our study of purpose as a contextual factor in mobile app permissions. Such methods can be applied to explore questions about whether there are other important contextual factors which have a strong influence on people's preferences, and if these should be incorporated into settings offered to users. It will be important to assess how these factors are perceived by users, and how well the settings offered subject to these factors can reflect users' mental models. It will also be important to consider the technical accuracy of the descriptions given to users, and evaluate their consistency of interpretation between contexts (especially new or changing contexts). In the case that newly introduced contextual factors also add complexity into the settings, applying machine learning (such as the technique that we have demonstrated in this work) should serve as a way to manage this complexity, which we show to be associated with the proliferation of context-specific options. It would be reasonable to expect that this proliferation will continue into the future, and researchers should determine the extent to which the methods we offer can scale over time.

6.4.2 Understanding the Evolution of Preferences

Though it may be the case that the ways in which privacy and security concepts are organized will change over time, future work should also explore how people's perceptions and preferences change over time as well. Taxonomies may be subject to many different interpretations, and may need to evolve as data practices, their underlying technologies, and societal values change. To do this in a way which upholds the values we seek to promote in this work, some key questions should be addressed: are there ways to define data practices in ways that are equally understandable, actionable, and robust to changes over time while remaining technically consistent? This question was hinted at in the previous section, as it leads to a more important question: how do people's perceptions, concerns, and preferences change over as these methods of organizing privacy and security concepts (and their associated interpretations) also change? These two questions can potentially be answered using many of the same methodologies in the studies presented in this work, but may also require adapting our methodologies to collect longitudinal data. Even though the

categories, classifications, and taxonomies we used may need to change to suit the projected status quo, the surveys and other instruments used in this work will likely still be applicable. At the very least, the instruments used in this work can form the basis of versions which are better adapted to reflect the changing circumstances into the future.

We recognize that it should be expected that people's privacy and security preferences will not remain static. Users' perceptions and preferences and more will likely change over time due to societal changes, events in the media, and so on. It will be important for future work to explore how these changes occur. Thus, future work should perform regular studies similar to those we have presented in this work, perhaps with additional data collected over time at regular intervals. This approach offers one way to explore the trends or patterns which may shift with time. However, this approach will be difficult because the effect of using different terminology and definitions which vary based on changes in technology can be unpredictable. As our studies conclude, the goal of this work should be to ensure that privacy and security concepts and related information are communicated in a way which people can understand, so that they can build mental models which reflect the reality of contemporary practices, and their options, based on real world data. Naturally, alternative classification and organization methods should carefully the design principles we have outlined in this work, and can use the methods of assessment we have used to determine whether the key dimensions we have highlighted are upheld. While we have proposed standardization in this work which should serve uphold these qualities to some extent, such standards will also need to be updated and evolve over time as well.

6.4.3 Exploring Alternative Interaction Designs

In this work, we have explored two modes of interaction with privacy and security controls: first, browser privacy and security settings, which generally take the form of interactive menus, buttons, toggles, and similar proactive options [96]. In contrast, the second mode of interaction we explored was smartphone app permissions, which are only offered as just-in-time blocking notifications. These permissions can be further augmented with recommendations made by machine learning based privacy assistants which could take either form [72, 26, 42]. In each of these cases, the expectation is that the user will be presented with a set of options, think about their preferences, and make a decision either in the moment that the permissions are requested, or when they review their settings. There are a host of other forms of

interaction that may have the potential to be even better aligned with users' mental models, preferences, and expectations, but these were not studied in this work.

Future work could explore many different alternative modes of interaction with privacy and security controls. For instance, one approach to configuring these settings which may have the potential to encourage users to regularly revisit their choices are through nudges [3]. These nudges could transform settings such as those seen in browsers into just-in-time options, which are much more similar to smartphone permissions. However, it is unclear whether this mode of interaction has the right trade-offs. Perhaps just-in-time controls could engage users more directly, but are they more likely to get in the way of the users' primary browsing tasks? While we have advocated for the design of privacy and security settings that are ideally well-aligned with users' preferences and mental models, it may be the case that the presentation of these options has a discernible impact on people's preferences, just as the various ways that contextual factors can be interpreted have a similar impact. One way such questions could be addressed would be to design controlled experiments which incorporate these alternative interaction styles against contemporary styles. These experiments would take the form of A/B testing, which could quantitatively focus on user burden metrics, such as time taken, or number of clicks [65]. Such experiments may also incorporate qualitative metrics such as those we explored in our studies. In A/B tests incorporating existing taxonomies of practices and contexts, alternative interactions would serve as the treatment, with the existing options and modes of interaction as a control. This form of study could have the potential to answer important questions about users, too. These questions include: how do users' impressions of their options change qualitatively as their options are presented differently? What makes their preferences quantitatively change, and are these changes associated with the measures of user burden each approach entails?

Alternative interactions need not fundamentally change the settings which are offered, but could potentially incorporate additional details in-situ which current user interfaces do not provide. This was a clear limitation of our first and second studies, which focused on browsers and the web generally, but did not incorporate browser add-ons or extensions. There are many examples of usability studies that focus on these add-ons in the literature [115, 98, 77]. However, more work is needed to elicit people's understanding and awareness of what they can provide – particularly in contrast to what is available in browsers without these add-ons.

There is also more work to be done studying alternative interaction designs that are not based on graphical user interfaces. For instance, a conversational assistant could verbally explain to users what the implications of changing their settings might

entail in real-time. These explanations could also potentially incorporate the ability for users to ask questions or propose "what-if" scenarios [92]. Such conversational assistants have already been proposed in the literature [32, 92, 91], and may have the potential to make configuring settings easier. However, care must be taken to ensure that these approaches are well studied within the contexts that they are intended to be deployed in. It may not be the case that a particular interaction style is suited to both web browsing and smartphone permissions, and it is not obvious which style is best for an individual user. Moreover, whatever approach is taken must address issues associated with generalizing to the large variety of different contexts online, or accommodating the limited resources available on mobile platforms. Regardless of what domain they are intended for, interactions must be both limited in scope and unobtrusive since privacy and security management is a secondary task [3]. Field studies, such as those seen in prior literature proposing personalized privacy assistants [72, 32], may be a good approach to evaluating various alternative interaction styles. Once a particular form of interaction style has been proven effective, it should be considered whether it might benefit from being made standard to ensure consistency across platforms. Encouraging adoption of such novel standards may prove difficult, which we discuss in the following section.

6.4.4 Addressing Challenges With Standards

In this work, we advocate for several possible types of standards which we show have the potential to alleviate many of the issues users face with managing privacy and security settings. Our results argue for standards to ensure that users have the ability to control potentially intrusive practices online as they browse; these ideally should take the form of standard APIs which enable browsers to capture users' preferences to allow or restrict potentially intrusive practices on websites. These preferences would then be communicated by browsers to websites on the user's behalf. Standards are also needed to ensure that website operators respect users' preferences, which would be specified as part of such APIs. Moreover, we identified that there is the potential for website operators to co-opt these standards by restricting or intentionally breaking website functionality for users who choose to restrict practices that website operators do not wish them to. This is a major challenge that must be addressed proactively, in order to ensure the consistent and fair adoption of standards. We believe that this challenge could perhaps be addressed with regulation, which would restructure the misalignment in incentives between website owners and users. With DNT serving as an example of the failure of purely voluntary standards [111], what

remains unclear is whether there are methods that can be effective in encouraging website operators to adopt standards at all without enforcing compliance through regulation. However, there are no simple answers when it comes to the form that such regulations should take, or how new standards could be written to avoid the need for heavy-handed regulation. What are the limits of standardization? Can standards be enforced on a voluntary basis, or are voluntary standards destined to fail without restructuring incentives? These are questions that public policy researchers and software engineering researchers must collaborate on, and the literature shows that there are ways in which these two groups can meaningfully engage to produce scientifically-derived policy recommendations [51, 31]. This is an area where there is a huge amount of future research that is needed.

Future work must also carefully consider the long-term consequences of novel regulations, exploring in detail how they can be written robustly amidst potentially unpredictable technological changes and advancements. Is it possible to make standards in domains like web browsing and the Internet of Things, which are able to tackle the rapid advancements in sensing, data collection, processing, and inference we are already seeing today? Answering this question in a way which can truly generalize may not be possible, but researchers should be encouraged to make attempts to draft new standards, and test them by developing systems that demonstrate compliance with the standards. Such future work should explore the possible design alternatives that can exist within the constraints imposed by proposed standards. Skeptics of emerging web and Internet of Things standards can take heart – we can see evidence of this type of work being effectively employed in the literature already [32, 42]. Evaluating prototypes and novel designs in conformance with emerging standards can (and should) be done in a systematic way which upholds privacy and security. The literature has also shown that the result of such work can serve as a point of reference for policymakers too [51, 67, 28].

Our work has identified that standards should incorporate factors which we show to help align the options users are provided, with their mental models. They should incorporate categories of practices, website categories, app categories, purpose, and other contextual factors which have been empirically validated. Regulations for data protection and consent such as GDPR [88] already incorporate this concept of contextual integrity [15] to a large degree. Yet, studies have shown that the ad hoc way website operators comply with technology-neutral regulations like GDPR is problematic, and clearly defined prescriptions for how to adequately design user interfaces which can meaningfully comply with these regulations are still needed [16, 14, 51]. This is a major challenge with standards adoption generally that must be

addressed before the novel standards we advocate for can be fleshed out.

As we previously discussed regarding alternative interaction designs, it is still unclear what the best approach is to elicit and communicate users' myriad preferences, particularly in such a massively heterogeneous context as the web or Internet of Things. We see that the fully-specified model of settings demonstrated in mobile app permissions is a promising place to start for future work, based on the fact that we have also identified (and attempted to rectify) issues that still exist in this highly specified domain. It is likely that future research will reveal that there are additional challenges inherent in creating uniform standards in other domains even as their standards become more fully specified. In particular, work within the domain of the Internet of Things shows that the proliferation of standards which are incompatible with one another is a real risk [32]. If standards cannot be applied, the machine learning approach we show in this work cannot be applied either. There are no clear solutions which fully mitigate this risk at present. Perhaps there are approaches to incorporate machine learning in different ways. For example, there may be some possibility to use machine learning techniques to collapse heterogeneous categories of data, data practices, and other contextual factors into smaller numbers of standard categories that retain a high degree of semantic association.

6.5 Final Thoughts

Many of the unanswered questions we propose in this chapter are rooted in the ever-changing technologies and practices that people encounter online. Security and privacy controls continue to proliferate amid new regulations, increasingly complex data practices, and people's urgent need to take control of their data. Yet in spite of this, managing privacy and security choices is getting harder, not easier – but this dissertation demonstrates that there are at least a few promising ways to combat this trend. We show that it is not inevitable that users must be baffled by complex arrays of opaque settings in order for their preferences to be adequately captured. Rather, we show that by improving alignment with users' mental models, thoughtfully considering contextual factors, and by applying machine learning, it is possible to refine existing privacy and security settings. In fact, we can improve them a great deal.

In this work, we explored security and privacy settings in web browsers and mobile apps, aiming to determine how effective they are at giving users the awareness and control they need, identifying areas in need of improvement, and informing

ways to improve. Through three studies, we determined whether users can identify risks, whether they are aware of privacy and security controls, and whether they can effectively use those controls to restrict undesirable behaviors and mitigate risks. We identified areas where standardization could make managing these controls easier and more effective. We revealed expectations which do not mesh with reality. We identified trade-offs impacting both the accuracy of the controls which are offered and the burden of configuring them. Finally, we proposed and evaluated novel machine learning techniques which show promise to help overcome these trade-offs. Combined with regulations to help normalize and standardize our recommendations, there is reason to be optimistic about what the future will offer for empowering users with awareness and control over emergent privacy and security risks.

Appendix A

Definitions and Descriptions of Data Practices

Table A.1: Risks and benefits associated with each potentially intrusive data practice which were provided as part of surveys and interviews seen in Chapters 3 and 4.

