Thesis Proposal

Understanding Users' Online Privacy Expectations

Ashwini Rao
Institute for Software Research
Carnegie Mellon University
arao@cmu.edu

February 2016

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Thesis Committee

Norman Sadeh (Chair)
Institute for Software Research
Carnegie Mellon University

Alessandro Acquisti
Heinz College
Carnegie Mellon University

James Herbsleb
Institute for Software Research
Carnegie Mellon University

Joel Reidenberg (External)
School of Law
Fordham University

Florian Schaub
Institute for Software Research
Carnegie Mellon University

Abstract

Online privacy notices are supposed to act as the primary mechanism to inform users about the data practices of online services. In practice, users ignore notices as they are too long and complex to read. Instead, users rely on expectations to determine which sites they feel comfortable interacting with. Mismatches between actual practices and users' expectations may result in users exposing themselves to unanticipated privacy risks. One approach for mitigating these risks is to highlight elements of privacy notices that users do not likely expect. However, to do so, we have to first identify (a) users expectations and (b) mismatches between users' expectations and data practices of online services. As part of this thesis, I propose to address these two challenges.

First, to understand users' online privacy expectations, I borrow and apply the concept of different types of expectations from non-privacy domain to privacy domain. I distinguish between two types of privacy expectations: Desired expectation ("should") and Likelihood expectation ("will"). I elicit both types of expectations via user studies. I discuss the importance of distinguishing between multiple types of expectations. I study how expectations vary by contextual factors (e.g. website type) and user characteristics (e.g. privacy knowledge and concern).

Second, I present an approach for identifying mismatches between users' expectations and data practices of online services, and investigate the impact of such mismatches on users' online privacy. I identify different types of mismatches that result from Desired and Likelihood expectation types. I study how mismatches vary by contextual factors and user characteristics.

Lastly, based on the understanding gained from studying expectations and mismatches, I design and test simplified privacy notices. I study whether simplified notices that highlight mismatches in users' expectations can be shorter and easier to comprehend for users. I also discuss the impact of the results on public policy management.

Contents

1. Introduction

The Internet has become an integral part of our lives, but it has come at a cost. The Internet has revolutionized and improved the way we live through services such as email, social networking and search that enable us to communicate, socialize and access information in novel ways. However, one of the important challenges today is addressing user privacy concerns that arise as users interact with online services. Users have expectations regarding data practices of online services, for example, types of data collected on websites, purposes for which websites share data and whether websites allow users to delete personal data. However, services may engage in data practices that are misaligned with user expectations. Mismatches between user expectations and data practices may result in concerns regarding violation of users' online privacy.

Currently, privacy policies serve as the primary mechanism for notifying users about data practices. However, privacy policies, written in natural language, can be long, time consuming to read [18, 27], and difficult to understand for users [36, 40]. They are therefore often ignored by users [8, 37]. Prior work has proposed simplifying privacy policies through summary notices that display data practices in an easy to understand visual format [9, 21, 44, 49]. Even with simplified notices, much of the information may not be relevant to users. Many data practices are expected and obvious, or may not create concern. For instance, it is obvious to users that when they explicitly provide their contact and payment details to an online store that information will be collected and is needed to fulfill the purchase. However, data practices that are unexpected may result in loss of trust and a sense that one's privacy has been violated, even if the practices in question were disclosed in a privacy notice [41].

The concept of contextual integrity highlights the importance of information flows between different contexts. The societal and transactional context in which data is collected shapes users' expectations of how the data will be used or whether it may be shared with other entities [31, 32]. For instance, collection of financial information on a banking website may be more expected than collection of health information. Privacy expectations are further influenced by an individual's personal, social and cultural background, as well as expectations in social roles and other "borders" that delineate spheres of privacy [26, 32]. For instance, depending on their technical knowledge, some users may expect that websites they visit can collect rough location information about them based on their IP address. For others, inference of their location may be completely unexpected.