Practice	Risks	Benefits
Identity/Sign-In Services	-This could be used to track you across many websites that may not be related -Allows inference of personal details that may be used for purposes other than logging in	+Don't need to remember as many passwords +Don't need to re-enter personal information or your account username and password with every new website you log in to
Targeted Advertising	-Data collected by advertisers may be used in ways you didn't anticipate, and for purposes other than advertisements	+Ads you are shown may be more relevant to your interests
Behavioral Profiling	-Facts may be inferred about you which are sensitive, or may make you feel uncomfortable -In some jurisdictions, profiles can be bought and sold and you have no rights to them	+May enable websites to improve products and services that they offer to you
Session Replay	-May reveal sensitive information, or information in a sensitive context	+May enable websites to improve products and services that they offer to you
Reporting and Analytics	-May reveal personal information, or information in a sensitive context	+May enable websites to improve products and services that they offer to you
Fingerprinting	-Can prevent you from remaining anonymous, by identifying you even when you've taken steps to hide your identity (e.g., after you've cleared cookies or used the privacy mode in a browser)	+May enable websites to offer better security features, which can protect your account and account information
"Nag" Screens	-Can prevent you from accessing content, even in the middle of reading it	+May help websites ensure that their business meets regulatory requirements in some jurisdictions +May help to ensure that the website earns enough revenue to continue operating
Crypto-Mining	-Can negatively affect the performance of your device, which can also disrupt your browsing experience	+May enable websites to improve your browsing experience +May enable websites to remove ads or give you access to premium content

Table A.2: Descriptions of data practices which were provided as part of surveys and interviews seen in Chapters 3 and 4.

Practice	Definition Provided To Participants
Identity/Sign-In Services	Identity/Sign-In Services help you log in to websites without relying on passwords specific to these websites. Examples are "Log in with Google", "Log in with Facebook", and "Sign in with your Apple ID". These services save you the effort of creating and remembering passwords for individual websites. Because they see the websites you access, these services might be able to infer details about you, such as your interests, education, income, and more. This information could be used for purposes that go beyond helping you log in.
Targeted Advertising	Targeted Advertising uses information collected about you to tailor the advertisements that are shown to you on a particular website.
Behavioral Profiling	Behavioral Profiling collects information about who you are, your interests, and the things you do, to categorize you into specific categories (or profiles). For example, a website might try to determine your age, whether you are an "impulse buyer," your political beliefs, and potentially much more. Sometimes, the profiles can be incorrect. The use of Behavioral Profiling does not necessarily mean that you will be subjected to advertisements, but it does mean that information may be collected and inferred about you.
Reporting and Analytics	Reporting and Analytics monitors what is happening as you are browsing websites, and generates technical information for the website developers. Often, this includes information about the state of your device, browser, and may also include technical information about what happened during your interaction with a particular website. This can help websites improve their products and services, but can potentially reveal sensitive information.
Session Replay	Session Replay creates detailed logs that record the actions you take while browsing a particular website and sends these logs to the website owners. This means that website owners can observe and replay exactly what you did and what you saw. Note that this is not a feature that enables you, as the person browsing a website, to replay what you did. Sometimes, sensitive information can be found in these recordings because it wasn't properly removed.
Fingerprinting	Fingerprinting is a technique which ensures that the website you are browsing always recognizes you, even if you are not signed in. Fingerprinting also enables websites to detect whether a device that the website doesn't recognize is interacting with the website. This can be useful for a variety of reasons, such as detecting when someone tries to access an account on a new or unrecognized device. This technique does not mean that the website is using biometrics (i.e. a fingerprint scanner) to identify you, and the technique has nothing to do with physical fingerprints. Rather, Fingerprinting refers to ways that your device can be picked out and recognized among others.
"Nag" Screens	Nag Screens can force you to see a popup, to watch an ad, to prevent you from viewing content, or otherwise to do something that disrupts your normal browsing experience. Sometimes "Nag" Screens appear when you're using an ad-blocker, or because the website needs you to interact with something, such as giving consent where required by law.
Crypto-Mining	Crypto-Mining uses your device to generate digital cash, such as Bitcoin, during the time you spend browsing a particular website. Generally this digital cash is sent to the owners of the websites, but in some circumstances you may get a share. Some websites use Crypto-Mining as a way of using your device to make money for the website instead of (or in addition to) advertisements. Since it uses your device's processing power to work, Crypto-Mining uses electricity or battery power on your device, and can affect device performance when you browse websites that employ Crypto-Mining.

Table A.3: Opt out scenarios for each data practice provided as part of surveys seen in Chapter 4.

Practice	Opt-Out Scenario (Specific Websites)	Opt-Out Scenario (All Websites)
Identity/Sign-In Services	Imagine that you are given a new setting in your browser that enables you to block Identity/Sign-In Services on specific websites you choose ("opting out" of Identity/Sign-In Services on these websites), requiring you to log in to these specific websites manually instead. This also requires you to log in to these specific websites separately. When enabled, the buttons and links to use Identity/Sign-In Services on the specific websites you opt out from are removed from the websites you opt out of, and the ability for these services to collect data is also removed on these websites. Assume that you will be able to reverse this setting on any website, at any time.	Imagine that you are given a new setting in your browser that enables you to block Identity/Sign-In Services on all websites ("opting out" of Identity/Sign-In Services), requiring you to log in manually instead on all websites. This also requires you to log in to all websites separately. When enabled, all the buttons and links to use Identity/Sign-In Services on websites are removed, and the ability for these services to collect data is also removed. Assume that you will be able to undo this setting at any time.
Targeted Advertising	Imagine that you are given a new setting in your browser that allows you to block Targeted Advertising on specific websites you choose ("opting out" of Targeted Advertising on these websites). When enabled, ads which use Targeted Advertising are blocked on websites which you opt out of. Ordinary ads which do not use Targeted Advertising are not affected by this setting. By default, you are still shown Targeted Advertising on websites which you are not opted out of. Assume that you will be able to reverse this setting on any website, at any time.	Imagine that you are given a new setting in your browser that enables you to block all Targeted Advertising ads ("opting out" of Targeted Advertising). When enabled, ads which use Targeted Advertising are blocked on all websites. Ordinary ads which do not use Targeted Advertising are not affected by this setting. Assume that you will be able to undo this setting at any time.
Behavioral Profiling	Imagine that you are given a new setting in your browser that enables you to block Behavioral Profiling from occurring on specific websites you choose ("opting out" of Behavioral Profiling on these websites). On the websites you opt out from, your browser hides your identity and blocks any information your browser might send in the background while you are browsing. This ensures the specific websites you opt out from cannot perform Behavioral Profiling on you. Assume that you will be able to reverse this setting on any website, at any time.	Imagine that you are given a new setting in your browser that enables you to block all websites from performing Behavioral Profiling ("opting out" of Behavioral Profiling). When enabled, the setting hides your identity and blocks any information your browser might send in the background while you are browsing. Assume that you will be able to undo this setting at any time.
Session Replay	Imagine you are given a new setting in your browser that enables you to block Session Replay from occurring on specific websites you choose ("opting out" of Session Replay on th ese websites), preventing the websites you opt out from collecting what is needed for Session Replay to occur. By default, on websites you have not opted out from, Session Replay will still occur normally. Assume that you will be able to reverse this setting on any website, at any time.	Imagine you are given a new setting in your browser that enables you to block Session Replay from occurring on all websites ("opting out" of Session Replay), preventing all websites from collecting what is needed for Session Replay to occur. Assume that you will be able to undo this setting at any time.
Reporting and Analytics	Imagine that you are given a new setting in your browser that enables you to block specific websites you choose from performing Reporting and Analytics ("opting out" of Reporting and Analytics on these websites). When enabled, your browser sends misleading signals to the websites that you opt out from, preventing the Reporting and Analytics mechanisms on websites from working there. By default, Reporting and Analytics will still work as it would normally on websites you do not choose to opt out from. Assume that you will be able to reverse this setting on any website, at any time.	Imagine that you are given a new setting in your browser that enables you to block all websites from performing Reporting and Analytics ("opting out" of Reporting and Analytics). When enabled, your browser sends misleading signals to all websites, to prevent the Reporting and Analytics mechanisms from working anywhere. Assume that you will be able to undo this setting at any time.
Fingerprinting	Imagine that you are given a new setting in your browser that enables you to block Fingerprinting from occurring on specific websites you choose ("opting out" of Fingerprinting on these websites). When enabled, the setting sends misleading signals to the websites you opt out from, which prevents Fingerprinting from taking place on those websites. By default, websites which you have not opted out from will still allow the Fingerprinting to take place as they would normally. Assume that you will be able to reverse this setting on any website, at any time.	Imagine that you are given a new setting in your browser that enables you to block Fingerprinting from occurring on all websites ("opting out" of Fingerprinting). When enabled, the setting sends misleading signals to all websites, which prevents Fingerprinting from occurring. Assume that you will be able to undo this setting at any time.
"Nag" Screens	Imagine that you are given a new setting in your browser that enables you to block "Nag" Screens on specific websites you choose ("opting out" of "Nag" Screens on these websites). When enabled, your browser blocks "Nag" Screens on specific websites, removing them from the contents of websites you opt out from. By default, on websites you have not opted out from, "Nag" Screens are still shown as they would normally be. Assume that you will be able to reverse this setting on any website, at any time.	Imagine that you are given a new setting in your browser that enables you to block "Nag" Screens ("opting out" of "Nag" Screens). When enabled, your browser blocks "Nag" Screens everywhere, removing them from all websites. Assume that you will be able to undo this setting at any time.
Crypto-Mining	Imagine you are given a new setting in your browser that enables you to block Crypto-Mining on specific websites you choose ("opting out" of Crypto-Mining on these websites), preventing Crypto-Mining from taking place on these websites in your browser. By default, on websites which you have not opted out from, Crypto-Mining is still allowed. Assume that you will be able to reverse this setting on any website, at any time.	Imagine you are given a new setting in your browser that enables you to block Crypto-Mining on all websites ("opting out" of Crypto-Mining), preventing any Crypto-Mining from taking place in your browser. Assume that you will be able to undo this setting at any time.

Appendix B

Surveys

B.1 Study: Managing Online Data Practices

The following surveys were used as part of the study seen in Chapter 4.

B.1.1 Qualitative Survey 1

Section 1

In the following section, you will be asked to provide examples of websites which you routinely browse, based on a number of categories. The questions concerning each category will be presented in random order. The categories are as follows: News and Information, Entertainment and Games, Shopping, Travel, Finance, Adult, Health and Wellbeing, and Social Media and Blogging

Later in the survey, we will be asking you questions which use the examples you provide us to set the context.

Section 2

If you are unable to provide 2 examples, or are uncomfortable with providing 2 examples for a category, you may proceed and examples will be provided for you.

Please note that if you do not provide 2 examples for at least 4 out of the 8 categories, you will be automatically withdrawn from the study.

[The website categories which follow are presented in random order.]

Take a moment to think of two **News and Information** pages you have visited, or which you browse routinely. These are websites which can include news papers, online journals, Wikis, and any other source of news or information. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Entertainment and Games** pages you have visited, or which you browse routinely. These are websites concerning digital, print, online, and other forms of media and entertainment, including video games, movies, gambling, and more. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Shopping** pages you have visited, or which you browse routinely. These are websites where you can purchase goods and services online, and browse for items you wish to purchase. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Travel** pages you have visited, or which you browse routinely. These are websites concerning booking travel and accommodations,

travel planning, reviews, hotels, and more. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Finance** pages you have visited, or which you browse routinely. These are websites which include trading, online banking, financial advice, market-related information, and more. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Adult** pages you have visited, or which you browse routinely. These are websites which include sexually (or otherwise) explicit materials, including videos, photos, and other material not intended for consumption by minors. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Health and Wellbeing** pages you have visited, or which you browse routinely. These are websites which concern medical, spiritual, dietary, and other forms of advice and discussion for the betterment of your physical, mental, and spiritual health. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Take a moment to think of two **Social Media and Blogging** pages you have visited, or which you browse routinely. These are websites which belong to social media networks, blogs, or other forms of online social interaction. If it helps, check your browsing history and see if you can find good examples. Enter the names of the two websites you think of into the fields below. **Please only enter the names of the websites into the fields below.**

[Participant is presented with free text entry fields.]

Section 3

The next part of the survey is intended to collect information about your thoughts and experiences with a particular web technology. Please read the text on the following page carefully. After, you will be asked a series of questions.

[Participant is presented with the description of one PIP.]

Prior to this survey, had you ever encountered any examples of *[PIP]*?

[Participant may choose between: Yes/No]

What do you think the **risks** might be for you when you browse a website with *[PIP]*?

[Participant is presented with a free text entry field.]

What do you think the **benefits** might be for you when you browse a website with *[PIP]*?

[Participant is presented with a free text entry field.]

Here are some examples of concrete risks and benefits associated with *[PIP]*:

[Participant is presented with the risks and benefits for the PIP.]

Section 4

In this section, you will be asked about *[PIP]* in a variety of scenarios. You can hover over the (i) symbol to remind you of the definition of *[PIP]*. Opting out of *[PIP]* means that:

[Participant is presented with the specific PIP scenario.]

[This form of question repeats for all of the specific websites, and website categories that the user provided in the priming exercise in the first part of the survey, in random order:]

Consider *[specific user-provided website]*, which is a *[website category]* website.

If you had a single one-click setting which enabled you to opt out of *[PIP]* on *[specific user-provided website]*, how likely would you be to use it? *[Participant is presented with 4-point Likert scale response options ranging from Very Unlikely to Very Likely with no neutral response.]*

Consider all the *[website category]* websites across the entire internet, which includes *[specific user-provided website]* and *[specific user-provided website]*.

If you had a single one-click setting that enabled you to opt out of *[PIP]* on all *[website category]* websites, how likely would you be to use it? *[Participant is presented with 4-point Likert scale response options ranging from Very Unlikely to Very Likely with no neutral response.]*

[Participant is presented with a randomized attention check question, which includes a reCAPTCHA test.]

Section 5

What benefits do you think that companies which have *[PIP]* on their website get from *[PIP]*?

[Participant is presented with a free text entry field.]

Are you aware of anything you can do to enable or disable *[PIP]* while browsing? Please explain.

[Participant is presented with a free text entry field.]

Have you ever tried to enable or disable *[PIP]*?

[Participant may choose between: Yes/No]

Why or why not? *[Participant is presented with a free text entry field.]*

[If the participant answered yes:] How did you know if you succeeded or failed? Would you want to be informed about the presence or absence of [PIP] on the websites you browse?

[Participant is presented with 5-point Likert scale response options ranging from Definitely Yes to Definitely not with a neutral response of I don't know.]

[Participant is presented with the post-survey.]

B.1.2 Quantitative Survey 2

Section 1

The next part of the survey is intended to collect your thoughts and experiences with a particular web technology.

Please read the description of the technology on the following page carefully. You will be asked a series of questions which depend on you having read the description.

[Participant is presented with the PIP description.]

Here are some examples of concrete risks and benefits for *[PIP]*.

Please take note of these risks and benefits and consider them carefully as you progress through the rest of the survey.

[Participant is presented with the list of PIP risks and benefits.]

Throughout the survey, you can click on the following button located at the top of each page, to see a reminder of the definition of *[PIP]* and the risks and benefits associated with it.

[Participant is presented with a button labeled with the name of the PIP, which is present throughout the survey on the top of each page.]

Section 2

In this section, you will be asked about *[PIP]* in a variety of scenarios.

Please carefully consider the definition of *[PIP]* and the associated risks and benefits you just saw when answering the questions which follow.

[The following section repeats for all the specific websites participants were asked about, across all website categories, in random order.]

[Participant is presented with a screenshot of a specific website, along with the name of the website, their logo, the date they were established, the country they are based in.]

[website name] is a *[website category]* website, established *[date]*, based in *[location]*. Please take a moment to familiarize yourself with the website if you aren't already familiar with it.

[Participant is presented with the scenario text for opting out of a specific website.]

Consider *[website]*, which is a *[website category]* website. How likely would you be to use the setting described above to **opt out of *[PIP]* on *[website]***?

[Participant is presented with 4-point Likert scale response options ranging from Very Unlikely to Very Likely with no neutral response.]

[Participant is presented with the scenario text for opting out of a website category.]

Consider all the *[website category]* websites across the internet, which includes the two websites you saw a moment ago, *[specific website]* and *[specific website]*, and many others.

How likely would you be to use the setting described above to opt out of *[PIP]*

for all *[website category]* websites?

This setting would not affect your separate choice to opt out (or to not opt out) for specific websites.

[Participant is presented with 4-point Likert scale response options ranging from Very Unlikely to Very Likely with no neutral response.]

Section 3

[Participant is presented with the scenario text for opting out of PIP on every website.]

Note that this would apply to every website you visit, no matter what category it belongs to.

How likely would you be to use the setting described above, to opt out of *[PIP]* **on every website you visit**?

[Participant is presented with 4-point Likert scale response options ranging from Very Unlikely to Very Likely with no neutral response.]

Section 4

Please answer the following questions about how often you would like to be notified about *[PIP]* on different categories of websites. *[The participant is presented with a matrix of questions from all website categories in randomized order.]*

On *[website category]* websites, how often would you like to be notified about *[PIP]*?

[Participant is presented with 5 response options: Notify every time I visit, Notify me only once per week, Notify me only once per month, Notify me only the first time I visit, Never notify me.]

[Participant is presented with the post-survey.]

Appendix C

Interview Scripts

C.1 Examining Browser Privacy and Security Settings (Interview Script)

This interview script was used during contextual interviews seen in Chapter 3. The interview script proceeds below. Instructions for the interviewer are *italicized*, and lines in the script which are intended to be read by the interviewer are surrounded with quotation marks. Portions of the script which are intended to be filled in with information from elsewhere are surrounded in square brackets. The interview begins after the participant has joined the screen-sharing session showing the browser they are assigned to, on the example page.

Before you begin, ensure the participant has completed their demographic pre-interview survey and that they are registered for the task on Prolific. First, confirm the participant's consent to record the interview, then begin recording. Ensure the meeting is configured such that the participant's video feed is disabled.

Introduction

"What you are currently seeing is a real web browser running on my computer. The web page you see in the browser is not real, it's just a made-up example. I'm going to ask you questions about what you're seeing, and I want you to know that you can ask me questions about what you're seeing at any time. You can also ask me to click on things for you in case you want to see what happens, and I'll let you know

if for some reason I'm not able to click on what you asked. We are going to work together. You can interrupt me to ask me anything you want at any point during the interview, don't be shy. The most important thing I need from you is to explain what you're looking at and what you're thinking, especially if you ask me to click on something."

"Do you recognize this browser?" "What about it is familiar?"

We want to confirm that this is the browser that the user says is their primary browser, to avoid any mistakes that may have resulted from failure to correctly screen participants.

"What browser do you use most often on your desktop or laptop computer?" "How often do you surf using this browser on an average week, roughly speaking?" "What other browsers do you use?" "How often?"

Now we should make it clear what browser it is that they are seeing in case they had trouble recognizing it. There is some possibility that participants may become confused due to variations in versions, or because people are not paying that close attention to what browser they use.

"The browser that you're seeing right now is [browser]."

If the participant is confused because they normally use a different operating system, proceed with: "Even though we're using this on a different operating system than you might be used to, [browser] looks and behaves in the same way here as it does on other systems. We can open up all the same settings, menus, and so on." "Do you have any questions about that?"

If the participant thought the browser was something other than what it was: "What made you say that it was [guess]?"

Interactive Tasks

In this section we'll be talking about data practices, and go through tasks to identify and restrict them in the simulated browser running on the interviewer's computer. Before this happens, we will ask questions to gauge the participants' baseline knowledge and assumptions about these tasks and data practices generally.

"When you browse online, there's a good chance that you will encounter many different ways that websites can collect or use your data. I'm going to refer to this as 'data practices'. Data practices can vary depending on the website you are on, and can have both risks and benefits. What is most important to me is how you

personally feel about them. I'm going to ask you some questions about data practices now."

Now we are going to focus on the task of identifying what is actually present and perspectives on whether the practices are a concern. First we will start on assumptions about what is present, without talking about the specific practices we have identified. This way we can gauge the participants' initial knowledge about different practices and how to identify them without revealing anything.

"Let's assume for a moment that this website you are seeing is real and not just an example. Do you think that you would be encountering any data practices here?" "Can you give me some examples, and what would they do?" "Can you think of or see any indications about what data practices there might be on this website?" "Is there anything that the browser might be trying to tell you about that? How can you tell what's going on, if anything?"

Now we will go through the different practices individually. We will go through the task of identifying the individual practices (to the extent that it is possible in each browser). We will also identify the default settings. Then we will attempt to block the practices, or at least identify the settings used to do so.

"Now we're going to focus on a list of specific practices that I have, because I'm interested in hearing what you have to say about these specific ones. We might cover ones you have already mentioned but that's okay because we're going to go into a little bit more detail."

For each practice, in random order, repeat the following: 'Let's talk about [practice]. This is defined as [practice description without risks/benefits]." "How do you feel about [practice]? Have you heard of this before? What do you think about it?" "Let's say for a moment that we wanted to find out if [practice] is on this website. Let's try to find out together about whether it's present on this website." "Can you walk me through the steps?"

This may be a point at which some participants get stuck. If they don't seem to understand how to get to the settings or dashboards, give them a hint to point them in the right direction. If the browser has settings/dashboard (Edge, Safari, Firefox, Brave) and they don't mention the settings/dashboard in response to the previous questions, say the following: "Some browsers have settings and interfaces for this kind of thing. Have you used those settings before?" "Have a look at this button. Did you notice this here before? Have you used it? What do you think it's for?"

If the participant mentions installing/using an add-on: "Let's say for a moment we did not have any add-ons or tools." "Is it still possible to find out what we want

to know, and how might we go about that? Can you walk me through the steps?"

"Now that we've attempted to find out whether [practice] is here or not, what do you think about [practice]?" "Do you think that there might be certain risks or benefits associated with [practice]? What might they be?"

Once the participant answers the question, or the ensuing discussion finishes: "Now that we have had a chance to talk about your feelings about [practice], I wanted to mention that [list of risks and benefits for practice] are also potential risks and benefits. " "Do you feel any differently about [practice], now that we have discussed this?" "As things are right now, do you think that the browser is doing anything to allow or stop [practice] from happening? Why or why not?"

Once all the practices have been covered: "Alright, now we're going to start looking at ways we can take control of the different practices we just talked about." "Right now the browser is using only the default settings. We haven't changed any settings and the browser is set up exactly as it would be when it's first installed."

We wanted to avoid any settings being changed up until this point, so that we can collect people's perspectives on what the default settings are.

For each practice, in the same random order as before: "This website has [practice]." "How would you tell the browser whether you wanted to allow or block [practice]? Can you walk me through the steps for that?"

If the participant mentions installing/using an add-on: "How would this be done with the add-on or tool? Tell me more." "Let's say for a moment we did not have any add-ons or tools. Is it still possible to get what we want, and how might we go about that? Can you walk me through the steps?"

If the participant seems to have been successful, or thinks they were successful at identifying the controls: "How successful were we, now that we changed the settings?" "Do you think it's possible for [practice] to still happen?"

If the participant seems to have been unsuccessful, or thinks they were unsuccessful at identifying the controls: "We tried our best, but let's stop for now." "What do you think went wrong?"

"Now that we've tried to stop [practice from happening], let's take a moment to reflect." "Is [practice] concerning for you?" "Do you feel like you know enough about [practice] and what's going on online, in your browser, and so on that you have control?" "Any other thoughts?"

User Impressions

Towards the end of the interview, we want to create opportunities for the participant to tell us directly what they think about the browser, and tell us about issues or concerns that they may have previously identified or are thinking about now that they have explored the browser in more detail. We also want to uncover whether there are other factors which are relevant to the participants' thought process and perceptions.

"Now that we have gone through all the different tasks together, let's reflect a bit more on what we saw and did." "Do you think that this browser did a good job of telling you about the kinds of data practices you might be encountering when browsing?" "Why or why not? Which ones in particular?"

"Do you think that this browser offered you the right kind of controls so that you could block or avoid the things that you would want to avoid? Why or why not?"

"Do you think that this browser presented you with information in a way that was easy for you to understand? Why or why not?"

"What do you think would have made this browser better? Do you have any other suggestions for improvement?"

The interview ends here. Thank the participant and open the floor to questions and comments.

Appendix D

Coding Manuals

D.1 Examining Browser Privacy and Security Settings (Thematic Analysis)

This coding manual contains 38 codes in 4 general categories of codes which represent themes connected to our research questions, which we saw in the interview transcripts. This manual was used in the analysis of qualitative data seen in Chapter 3. "Understanding" refers to a category of codes which roughly correspond to RQ1, which is primarily about what users understand with respect to the interview tasks, definitions, and making connections. "Features" is a category based codes related to browser features, their usage, and issues encountered. "Expectations" is a category of responses related to users' approval/disapproval of practices, experiences, expectations, and whether these expectations seemed to be in alignment with reality. "Suggestions" is a category which reflects different ways users suggested that their browser might work better, offer better awareness, or offer more control. These suggestions were all related to the interview tasks. It is worth noting that codes were not mutually exclusive; many of the responses contained multiple layers of meanings and thus were assigned several codes simultaneously.