Although unexpected data practices may be described in a privacy policy, they often get lost between descriptions of practices that are expected or irrelevant to the user's current transactional context. The verbosity of privacy policies may be necessary to comply with legal and regulatory requirements, but it also means that privacy policies are not helpful to users in making informed privacy decisions [8]. In order to provide transparency to users, compliance-oriented privacy policies need to be complemented with short form notices tailored to the user's transactional context [44, 45] that should warn users about unexpected practices in particular [14]. The challenge, however, lies in identifying unexpected practices. Much work has studied users' privacy preferences in different contexts [22, 35]. However, privacy behavior differs from stated preferences [33] and they are not reliable for identifying mismatches between expectations, in the sense of "stated preferences," and actual site practices.

To advance toward more practical solutions that can impact privacy notice design, I propose to do the following. First, I outline an approach for determining mismatches between users' expectations and sites' data practices, as stated in their privacy policies. Research in other fields e.g. marketing has highlighted that the term "expectations" can have several meanings in consumers' context [29]. However, in the privacy context most work has focused on expectations in the desired sense or preferences [22, 30] or has not clarified the meaning of expectation [12, 16, 24]. I disambiguate the meaning of expectation by differentiating between more subjective expectation in the sense of "desire" and more objective expectation in the sense of "likelihood" or "expected frequency." Henceforth, I will use the terminology Desired expectation to refer to

the former and Likelihood expectation to refer to the latter. Focusing on Likelihood expectation could avoid problems resulting from Desired expectation, for example, users stated Desired expectation may not match their actual behavior [6]. I elicit both types of expectations and do a three-way comparison between Desired expectations, Likelihood expectations and actual data practices to identify mismatches in expectations. I identify different types of mismatches for each type of expectation and discuss how the meaning of mismatch changes when we consider both expectation types simultaneously. I study the impact of mismatches on users, for example, whether users attach more importance to certain types of mismatches. Lastly, I identify factors that impact user expectations and mismatches. To evaluate whether the approach developed in context of website data practices is generalizable, I propose to do an optional study in the context of Internet-of-Things. Second, based on the understanding gained from studying user expectations and mismatches for website data practices, I design simplified website privacy notices and evaluate whether they can reduce user burden.

In addition to simplifying privacy notices, understanding user privacy expectations and mismatches regarding data practices can have other benefits. Service providers could understand data practices that users will likely not expect and that may therefore become cause for privacy concern. They could then improve their user-facing privacy notices to emphasize these practices and, at the same time, explain the rationale behind those practices to assuage user concerns. Making their websites data practices and privacy policies easier to understand could allow them to obtain a competitive advantage. Results could inform the design of privacy services and tools, such as browser extensions, that aim to improve privacy transparency online. Regulatory agencies such as the Federal Trade Commission work on protecting users privacy, and mismatched expectations could indicate to them important public issues that need attention.

So far, I have conducted a user study to gain insights into users' Likelihood expectations. I compared their Likelihood expectation to data practices of websites to identify mismatches. My study focused on website scenarios and important collection, sharing and deletion website data practices. I compared users' Likelihood expectations to data practices disclosed in website privacy policies. Using empirical data from the user study, I have identified several mismatches regarding website data practices. I have also identified the existence of different types of mismatches for Likelihood expectation type. My analysis shows that characteristics of a website, such as its type, as well as user characteristics, such as privacy knowledge and concern, are strong predictors of data practices that are likely to be unexpected. My results show that the number of mismatches may be small and designing privacy notices that highlight mismatches can shorten privacy notices. They also show the potential for contextualizing and personalizing privacy notices.

2. Thesis

My thesis is the following:

> **We can understand users' online privacy expectations and corresponding mismatches by (i) distinguishing between different types of expectations (ii) considering factors that impact expectations and mismatches and (iii) studying impact of different types of mismatches on users.**

I elaborate on the different elements of the thesis below.