Understanding

RESIGNATION: The participant was resigned about their ability to understand, comprehend, recognize, or control a particular data practice. They just could not see any way to get what they wanted or to take control, and end up giving up (or

mention that they feel like giving up).

UNCLEAR_DEFINITION: The participant seemed to be confused about the terminology we gave them for a practice, and was unable to associate it with the terminology that was mentioned in their browser – this is specifically about the terminology, not the actual definition we gave them.

CONNECTION_DEFINITION: The participant seemed to make a valid connection between the definition of a practice that we gave them, language used by the browser and elsewhere to describe or identify it, a prior experience, or something else which made it clear that they understood the concept.

CONFIDENCE_LACKING: The participant expressed that they did not feel confident in their own ability to identify, understand, recognize, or allow/block a particular practice.

NOTICE_LACKING: The participant mentioned that there was no way for them to find out information about the practice they were looking for in their browser.

CONTROL_LACKING: The participant mentioned that they did not seem to have a way to control the practice that they were trying to control in their browser, which includes a way of knowing whether it was allowed or blocked (this is different from notice, which is only referring to knowing whether the practice is present).

ASSUME_CONTROL: The participant assumed that there was some way to control the practice they were trying to control in their browser, even if they were not able to find or identify the specific controls.

NOTICE_WEBSITE: The participant thought that the information they were looking for to tell if a practice was present is located on the website they are browsing, rather than in the browser.

UNCONCERNED: The participant felt unconcerned about a data practice, which may or may not have had an impact on their sense of feeling in control.

CONTROL_WEBSITE: The participant thought that the settings to control a practice they were looking for is located on the website they are browsing, rather than in the browser.

NOVELTY: The participant expressed that they had learned or experienced something new.

TOOL_ADDON: The participant made reference to an add-on for their browser or something similar, like searching on another website, anti-virus software or other third party tools that would be useful or necessary for them to perform a task.

HINT_ATTENTION: The interviewer directed the participant's attention towards something on the screen sharing session, or brought something up that had happened recently in the interview in order to prompt/probe the participant.

Features

UI_EFFECTIVE_NOTICE: The participant seemed to use the user interface in their browser to understand whether a practice was present or not, and was cognizant of this (rather than making a guess, or using a heuristic that was not based on evidence provided by the browser).

GUESS_NOTICE: The participant openly speculated, or relied on a guess to determine whether a practice was present or not, rather than using evidence provided by their browser UI.

UI_EFFECTIVE_CONTROL: The participant seemed to use the user interface in their browser to allow or block a practice, and was cognizant of this (rather than making a guess, or using a heuristic that was not based on evidence provided by the browser).

GUESS_CONTROL: The participant openly speculated, or relied on a guess to determine whether a practice was allowed or blocked, rather than using evidence provided by their browser UI.

EXPLORATION: The participant asked the interviewer to interact with one or more UI elements with the goal of trying to find out something, because they were unsure how to approach a task or because they were curious.

STRUGGLE: The participant seemed to be struggling; there was a significant pause (>2 sec) while they attempted to either find something, do something, or otherwise expressed frustration with the task at hand.

REQUIRED_PROMPT: The participant needed to be prompted by the interviewer in order to move forward with the task at hand, because they were stuck.

SAVVY: The participant seemed to be able to demonstrate that they had a good technical understanding of what is going on, based on prior knowledge or experience; the participant jumps ahead and/or correctly anticipates the answers to the interviewer's next questions, because they clearly understand what is going on.

UNSAVVY: The participant referenced their own lack of technical knowledge, experience, or understanding for their difficulty in engaging with a task.

CONFIDENCE: The participant expressed confidence with their reply; it was definitive, even if they were incorrect about what was actually going on, or if the answer to the questions they were asked were incorrect.

CHANGE_CONCERN: The participant expressed a change in their level of concern about a practice, as a result of the interview because of either the definition that was provided, or after having attempted to perform a task. The change can be positive or negative.

FAMILIAR_BROWSER: The participant was able to clearly recognize the browser or browser-specific feature that they were shown, and displayed some level of familiarity with it.

UNFAMILIAR_BROWSER: The participant was not able to clearly recognize the browser or browser-specific feature that they were shown, or admitted that they felt unfamiliar with it.

Expectations

EXPECT_MISMATCH_CONTROL: The participant seemed to have expectations about a control that were different compared to what the control actually did (e.g., the participant expects to use the controls to allow/block a practice in order to determine if it's there, or the participant just does not expect the controls to do what they actually do).

EXPECT_ALIGNED_CONTROL: The participant seemed to have a good understanding of how the controls worked, and their expectations for what the controls did were well aligned with reality.

EXPECT_MISMATCH_DEFAULT: The participant seemed to have expectations about the default settings that were different compared to what they actually are.

EXPECT_ALIGNED_DEFAULT: The participant seemed to have a good understanding of the default settings, and their expectations for how the settings were configured were well aligned with reality.

MISUNDERSTANDING_CONTROL: The participant did not seem to understand what kind of controls were available to them for addressing a specific practice.

MISUNDERSTANDING_DEFINITION: The participant did not seem to understand the definition we provided them for a practice, and either thought that

it meant something else, or was otherwise confused about what we meant.

EXPERIENCE: The participant described a prior experience (outside of the interview) with one or more data practices, in terms of being aware of them or controlling them.

APPROVAL_PRACTICE: The participant expressed approval towards a data practice; they saw it as helpful, or beneficial, or at least benign.

DISAPPROVAL_PRACTICE: The participant expressed disapproval towards a data practice; regardless of their level of concern about it, they saw it as bad, or harmful, or unethical, or evil, or malicious, or at least unnecessary.

Suggestions

IMPROVE_AWARENESS_CONTROL: The participant pointed out a way that their awareness of what is going on in the browser, or their options to control it, could be improved.

IMPROVE_AUTOMATION: The participant pointed out a way that their browser should automatically help them, make a decision for them, or perform an action without needing to be prompted.

IMPROVE_OPTIONS: The participant pointed out an option or setting that they believed should be present, but that they did not currently have, so they think it should be added.

D.2 Managing Online Data Practices (Grounded Analysis)

This coding manual contains 26 codes in 7 categories of codes which represent overarching themes seen in the responses. This manual was used in the analysis of qualitative data seen in Chapter 4. "Trends" refers to a category of codes which were generated in second-cycle coding, which had specific relevance to trends in responses seen after first-cycle coding. "Understanding" is a category based around questions concerning what practices participants seemed to understand, or misunderstand. "Bad Assumptions" is a category of responses created in second-cycle coding which was intended to identify specific assumptions that participants were making in their responses that were at times based on misunderstandings, lack of knowledge, or in-

correct perceptions. "Opposition" is a category which reflects attitudes, actions, and concerns participants expressed in opposition to intrusive practices. "Acceptance" highlights reasons, experiences, and expressions of ambivalence or ignorance towards practices which led to accepting them in certain circumstances. "Experience" is a category which highlights specific experiences, incidents, and their circumstances which participants shared, as well as expressions of lacking experience. Finally, "Miscellaneous" was a category with only one code, used to highlight responses which were selected for removal from the data set due to poor quality or survey abuse which was not detected by automated measures. It is worth noting that some codes were not mutually exclusive; many of the responses contained multiple layers of meanings and thus were assign several codes simultaneously.

Trends

SECURITY_THINKING: participant expresses evidence of thinking that is directly related to security, protection from security threats, protecting accounts and preventing fraud/scams (10 instances in 10 responses)

PROFILING_MENTIONS_ADS: participant explicitly seems to be making a connection between behavioral profiling and advertisements, targeted or otherwise (3 instances in 3 responses)

BREAKAGE: participant explicitly mentions parts of a website not functioning correctly (7 instances in 7 responses)

Understanding

UNDERSTANDING_DEMONSTRATES_KNOWLEDGE: participant expresses factual or operational knowledge of the technology and/or ramifications of their interactions with it (371 instances in 160 responses)

UNDERSTANDING_VAGUE: participant seems to express a vague or incomplete understanding of the technology or their interactions with it, such that it is difficult to gauge their level of understanding or expertise (123 instances in 83 responses)

UNDERSTANDING_MISCONCEPTION: participant seems to demonstrate a lack of knowledge about the technology and/or ramifications of their interactions with it, either by expressing factual inaccuracies, or other errors such as mixed-up terminology (110 instances in 74 responses)

Assumptions

bad_assumption: participant seems to be making an incorrect assumption (129 instances in 99 responses)

ADBLOCKER_EFFECTIVENESS: participant seems to be making a bad assumption, specifically about the effectiveness of ad blocking tools (21 instances in 19 responses)

INCOGNITO_MODE: participant seems to be making an assumption, specifically about the effectiveness of incognito mode/private browsing mode or similar features offered by private browsers, Tor, VPNs, general privacy extensions which are not ad-blockers, clearing cookies/history (58 instances in 51 responses)

ANTIVIRUS: participant seems to be making an assumption about the effectiveness of antivirus tools or firewalls in blocking privacy threats (3 instances in 3 responses)

HAS_CONTROL: participant seems to be making an assumption about having control over a setting which does not actually exist, or over a variable which they do not actually have control over (60 instances in 62 responses)

SAFE_BROWSING: participant seems to be making an assumption about being protected based on their own special browsing behavior, which makes them safe (15 instances in 13 responses)

MALWARE_RISK: participant seems to be making an assumption about the risk of being infected with malware (12 instances in 12 responses)

CONCERNED_ADS: participant is explicitly concerned with advertisements, either thinking that this is the way they can tell there is a problem, or that they are safe (16 instances in 16 responses)

UNCONCERNED: participant seems to be totally unconcerned with any privacy or security risk that may come about as a result of this practice (7 instances in 7 responses)

Opposition

OPPOSITION_ACTION: a specific action or mitigation strategy that a participant employs to oppose an intrusive practice (83 instances in 80 responses)

OPPOSITION_DISABLE_ATTEMPT: participants experiences with disabling/attempting to disable a practice (54 instances in 51 responses)

OPPOSITION_CONCERN: participants expressing a specific concern that they were attempting to address/mitigate (138 instances in 113 responses)

Acceptance

ACCEPTANCE_APPROVAL: reasons why participants seem to express explicit or tacit approval of a practice; they like it, and they don't believe that there are negatives/risks for them (50 instances in 44 responses)

ACCEPTANCE_AMBIVALENCE: reasons why participants seem to express explicit or tacit acceptance toward a practice; they recognize it is/might be bad or intrusive, but it does not bother them, will not get in the way, etc. (34 instances in 33 responses)

ACCEPTANCE_IGNORANCE: reasons where participants express ignorance about a practice and/or the ramifications of their interactions with it, suggesting that they are okay with the practice because they do not understand it or know enough about it to form an opinion (108 instances in 107 responses)

Experience

EXPERIENCE_POSITIVE: participants express a positive experience when interacting with a practice, including acknowledging that they received a benefit (9 instances in 8 responses)

EXPERIENCE_NEGATIVE: participants express a negative experience when interacting with a practice, including fears of repercussions, "creep factor" and other concerns or harms that they directly or indirectly experienced (29 instances in 25 responses)

EXPERIENCE_NEUTRAL: participants express some form of experience with interacting with a practice, but without obvious or apparent risks or benefits; they just acknowledge that there was some kind of experience without making a judgment about it (65 instances in 65 responses)

EXPERIENCE_LACKING: participants expressing a lack of experience and/or ignorance about whether they actually had an experience with a practice (9 instances in 8 responses)

Appendix E

Additional Results

This appendix contains regression tables referenced in Chapters 3, 4 and 5.

Table E.1: This table shows the χ^2 test results for the factors which we tested to determine if they influenced our participants' smartphone app permissions seen in Chapter 5. Factors with very strong significance ($pr(\chi^2) \leq 0.01$) are marked with **. Factors with strong significance ($pr(\chi^2) \leq 0.05$) are marked with *. Factors found to not be statistically significant are marked ns.

Factors	df	Calendar		Location		Contacts		Calendar x Internal		Calendar x Ads		Calendar x Other	
		χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
App Familiarity	4	312.50	**	341.23	**	308.76	**	238.26	**	222.32	**	169.75	**
App Usage Frequency	5	450.22	**	413.52	**	394.26	**	262.22	**	350.66	**	248.20	**
App Category	24	82.75	**	138.02	**	76.29	**	74.08	**	56.92	**	0.00	ns
Age	3	27.96	**	6.46	ns	29.98	**	11.07	*	27.60	**	6.19	ns
Education	7	24.61	**	42.51	**	29.99	**	19.97	**	37.92	**	0.00	ns
Gender	2	1.05	ns	1.33	ns	0.40	ns	0.26	ns	0.93	ns	0.84	ns
City Size	3	31.30	**	17.00	**	27.85	**	28.01	**	38.87	**	8.40	*
Marital Status	5	40.96	**	46.00	**	59.94	**	26.02	**	65.40	**	19.84	**
Employment	8	13.67	ns	12.36	ns	15.51	ns	6.93	ns	23.74	ns	0.00	ns
Phone Usage Frequency	3	12.74	**	5.53	ns	11.17	*	9.12	*	9.02	*	6.19	ns
Phone Usage Duration	3	30.64	ns	15.14	**	30.35	ns	15.52	ns	24.36	ns	10.59	ns
Number of Apps Installed	3	17.78	**	20.77	**	22.76	**	10.31	*	9.80	*	2.40	ns
Number of Apps Used	3	70.02	**	50.00	**	87.79	**	26.24	**	96.42	**	43.32	**
Android Version	9	40.41	ns	35.40	ns	34.19	ns	37.19	ns	35.28	ns	8.01	ns
Number of Privacy Surveys	3	5.81	ns	12.44	**	20.34	**	7.71	ns	20.69	**	6.54	ns

Factors	df	Location x Internal		Location x Ads		Location x Other		Contacts x Internal		Contacts x Ads		Contacts x Other	
		χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
App Familiarity	4	238.80	**	271.31	**	232.24	**	265.09	**	216.17	**	186.84	**
App Usage Frequency	5	246.27	**	415.82	**	336.80	**	297.34	**	372.51	**	238.93	**
App Category	24	127.62	**	86.11	**	53.21	**	74.40	**	18.73	ns	0.00	ns
Age	3	2.09	ns	13.71	**	5.87	ns	7.91	*	8.77	ns	4.65	ns
Education	7	25.15	**	39.57	**	28.78	**	24.87	**	0.00	ns	4.87	ns
Gender	2	0.75	ns	1.42	ns	5.17	ns	2.43	ns	0.40	ns	0.48	ns
City Size	3	15.65	**	31.44	**	12.22	**	25.16	**	19.75	**	6.31	ns
Marital Status	5	23.72	**	57.66	**	36.61	**	38.65	**	36.34	**	19.45	**
Employment	8	7.36	ns	19.53	ns	0.00	ns	6.93	ns	0.00	ns	0.00	ns
Phone Usage Frequency	3	8.75	*	9.83	*	7.93	*	5.76	ns	6.64	ns	6.09	ns
Phone Usage Duration	3	8.01	ns	23.23	**	13.11	**	20.37	**	18.49	**	7.96	ns
Number of Apps Installed	3	14.08	**	11.65	**	5.59	ns	8.71	ns	1.74	ns	1.76	ns
Number of Apps Used	3	16.35	**	91.59	**	81.58	**	35.21	**	60.96	**	37.51	**
Android Version	9	26.00	ns	54.68	ns	26.13	**	24.41	ns	16.09	ns	0.00	ns
Number of Privacy Surveys	3	14.18	**	26.16	**	11.27	*	8.31	*	11.47	**	6.91	ns

Table E.2: Z-Test odds-ratios for opt-out likelihood, for all PIPs studied in Chapter 4 (with respect to the intercept). Levels with no data points are marked with ∅. Levels which did not converge are marked NC. Factors with statistically significant p-values are darkened. Intercept: AgeRange [18-24], EducationLevel [Associates], CitySize [City], MaritalStatus [Divorced], EmploymentStatus [Employed], EmploymentField [Non-STEM], PrivacySettings_LastLooked [Past month], PrivacySettings_LastChanged [Past month], Browser [Chrome], PrivacySurveys_PastYear [6-9], Privacy_AtRisk [FALSE], website_category [ADULT].

	Behav. Profiling (n=113)		**Reporting (n=113)**		**Session Replay (n=99)**		**Targeted Ads (n=103)**	
Factors	*Odds Ratios*	*p*	*Odds Ratios*	*p*	*Odds Ratios*	*p*	*Odds Ratios*	*p*
(Intercept)	22.86	0.009	2.39	0.481	43.54	0.043	1.25	0.816
AgeRange [25-44]	1.09	0.887	0.83	0.834	2.32	0.567	3.8	0.016
AgeRange [45-64]	1.56	0.505	3.53	0.169	1.93	0.660	4.41	0.021
AgeRange [65+]	5.34	0.096	4.8	0.339	NC	0.648	7.87	0.045
EducationLevel [Bachelors]	0.38	0.043	1.79	0.258	0.55	0.427	0.86	0.659
EducationLevel [PhD]	∅	∅	19.53	0.019	0.14	0.319	1.61	0.770
EducationLevel [High School]	0.41	0.150	2.29	0.192	0.35	0.250	0.62	0.333
EducationLevel [<High School]	∅	∅	∅	∅	0.12	0.295	∅	∅
EducationLevel [Masters]	0.30	0.053	1.45	0.540	0.34	0.261	1.2	0.782
EducationLevel [JD, MD]	1.01	0.995	NC	0.950	1.18	0.944	5.1	0.249
EducationLevel [Some College]	0.56	0.303	2.58	0.080	1.03	0.976	1.05	0.920
EducationField [STEM]	0.72	0.396	4.09	0.007	0.71	0.558	0.45	0.047
CitySize [Large City]	2.84	0.018	2.08	0.114	1.29	0.664	1.24	0.564
CitySize [Rural Area]	1.81	0.199	1.5	0.435	1.27	0.688	0.84	0.735
CitySize [Town or Suburb]	2.35	0.022	1.05	0.910	3.93	0.006	1.18	0.661
MaritalStatus [Married]	1.45	0.599	0.88	0.856	0.28	0.109	3.17	0.059
MaritalStatus [Never married]	0.85	0.829	1.44	0.612	0.37	0.206	1.72	0.363
MaritalStatus [Prefer not to disclose]	0.16	0.242	0.74	0.851	0.35	0.452	93.95	0.527
MaritalStatus [Separated]	3.14	0.304	1.42	0.824	4.16	0.423	20.74	0.002
MaritalStatus [Widowed]	2.22	0.611	3.85	0.503	0.13	0.114	∅	∅
EmploymentStatus [Student]	∅	∅	0.2	0.026	0.04	0.163	1.85	0.550
EmploymentStatus [Unemployed]	2.32	0.132	0.16	0.003	0.44	0.451	0.64	0.388
EmploymentStatus [Prefer not to answer]	1.19	0.920	4.01	0.419	∅	∅	∅	∅
EmploymentField [STEM]	0.79	0.538	0.33	0.031	1.1	0.869	1.46	0.349
PrivacySettings_LastLooked [Past week]	3.20	0.015	1	0.992	2.58	0.126	2.09	0.033
PrivacySettings_LastLooked [Past year]	1.07	0.872	1.13	0.797	1.21	0.719	2.64	0.042
PrivacySettings_LastLooked [Never]	4.92	0.084	99.44	0.007	0.55	0.572	0.99	0.994
PrivacySettings_LastChanged [Past week]	0.09	<0.001	1.94	0.238	0.88	0.842	1.3	0.562
PrivacySettings_LastChanged [Past year]	0.31	0.004	1.11	0.797	1.3	0.580	1.11	0.758
PrivacySettings_LastChanged [Never]	0.05	<0.001	0.67	0.526	1.49	0.721	1.93	0.597
Browser [Edge]	∅	∅	0.87	0.932	∅	∅	∅	∅
Browser [Firefox]	1.92	0.118	1.94	0.114	0.47	0.182	3.75	0.011
Browser [IE]	5.19	0.262	0.03	0.020	∅	∅	2.25	0.476
Browser [Other]	4.37	0.224	15.99	0.007	NC	0.949	2.36	0.297
Browser [Safari]	1.32	0.805	1.22	0.807	0.02	0.024	10	0.011
PrivacySurveys_PastYear [<5]	0.68	0.337	1.2	0.704	0.4	0.135	1.33	0.450
PrivacySurveys_PastYear [>10]	1.91	0.414	2.2	0.268	0.72	0.685	2.32	0.199
PrivacySurveys_PastYear [0]	0.71	0.429	0.97	0.941	0.33	0.093	0.35	0.012
Privacy_AtRisk [TRUE]	2.66	0.028	3.05	0.007	3.22	0.016	1.32	0.484
website_category [FINANCE]	0.60	0.082	0.29	<0.001	0.69	0.298	0.34	0.001
website_category [GAMES]	0.14	<0.001	0.17	<0.001	0.12	<0.001	0.12	<0.001
website_category [HEALTH]	0.20	<0.001	0.11	<0.001	0.12	<0.001	0.1	<0.001
website_category [NEWS]	0.27	<0.001	0.22	<0.001	0.1	<0.001	0.3	<0.001
website_category [SHOPPING]	0.15	<0.001	0.13	<0.001	0.23	<0.001	0.08	<0.001
website_category [SOCIAL]	1.05	0.877	0.64	0.135	0.88	0.720	0.53	0.05
website_category [TRAVEL]	0.46	0.008	0.39	0.001	0.53	0.065	0.28	<0.001

	Crypto-Mining (n=102)		**Identity Sign-In (n=133)**		**Fingerprinting (n=121)**		**Nag Screens (n=104)**	
Factors	*Odds Ratios*	*p*	*Odds Ratios*	*p*	*Odds Ratios*	*p*	*Odds Ratios*	*p*
(Intercept)	1423.54	0.012	642.25	<0.001	27.39	0.001	141.75	<0.001
AgeRange [25-44]	0.32	0.337	0.13	0.009	0.57	0.299	0.48	0.303
AgeRange [45-64]	0.44	0.494	0.11	0.009	0.56	0.331	0.45	0.280
AgeRange [65+]	4.38	0.343	0.09	0.063	1.6	0.670	0.27	0.205
EducationLevel [Bachelors]	1.2	0.721	1.38	0.488	1.7	0.182	0.51	0.157
EducationLevel [PhD]	0.08	0.043	1.99	0.613	1.55	0.648	∅	∅
EducationLevel [High School]	0.3	0.036	1.04	0.953	0.64	0.398	0.26	0.041
EducationLevel [<High School]	∅	∅	NC	0.421	∅	∅	∅	∅
EducationLevel [Masters]	2.51	0.159	1.79	0.310	5.77	0.002	1.04	0.958
EducationLevel [JD, MD]	0.17	0.047	7.84	0.092	1.94	0.544	2.95	0.554
EducationLevel [Some College]	0.52	0.236	0.98	0.973	3.01	0.020	0.57	0.348
EducationField [STEM]	0.57	0.152	1.68	0.188	2.47	0.118	0.71	0.517
CitySize [Large City]	0.65	0.332	0.69	0.342	0.47	0.049	0.62	0.310
CitySize [Rural Area]	3.67	0.032	0.78	0.590	0.43	0.110	0.39	0.091
CitySize [Town or Suburb]	0.36	0.003	0.89	0.722	0.97	0.926	0.54	0.146
MaritalStatus [Married]	0.01	0.096	0.41	0.115	0.27	0.011	0.6	0.393
MaritalStatus [Never married]	0.02	0.129	0.22	0.013	0.41	0.094	0.66	0.497
MaritalStatus [Prefer not to disclose]	0	0.019	5.01	0.369	0.54	0.519	NC	0.989
MaritalStatus [Separated]	0.03	0.244	∅	∅	∅	∅	746.3	0.530
MaritalStatus [Widowed]	0.2	0.612	0.03	0.063	∅	∅	1.36	0.801
EmploymentStatus [Student]	83.24	0.007	0.21	0.180	13.42	0.041	3.37	0.200
EmploymentStatus [Unemployed]	0.83	0.729	1.65	0.225	1.28	0.686	1.47	0.478
EmploymentStatus [Prefer not to answer]	∅	∅	1.12	0.933	2.62	0.361	0.15	0.214
EmploymentField [STEM]	4.93	<0.001	1.08	0.830	0.5	0.237	1.18	0.763
PrivacySettings_LastLooked [Past week]	0.25	0.002	1.87	0.127	1.42	0.309	0.34	0.015
PrivacySettings_LastLooked [Past year]	0.69	0.400	2.03	0.086	0.79	0.516	0.71	0.479
PrivacySettings_LastLooked [Never]	1.89	0.643	3.36	0.145	0.22	0.172	0.03	0.014
PrivacySettings_LastChanged [Past week]	1.33	0.554	0.51	0.165	0.63	0.301	1.75	0.306
PrivacySettings_LastChanged [Past year]	0.93	0.853	0.38	0.010	0.62	0.195	0.71	0.393
PrivacySettings_LastChanged [Never]	0.04	0.006	0.07	<0.001	1.31	0.757	3.69	0.282
Browser [Edge]	8.93	0.021	0.88	0.911	0.27	0.393	∅	∅
Browser [Firefox]	2.83	0.013	1.83	0.178	1.78	0.182	2.85	0.022
Browser [IE]	∅	∅	27.42	0.080	3.79	0.257	13.27	0.159
Browser [Other]	0	0.444	0.22	0.018	0.25	0.387	7.75	0.195
Browser [Safari]	NC	0.502	0.54	0.719	0.22	0.076	0.48	0.460
PrivacySurveys_PastYear [<5]	1.14	0.782	0.39	0.020	1.43	0.463	0.6	0.297
PrivacySurveys_PastYear [>10]	2.06	0.237	0.99	0.987	1.74	0.409	2.35	0.149
PrivacySurveys_PastYear [0]	2.27	0.124	0.46	0.067	0.91	0.857	0.61	0.344
Privacy_AtRisk [TRUE]	2.35	0.029	0.79	0.586	1.53	0.358	0.84	0.693
website_category [FINANCE]	1.42	0.307	0.25	<0.001	0.09	<0.001	0.21	<0.001
website_category [GAMES]	0.9	0.740	0.16	<0.001	0.14	<0.001	0.56	0.036
website_category [HEALTH]	0.33	<0.001	0.19	<0.001	0.13	<0.001	0.22	<0.001
website_category [NEWS]	0.5	0.029	0.16	<0.001	0.32	<0.001	0.63	0.096
website_category [SHOPPING]	0.81	0.513	0.16	<0.001	0.16	<0.001	0.28	<0.001
website_category [SOCIAL]	1.84	0.083	0.23	<0.001	0.39	<0.001	1.05	0.876
website_category [TRAVEL]	1.33	0.397	0.34	<0.001	0.28	<0.001	0.5	0.012

Table E.3: Matrix showing the results of contextual interview tasks seen in Chapter 3. In this table, the following are symbolized: ✅ indicates success, ❌ indicates failure, 😒❌ indicates that the participant assumed they were successful but failed, 😒✅ indicates that the participant assumed they had failed but succeeded, 🤷 indicates unclear or unsure, 😀 indicates approval or feeling in control, 😡 indicates disapproval or feeling not in control, 😎 indicates overconfidence.

		Identity/Sign-In Services					
ID	Browser	Understood Definition	Approved/Disapproved	Understood Defaults	Determined Presence	Took Control	Felt In Control
P1	Brave	✅	😡	✅	✅	✅	😀
P2	Brave	✅	😡	❌	❌	✅	😀
P3	Brave	✅	🤷	✅	❌	✅	😀
P4	Brave	😒❌	🤷	✅	✅	😒✅	😡
P5	Brave	✅	😡	❌	❌	✅	😡
P6	Chrome	✅	😀	✅	❌	❌	😡
P7	Chrome	✅	😀	✅	❌	❌	😡
P8	Chrome	✅	😡	✅	❌	😒❌	😡
P9	Chrome	😒❌	🤷	✅	✅	😒❌	😎
P10	Chrome	✅	😀	❌	❌	✅	😎
P11	Edge	✅	🤷	✅	❌	😒❌	🤷
P12	Edge	✅	🤷	✅	❌	❌	😡
P13	Edge	✅	😡	❌	❌	😒❌	😎
P14	Edge	✅	😡	❌	❌	❌	😡
P15	Edge	✅	😀	❌	❌	😒❌	😎
P16	Firefox	✅	😀	❌	❌	😒❌	😡
P17	Firefox	✅	😡	✅	✅	❌	😡
P18	Firefox	✅	😀	❌	❌	✅	😀
P19	Firefox	✅	😡	✅	✅	😒❌	🤷
P20	Firefox	😒❌	😡	❌	❌	❌	😡
P21	Safari	✅	😀	🤷	❌	❌	😎
P22	Safari	✅	😡	🤷	😒✅	😒❌	😎
P23	Safari	❌	🤷	🤷	❌	❌	🤷
P24	Safari	✅	😀	✅	❌	❌	😎
P25	Safari	✅	😡	❌	❌	❌	😡

Behavioral Profiling

ID	Browser	Understood Definition	Approved/Disapproved	Understood Defaults	Determined Presence	Took Control	Felt In Control
P1	Brave	😒❌	😡	✅	❌	✅	😀
P2	Brave	✅	😀	❌	😒✅	✅	😡
P3	Brave	✅	😡	✅	✅	✅	😀
P4	Brave	✅	🤦	✅	❌	✅	😀
P5	Brave	✅	😡	✅	❌	❌	😀
P6	Chrome	✅	😡	🤦	❌	❌	😡
P7	Chrome	✅	😡	✅	❌	❌	🤦
P8	Chrome	✅	😡	✅	❌	😒❌	😡
P9	Chrome	✅	😡	✅	✅	❌	🤦
P10	Chrome	✅	😡	✅	✅	✅	😡
P11	Edge	✅	😡	✅	✅	✅	😀
P12	Edge	✅	😡	✅	✅	😒✅	😡
P13	Edge	❌	😡	❌	❌	❌	😡
P14	Edge	✅	🤦	✅	✅	✅	😀
P15	Edge	✅	🤦	❌	❌	✅	🤦
P16	Firefox	✅	😡	❌	😒❌	😒❌	😡
P17	Firefox	❌	😡	✅	✅	✅	🤦
P18	Firefox	✅	😡	🤦	✅	✅	😀
P19	Firefox	✅	😡	✅	😒❌	😒✅	🤦
P20	Firefox	✅	😡	✅	✅	✅	😡
P21	Safari	✅	😡	✅	❌	😒❌	🤦
P22	Safari	✅	😡	✅	❌	😒❌	😎
P23	Safari	😒❌	😡	❌	❌	❌	😡
P24	Safari	✅	😡	✅	✅	✅	😡
P25	Safari	✅	😡	✅	😒❌	✅	😡

Targeted Advertising

ID	Browser	Understood Definition	Approved/Disapproved	Understood Defaults	Determined Presence	Took Control	Felt In Control
P1	Brave	✅	😡	✅	✅	✅	😀
P2	Brave	✅	😡	✅	❌	😒✅	😡
P3	Brave	✅	😡	✅	✅	✅	😀
P4	Brave	✅	😡	✅	✅	✅	😀
P5	Brave	✅	😡	✅	❌	✅	🤦
P6	Chrome	✅	😡	✅	❌	❌	😡
P7	Chrome	✅	😡	🤦	❌	❌	😡
P8	Chrome	✅	😡	✅	✅	😒❌	😡
P9	Chrome	✅	🤦	✅	✅	❌	😎
P10	Chrome	✅	🤦	✅	❌	✅	😡
P11	Edge	✅	😡	✅	✅	✅	😀
P12	Edge	✅	🤦	✅	😒✅	😒✅	😡
P13	Edge	😒❌	😡	✅	✅	✅	😎
P14	Edge	✅	😡	✅	✅	✅	😀
P15	Edge	✅	😡	✅	😒✅	❌	😡
P16	Firefox	✅	😡	❌	✅	✅	😡
P17	Firefox	✅	😡	🤦	❌	✅	😀
P18	Firefox	✅	😡	🤦	😒✅	✅	😀
P19	Firefox	✅	🤦	✅	😒❌	✅	🤦
P20	Firefox	✅	😡	✅	✅	✅	😀
P21	Safari	✅	😡	❌	❌	😒❌	😎
P22	Safari	✅	😡	✅	😒✅	😒❌	😎
P23	Safari	✅	😡	🤦	❌	😒❌	😎
P24	Safari	✅	😡	✅	✅	✅	😎
P25	Safari	✅	😡	✅	✅	❌	😡

Fingerprinting

ID	Browser	Understood Definition	Approved/Disapproved	Understood Defaults	Determined Presence	Took Control	Felt In Control
P1	Brave	✅	🤬	✅	✅	✅	😀
P2	Brave	❌	🤬	✅	😒❌	😒❌	🤬
P3	Brave	😒✅	🤦	✅	✅	✅	😀
P4	Brave	✅	🤦	✅	✅	✅	😀
P5	Brave	✅	🤬	✅	✅	✅	🤬
P6	Chrome	😒❌	🤦	🤦	❌	❌	🤬
P7	Chrome	❌	🤦	✅	❌	❌	🤬
P8	Chrome	✅	🤬	🤦	❌	❌	🤬
P9	Chrome	✅	😀	✅	😒❌	❌	🤦
P10	Chrome	✅	🤬	✅	✅	✅	😀
P11	Edge	✅	🤬	🤦	❌	❌	🤬
P12	Edge	😒❌	🤬	✅	❌	❌	🤬
P13	Edge	❌	🤬	❌	❌	😒❌	🤬
P14	Edge	✅	🤬	❌	❌	😒❌	😎
P15	Edge	❌	🤬	✅	❌	❌	🤬
P16	Firefox	✅	🤬	✅	✅	✅	😀
P17	Firefox	❌	🤬	✅	✅	✅	😀
P18	Firefox	❌	😀	✅	✅	✅	😀
P19	Firefox	😒✅	🤬	✅	✅	😒✅	🤬
P20	Firefox	✅	😀	✅	✅	✅	😀
P21	Safari	😒❌	🤦	❌	❌	❌	🤦
P22	Safari	❌	🤦	🤦	❌	❌	🤬
P23	Safari	❌	🤦	🤦	❌	❌	🤦
P24	Safari	❌	🤬	✅	❌	😒❌	🤬
P25	Safari	✅	🤬	❌	❌	❌	🤬

Crypto-Mining

ID	Browser	Understood Definition	Approved/Disapproved	Understood Defaults	Determined Presence	Took Control	Felt In Control
P1	Brave	✅	🤦	✅	❌	✅	😀
P2	Brave	✅	🤦	❌	❌	❌	🤬
P3	Brave	✅	🤦	✅	❌	✅	😀
P4	Brave	✅	🤬	✅	✅	✅	😀
P5	Brave	❌	🤬	✅	❌	❌	🤬
P6	Chrome	❌	🤦	🤦	❌	😒❌	🤬
P7	Chrome	❌	🤦	✅	❌	❌	🤬
P8	Chrome	❌	🤬	✅	❌	😒❌	🤬
P9	Chrome	✅	🤦	❌	😒✅	❌	🤬
P10	Chrome	✅	🤬	🤦	✅	😒✅	🤬
P11	Edge	✅	🤬	✅	✅	✅	😀
P12	Edge	✅	🤬	❌	❌	❌	🤬
P13	Edge	❌	🤬	❌	✅	✅	🤦
P14	Edge	✅	🤦	✅	✅	✅	😀
P15	Edge	❌	🤦	❌	❌	❌	🤬
P16	Firefox	✅	🤬	✅	✅	✅	🤦
P17	Firefox	✅	🤬	✅	✅	✅	😀
P18	Firefox	✅	😀	✅	✅	✅	😀
P19	Firefox	✅	🤬	✅	✅	✅	😀
P20	Firefox	✅	🤬	✅	✅	✅	😀
P21	Safari	✅	🤬	❌	✅	😒❌	🤦
P22	Safari	✅	🤬	✅	❌	❌	🤬
P23	Safari	❌	🤦	🤦	❌	❌	🤦
P24	Safari	❌	🤬	✅	❌	❌	🤬
P25	Safari	✅	🤬	✅	❌	❌	🤬

Bibliography

[1] Ruba Abu-Salma. *Designing User-Centered Privacy-Enhancing Technologies*. PhD thesis, UCL (University College London), 2020. 2.2, 2.3

[2] Ruba Abu-Salma and B. Livshits. Evaluating the end-user experience of private browsing mode. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2020, 2020. 2.2, 3.3.2, 4.3

[3] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, et al. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys*, 50(3):1–41, 2017. 1.3, 3, 2.3, 3.1, 6.4.1, 6.4.3

[4] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. In *IEEE Security and Privacy*, S&P 2005, pages 26–33. IEEE, Jan 2005. 1.3

[5] Alessandro Acquisti, Curtis Taylor, and Liad Wagman. The economics of privacy. *Journal of Economic Literature*, 54(2):442–92, June 2016. 1.2, 1.3, 1.4, 3, 4, 4.1, 6.4.1

[6] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass: Re-examining user concerns for online tracking and advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS 2013, New York, NY, USA, 2013. ACM. 1.2, 3.1, 4.1

[7] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *22nd USENIX Security Symposium*, USENIX Security 2013, pages 257–272, 2013. 2.2

[8] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj 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 2015, pages 787–796, New York, NY, USA, 2015. ACM. 1, 2.4, 3.1, 3.4, 5

[9] Ali Alshehri, Pawel Marcinek, Abdulrahman Alzahrani, Hani Alshahrani, and Huirong Fu. Puredroid: Permission usage and risk estimation for android applications. In *Proceedings of the 3rd International Conference on Information System and Data Mining*, pages 179–184, 2019. 2.4

[10] Amazon. Alexa top sites. `https://www.alexa.com/topsites`, 2020. 4.2, 6.4.1

[11] Reyhan Aydoğan, Pinar Øzturk, and Yousef Razeghi. Negotiation for incentive driven privacy-preserving information sharing. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 486–494. Springer, 2017. 2.4

[12] Tim Baarslag, Alan Alper, Richard Gomer, Muddasser Alam, Perera Charith, Enrico Gerding, and m.c. schraefel. An automated negotiation agent for permission management. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, AAMAS 2017, pages 380–390. ACM, May 2017. 2.4

[13] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, CCS 2015, pages 63–74. ACM, 2015. 2.1

[14] Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, Florian Schaub, and Norman Sadeh. Finding a choice in a haystack: Automatic extraction of opt-out statements from privacy policy text. In *Proceedings of The Web Conference*, WWW 2020, page 1943–1954, New York, NY, USA, 2020. ACM. 2.2, 4.1, 6.4.4

[15] Adam Barth, Anupam Datta, John C Mitchell, and Helen Nissenbaum. Privacy and contextual integrity: Framework and applications. In *IEEE Symposium*

on Security and Privacy, S&P 2006, pages 15–pp. IEEE, 2006. 1, 1.3, 2.4, 5, 6.4.4

[16] David Basin, Søren Debois, and Thomas Hildebrandt. On purpose and by necessity: compliance under the GDPR. *Proceedings of Financial Cryptography and Data Security*, 18, 2018. 6.4.4

[17] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. 4.2.3, 5.2.2

[18] Michael Benisch, Patrick Gage Kelley, Norman Sadeh, and Lorrie Faith Cranor. Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. *Personal Ubiquitous Computing*, 15(7):679–694, October 2011. 1.3, 1, 2.2, 2.4

[19] Christoph Bier, Kay Kühne, and Jürgen Beyerer. Privacyinsight: the next generation privacy dashboard. In *Annual Privacy Forum*, pages 135–152. Springer, 2016. 2.2

[20] Michael Bloor and Fiona Wood. Purposive sampling. In *Keywords in qualitative methods*, pages 143–144. SAGE Publications, 2006. 3.2.1

[21] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, 4(3):340–347, 2013. 1.4

[22] Travis Breaux, Lujo Bauer, Lorrie Faith Cranor, Simson Garfinkel, and David Gordon. *An Introduction to Privacy for Technology Professionals*. International Association of Privacy Professionals, 2020. 2.1

[23] Johana Cabinakova, Christian Zimmermann, and Guenter Mueller. An empirical analysis of privacy dashboard acceptance: the google case. *Proceedings of the 2016 European Conference on Information Systems*, 2016. 2.2

[24] Yanto Chandra and Liang Shang. An RQDA-based constructivist methodology for qualitative research. *Qualitative Market Research: An International Journal*, 2017. 3.2.4

[25] Hongliang Chen, Christopher E. Beaudoin, and Traci Hong. Securing online privacy: An empirical test on internet scam victimization, online privacy concerns, and privacy protection behaviors. *Computers in Human Behavior*, 70:291 – 302, 2017. 1.2, 3.1, 4.1

[26] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Informing the design of a personalized privacy assistant for the internet of things. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2020, pages 1–13, 2020. 2.1, 6.4.3

[27] Lorrie Cranor and Rigo Wenning. Platform for privacy preferences (P3P) project. `https://www.w3.org/P3P/`, Feb 2018. 3.1

[28] Lorrie Faith Cranor. What do they "indicate?" evaluating security and privacy indicators. *Interactions*, 13(3):45–47, 2006. 2.2, 6.4.4

[29] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, UPSEC 2008, USA, 2008. USENIX Association. 1, 1.4, 2.1

[30] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *JTHTL*, 10:273–308, 2012. 1.2, 1.3, 1.4, 2.1, 3.1

[31] Lorrie Faith Cranor. Informing california privacy regulations with evidence from research. *Communications of the ACM*, 64(3):29–32, February 2021. 6.4.4

[32] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Pervasive Computing*, 17(3):35–46, 2018. 6.3, 6.4.3, 6.4.4

[33] Willem De Groef, Dominique Devriese, Nick Nikiforakis, and Frank Piessens. Flowfox: A web browser with flexible and precise information flow control. In *Proceedings of the ACM Conference on Computer and Communications Security*, CCS 2012, page 748–759. ACM, 2012. 2.2

[34] Nicolás E Díaz Ferreyra, Tobias Kroll, Esma Aïmeur, Stefan Stieglitz, and Maritta Heisel. Preventative nudges: Introducing risk cues for supporting online self-disclosure decisions. *Information*, 11(8):399, 2020. 2.3

[35] Joseph Dickinson. Tracking changes in browser security indicators. In *The best of ECE undergraduate research*. Illinois Digital Environment for Access to Learning and Scholarship, 2018. 2.2

[36] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the experts: What should be on an IoT privacy and security label? In

IEEE Symposium on Security and Privacy, S&P 2020, pages 447–464. IEEE, 2020. 2.3

[37] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the ACM Conference on Computer and Communications Security*, CCS 2016, page 1388–1401, New York, NY, USA, 2016. ACM. 4.1

[38] Facebook. When I post something on Facebook, how do I choose who can see it? `https://www.facebook.com/help/120939471321735`, 2021. 3.4, 4.1

[39] Cori Faklaris, Laura Dabbish, and Jason I. Hong. A self-report measure of end-user security attitudes (SA-6). In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*, SOUPS 2019, page 61–77, USA, 2019. USENIX Association. 4.2.1, 4.5

[40] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th International Conference on the World Wide Web*, WWW 2010, pages 351–360, New York, NY, USA, 2010. ACM. 2.2

[41] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2):175–191, May 2007. 5.2.2

[42] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. A design space for privacy choices: Towards meaningful privacy control in the internet of things. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2021, page 1–15, New York, NY, USA, 2021. ACM. 2.1, 3.4, 6.4.3, 6.4.4

[43] Denis Feth, Andreas Maier, and Svenja Polst. A user-centered model for usable security and privacy. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 74–89. Springer, 2017. 2.3

[44] Barney G Glaser and Anselm L Strauss. *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017. 2.1, 4.2.1

[45] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj 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, 2016. 2.2

[46] Google. Android permissions overview. https://developer.android.com/guide/topics/permissions/overview, Jan 2019. Accessed: 2019-02-24. 5.2.1

[47] Google. Choose your privacy settings. https://support.google.com/chrome/answer/114836, 2021. 3.1, 4.2.4

[48] Peiqing Guan and Wei Zhou. Business analytics generated data brokerage: Law, ethical and social issues. In Robin Doss, Selwyn Piramuthu, and Wei Zhou, editors, *Future Network Systems and Security*, pages 167–175. Springer International Publishing, 2017. 4.1

[49] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. "it's a scavenger hunt": Usability of websites' opt-out and data deletion choices. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2020, pages 1–12, 2020. 2.1

[50] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. An empirical analysis of data deletion and opt-out choices on 150 websites. In *Fifteenth Symposium on Usable Privacy and Security*, SOUPS 2019, pages 387–406, 2019. 2.1

[51] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. Toggles, dollar signs, and triangles: How to (in) effectively convey privacy choices with icons and link texts. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2021, pages 1–25, 2021. 6.4.4

[52] Margaret Hagen. User-centered privacy communication design. In *Twelfth Symposium on Usable Privacy and Security*, SOUPS 2016. USENIX Association, 2016. 2.3

[53] Karen Holtzblatt and Sandra Jones. Conducting and analyzing a contextual interview (excerpt). In *Readings in Human–Computer Interaction*, pages 241–253. Elsevier, 1995. 3.2

[54] Corey Brian Jackson and Yang Wang. Addressing the privacy paradox through personalized privacy notifications. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–25, 2018. 2.3

[55] Carlos Jensen, Colin Potts, and Christian Jensen. Privacy practices of internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, 2005. 2.2

[56] Laurie Kantner, Deborah Hinderer Sova, and Stephanie Rosenbaum. Alternative methods for field usability research. In *Proceedings of the 21st Annual International Conference on Documentation*, SIGDOC 2003, page 68–72, New York, NY, USA, 2003. ACM. 2.1, 3.2

[57] Soroush Karami, Panagiotis Ilia, Konstantinos Solomos, and Jason Polakis. Carnus: Exploring the privacy threats of browser extension fingerprinting. In *Proceedings of the Symposium on Network and Distributed System Security*, NDSS 2020, 2020. 2.2, 3.1, 4.1

[58] Mark J Keith, Courtenay Maynes, Paul Benjamin Lowry, and Jeffry Babb. Privacy fatigue: The effect of privacy control complexity on consumer electronic information disclosure. In *International Conference on Information Systems*, ICIS 2014, pages 14–17, December 2014. 2.2

[59] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS 2009, pages 1–12, 2009. 2.3

[60] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2010, pages 1573–1582, 2010. 2.3

[61] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. A conundrum of permissions: installing applications on an android smartphone. In *International conference on financial cryptography and data security*, pages 68–79. Springer, 2012. 2.1

[62] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2013, pages 3393–3402, 2013. 2.3

[63] Patrick Gage Kelley, Paul Hankes Drielsma, Norman Sadeh, and Lorrie Faith Cranor. User-controllable learning of security and privacy policies. In *Proceedings of the 1st ACM Workshop on Workshop on AISec*, AISec 2008, page 11–18, New York, NY, USA, 2008. Association for Computing Machinery. 2.4

[64] Eunjin Kim and Byungtae Lee. E-service quality competition through personalization under consumer privacy concerns. *Electronic Commerce Research and Applications*, 8(4):182 – 190, 2009. Special Issue: Economics and Electronic Commerce. 1.2, 4, 4.1

[65] Ron Kohavi and Roger Longbotham. Online controlled experiments and a/b testing. *Encyclopedia of machine learning and data mining*, 7(8):922–929, 2017. 6.4.3

[66] Kat Krol and Sören Preibusch. Control versus effort in privacy warnings for webforms. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, pages 13–23. ACM, 2016. 2.2

[67] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2012, page 589–598, New York, NY, USA, 2012. ACM. 2.2, 3.1, 6.4.4

[68] Tong Li, Mingyang Zhang, Hancheng Cao, Yong Li, Sasu Tarkoma, and Pan Hui. "what apps did you use?": Understanding the long-term evolution of mobile app usage. In *Proceedings of The Web Conference*, WWW 2020, pages 66–76, 2020. 2.4

[69] Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the ACM Conference on Ubiquitous Computing*, UbiComp 2012, pages 501–510, New York, NY, USA, 2012. ACM. 1.1, 1, 2.4, 5

[70] Jialiu Lin, Michael Benisch, Norman Sadeh, Jianwei Niu, Jason Hong, Banghui Lu, and Shaohui Guo. A comparative study of location-sharing privacy preferences in the united states and china. *Personal Ubiquitous Computing*, 17(4):697–711, April 2013. 2.4

[71] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *10th Symposium On Usable Privacy and Security*, SOUPS 2014, pages 199–212, Menlo Park, CA, 2014. USENIX Association. 1.3, 2.1, 4, 2.4, 5.1, 5.3.1, 6.4.1

[72] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun (Aerin) Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Twelfth Symposium on Usable Privacy and Security*, SOUPS 2016, pages 27–41, Denver, CO, 2016. USENIX Association. 1.1, 2.1, 2, 4, 2.4, 5, 5.1, 5.2.1, 5.3.1, 6.4.3

[73] Bin Liu, Jialiu Lin, and Norman Sadeh. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help? In *Proceedings of the 23rd International Conference on the World Wide Web*, WWW 2014, pages 201–212, New York, NY, USA, 2014. ACM. 2, 2.4, 5.1, 5.3.4

[74] Awio Web Services LLC. Web browser market share. `http://www.w3counter.com/globalstats.php?year=2021&month=1`, Jan 2021. 3.2.1

[75] Yemian Lu, Qi Li, Purui Su, Juan Pan, Jia Yan, Pengyi Zhan, and Wei Guo. A comprehensive study of permission usage on android. In *International Conference on Network and System Security*, pages 64–79. Springer, 2018. 2.4

[76] Kirsten Martin and Katie Shilton. Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society*, 32(3):200–216, 2016. 4.1.1

[77] Arunesh Mathur, Jessica Vitak, Arvind Narayanan, and Marshini Chetty. Characterizing the use of browser-based blocking extensions to prevent online tracking. In *Fourteenth Symposium on Usable Privacy and Security*, SOUPS 2018, pages 103–116, Baltimore, MD, August 2018. USENIX Association. 2.2, 3.1, 3.4, 4.1, 6.4.3

[78] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research. *Proceedings of the ACM on Human Computer Interaction*, 3:1–23, 2019. 3.2.4, 4.2.1

[79] G. Merzdovnik, M. Huber, D. Buhov, N. Nikiforakis, S. Neuner, M. Schmiedecker, and E. Weippl. Block me if you can: A large-scale study of tracker-blocking tools. In *IEEE European Symposium on Security and Privacy*, Euro S&P 2017, pages 319–333. IEEE, 2017. 4.1

[80] Microsoft. Microsoft edge, browsing data, and privacy. `https://support.microsoft.com/en-us/windows/microsoft-edge-browsing-data-and-privacy-bb8174ba-9d73-dcf2-9b4a-c582b4e640dd`. 3.1

[81] Heather Molyneaux, Elizabeth Stobert, Irina Kondratova, and Manon Gaudet. Security matters... until something else matters more: Security notifications on different form factors. In *International Conference on Human-Computer Interaction*, pages 189–205. Springer, 2020. 2.2

[82] Mozilla. Enhanced tracking protection in firefox. `https://support.mozilla.org/en-US/kb/enhanced-tracking-protection-firefox-desktop`, 2021. 3.1, 3.2.2

[83] Jonathan Mugan, Tarun Sharma, and Norman Sadeh. Understandable learning of privacy preferences through default personas and suggestions. `http://reports-archive.adm.cs.cmu.edu/anon/isr2011/abstracts/11-112.html`, Aug 2011. 2.2, 2.4, 5.1

[84] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In *Thirteenth Symposium on Usable Privacy and Security*, SOUPS 2017, pages 399–412. USENIX Association, 2017. 3.3.2, 4.1.1

[85] Toru Nakamura, Shinsaku Kiyomoto, Welderufael B Tesfay, and Jetzabel Serna. Easing the burden of setting privacy preferences: A machine learning approach. In *International Conference on Information Systems Security and Privacy*, pages 44–63. Springer, 2016. 2.4

[86] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009. 6.4.1

[87] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs*, 41(1):100–126, 2007. 1.3

[88] Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119:1–88, May 2016. 5, 5.3.5, 6.4.4

[89] Thanasis Petsas, Antonis Papadogiannakis, Michalis Polychronakis, Evangelos P Markatos, and Thomas Karagiannis. Rise of the planet of the apps: A

systematic study of the mobile app ecosystem. In *Proceedings of the Conference on Internet Measurement*, pages 277–290, 2013. 2.4

[90] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Twelfth Symposium on Usable Privacy and Security*, SOUPS 2016, pages 77–96, Denver, CO, June 2016. USENIX Association. 2.2, 3.1, 3.3.2, 4.1.1, 6.4.1

[91] Abhilasha Ravichander, Alan W Black, Thomas Norton, Shomir Wilson, and Norman Sadeh. Breaking down walls of text: How can NLP benefit consumer privacy? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4125–4140, Online, August 2021. Association for Computational Linguistics. 6.4.3

[92] Abhilasha Ravichander, Alan W Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. Question answering for privacy policies: Combining computational and legal perspectives, 2019. 6.4.3

[93] Ramprasad Ravichandran, Michael Benisch, Patrick Gage Kelley, and Norman Sadeh. Capturing social networking privacy preferences:. In Ian Goldberg and Mikhail J. Atallah, editors, *Privacy Enhancing Technologies*, pages 1–18, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. 2.2, 2.4

[94] Joel R Reidenberg, N Cameron Russell, Alexander J Callen, Sophia Qasir, and Thomas B Norton. Privacy harms and the effectiveness of the notice and choice framework. *ISJLP*, 11:485, 2015. 1.3

[95] Odnan Ref Sanchez, Ilaria Torre, and Bart P Knijnenburg. Semantic-based privacy settings negotiation and management. *Future Generation Computer Systems*, 111:879–898, 2020. 2.4

[96] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. Designing effective privacy notices and controls. *IEEE Internet Computing*, 2017. 1.1, 1.4, 2.1, 3, 3.1, 6.4.3

[97] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS 2015, pages 1–17, Berkeley, CA, USA, 2015. USENIX Association. 1.2, 2.1, 3.1

[98] Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. Watching them watching me: Browser extensions' impact on user privacy awareness and concern. In *NDSS Workshop on Usable Security*, NDSS 2016, pages 1–10, 2016. 2.2, 3.4, 6.4.3

[99] Sebastian Schelter and Jérôme Kunegis. On the ubiquity of web tracking: Insights from a billion-page web crawl. *Journal of Web Science*, 4:53–66, 2018. 1.2, 3.1, 4, 4.1

[100] Tanusree Sharma, Hunter A Dyer, and Masooda Bashir. Enabling user-centered privacy controls for mobile applications: Covid-19 perspective. *ACM Transactions on Internet Technology*, 21(1):1–24, 2021. 2.4

[101] Yun Shen and Pierre-Antoine Vervier. Iot security and privacy labels. In *Privacy Technologies and Policy*, pages 136–147. Springer International Publishing, 2019. 2.3

[102] Michael Simon. Apple is removing the do not track toggle from safari, but for a good reason. `https://www.macworld.com/article/3338152/apple-safari-removing-do-not-track.html`, Feb 2019. 1.1, 1.5.1

[103] Daniel Smullen, Yuanyuan Feng, Shikun Aerin Zhang, and Norman Sadeh. The best of both worlds: Mitigating trade-offs between accuracy and user burden in capturing mobile app privacy preferences. *Proceedings on Privacy Enhancing Technologies Symposium*, 2020(1):195 – 215, 01 Jan. 2020. 3.1, 4.2.2, †

[104] Daniel Smullen, Yaxing Yao, Yuanyuan Feng, Norman Sadeh, Arthur Edelstein, and Rebecca Weiss. Managing potentially intrusive practices in the browser: A user-centered perspective. *Proceedings on Privacy Enhancing Technologies*, 2021(4):500–527, 2021. 3.2.2, 3.3.2, 3.3.3, †

[105] Daniel Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154:477, 2005-2006. 1.2, 4.1, 4.2

[106] Daniel J Solove. *Nothing to hide: The false tradeoff between privacy and security*. Yale University Press, 2011. 1

[107] Aditya K Sood and Richard J Enbody. Malvertising–exploiting web advertising. *Computer Fraud & Security*, 2011(4):11–16, 2011. 4.3.2

[108] B. Stanton, M. F. Theofanos, S. S. Prettyman, and S. Furman. Security fatigue. *IT Professional*, 18(5):26–32, 2016. 2.2

[109] Peter Story, Daniel Smullen, Yaxing Yao, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. Awareness, adoption, and misconceptions of web privacy tools. In *Proceedings on Privacy Enhancing Technologies Symposium*, volume 3 of *PoPETS 2021*, July 2021. 2.2, 3.3.2, 3.3.3, 4.3, 6.4.1

[110] Helmut Strasser and Christian Weber. On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics*, 8, 02 1970. 5.2.2

[111] Berin Michael Szoka. The paradox of privacy empowerment: The unintended consequences of do not track. In *The 41st Research Conference on Communication, Information and Internet Policy*, TPRC 2013, 2013. 1.5.1, 6.2, 6.4.4

[112] Joshua Tan, Khanh Nguyen, Michael Theodorides, Heidi Negrón-Arroyo, Christopher Thompson, Serge Egelman, and David Wagner. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2014, pages 91–100, New York, NY, USA, 2014. ACM. 1.1, 5.3.4

[113] Richard H. Thaler and Cass R. Sunstein. Libertarian paternalism. *American Economic Review*, 93(2):175–179, May 2003. 1.4

[114] Richard H. Thaler, Cass R. Sunstein, and John P. Balz. *Chapter 25: Choice Architecture*, pages 428–439. Princeton University Press, 2013. 1.4

[115] Nikolaos Tsalis, Alexios Mylonas, and Dimitris Gritzalis. An intensive analysis of security and privacy browser add-ons. In *International Conference on Risks and Security of Internet and Systems*, pages 258–273. Springer, 2015. 2.2, 3.4, 6.4.3

[116] Randika Upathilake, Yingkun Li, and Ashraf Matrawy. A classification of web browser fingerprinting techniques. In *7th International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5, 2015. 1.2, 4

[117] Max Van Kleek, Ilaria Liccardi, Reuben Binns, Jun Zhao, Daniel J. Weitzner, and Nigel Shadbolt. Better the devil you know: Exposing the data sharing practices of smartphone apps. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2017, pages 5208–5220, New York, NY, USA, 2017. ACM. 2.4

[118] Anthony Vance, David Eargle, Jeffrey L Jenkins, C Brock Kirwan, and Bonnie Brinton Anderson. The fog of warnings: how non-essential notifications blur with security warnings. In *Fifteenth Symposium on Usable Privacy and Security*, SOUPS 2019. USENIX Association, 2019. 2.2

[119] Daniel Votipka, Seth M. Rabin, Kristopher Micinski, Thomas Gilray, Michelle L. Mazurek, and Jeffrey S. Foster. User comfort with android background resource accesses in different contexts. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, SOUPS 2018, pages 235–250, Baltimore, MD, August 2018. USENIX Association. 2.4

[120] Diane Walker and Florence Myrick. Grounded theory: An exploration of process and procedure. *Qualitative Health Research*, 16(4):547–559, 2006. PMID: 16513996. 4.2, 4.2.1

[121] Rui Wang, Shuo Chen, and Xiao Feng Wang. Signing me onto your accounts through facebook and google: A traffic-guided security study of commercially deployed single-sign-on web services. In *IEEE Symposium on Security and Privacy*, S&P 2012, pages 365–379. IEEE, 2012. 1.2, 4

[122] Alan F. Westin. *Privacy and freedom*. Atheneum, 1970. 1

[123] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *IEEE Symposium on Security and Privacy*, S&P 2017, pages 1077–1093. IEEE, May 2017. 2.4

[124] Daricia Wilkinson, Moses Namara, Karla Badillo-Urquiola, Pamela J. Wisniewski, Bart P. Knijnenburg, Xinru Page, Eran Toch, and Jen Romano-Bergstrom. Moving beyond a "one-size fits all": Exploring individual differences in privacy. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA 2018, page 1–8, New York, NY, USA, 2018. ACM. 1, 1.3, 1.4

[125] Shikun Zhang, Yuanyuan Feng, Lujo Bauer, Lorrie Faith Cranor, Anupam Das, and Norman M Sadeh. "did you know this camera tracks your mood?": Understanding privacy expectations and preferences in the age of video analytics. *Procedings on Privacy Enhancing Technologies Symposium*, 2021(2):282–304, 2021. 2.4

[126] Shikun Zhang, Yuanyuan Feng, Anupam Das, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Understanding people's privacy attitudes towards video analytics technologies. *Proceedings of the Federal Trade Commission Privacy Conference*, pages 1–18, 2020. 2.4

[127] Yixin Zou, Kevin Roundy, Acar Tamersoy, Saurabh Shintre, Johann Roturier, and Florian Schaub. Examining the adoption and abandonment of security, privacy, and identity theft protection practices. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI 2020, page 1–15, New York, NY, USA, 2020. ACM. 1.2, 3.1, 4.1