> **...(i) distinguishing between different types of expectations...**

Expectation can have different types [15, 29, 46, 48], for example, Miller proposed four conceptual types, Desired, Expected, Ideal and Minimally Acceptable in the context of marketing [29]. However, in the privacy context most work has either not clarified the meaning of expectation [12, 16, 24] or has focused on expectations in the desired sense or preferences [22, 30]. Not

clarifying the meaning of expectation can lead to ambiguous results. Measuring only desired expectation is not ideal. First, privacy notices inform users whether a data practice is likely, and the data practice may or may not match users' desired expectation. Second, privacy behavior can differ from users' desired expectations [33] and may not be reliable for identifying mismatches between expectations and actual site practices. I measure two types of expectations: more subjective expectation in the sense of "desire" and more objective expectation in the sense of "likelihood" or "expected frequency." I refer to the former as Desired expectation and latter as Likelihood expectation. Eliciting both expectation types will enable us to identify and analyze the impact of mismatches resulting from both expectation types.

...(ii) considering factors that may impact expectations and mismatches...

Users expectations may vary based on contextual factors such as type of website, for example, users may expect financial websites to collect financial information, but not health information. If expectations vary based on contextual factors, mismatches can also vary based on contextual factors. I identify contextual factors that impact expectations and mismatches. User expectations and mismatches may also vary based on user characteristics such as demographic and privacy knowledge. I identify user characteristics that impact expectations and mismatches.

...(iii) impact of different types of mismatches on users.

Users may perceive different types of mismatches differently, for example, some mismatches may be more concerning than others.

To understand the practical implications of studying expectations and mismatches, I design and test privacy notices that highlight mismatches in users' expectations. I study whether such notices can be shorter than full privacy notices and whether they are easier to comprehend for users. Currently, users find website privacy notices, written in natural language, long and time consuming to read [18, 27], and difficult to understand [36, 40]. Work on simplifying privacy notices has focused on more effective visual formats [9, 21, 44, 49]. However, many of the data practices displayed in such simplified notices may be expected, for instance, collection of health information on health websites. A privacy notice may not need to display information that users already expect or know, but only display those data practices that users do not expect or do not know.I also discuss the impact of the results on public policy management. While designing notices, I consider whether we could display only mismatches that are specific to a given context, and, further, we could personalize a notice by displaying mismatches that are specific to certain users. I also consider whether only a subset of mismatches that have greater impact on users' privacy could be displayed.

2.1. Contributions

The main contributions of this thesis are as follows:

- It will show that it is important for privacy research to consider expectation as a conceptual variable that has multiple types or levels.
- It will demonstrate that mismatches in user privacy expectations can be of different types and can impact user privacy differently.
- It will outline an approach for identifying mismatches between user expectations regarding data practices of online services and the actual practices of online services.
- It will identify important mismatches in online privacy expectations, which can inform design of better privacy tools and techniques as well as privacy regulations.
- It will study whether by focusing on mismatches, it is possible to design privacy notices that are shorter and more easy to comprehend for users.

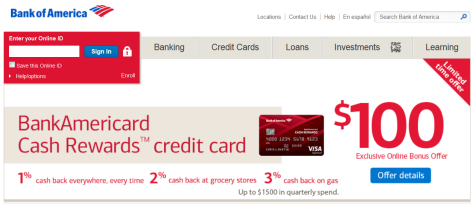

Figure 1. *Users may have expectations regarding data practices of Bankofamerica.com website.*

3. Example and Problem Description

The following example will illustrate the need for understanding users' online privacy expectations. Consider a scenario where users are using Bank of America website shown in Fig. 1. Users may consider criteria such as utility to decide whether to use or not use the website. When thinking about privacy, users may also consider the website's data practices such as types of data collected, with whom data is shared and deletion of user data. User expectations regarding data practices can impact the decision to use or not use a website because expectations influence decision making [17]. Since website privacy policies are difficult to read [27], users may infer data practices based on heuristics such as website type. For example, users may expect that a banking website *will* collect financial information, but not health information. In addition to what a website *would* do, they may have opinions about what it *should* or should not do. For example, they may expect that a banking website *should* not collect health information. The former is a more objective expectation than the latter expectation, which is more subjective. Expectations, regarding website data practices, formed based on heuristics may be inaccurate because they may not match websites' actual data practices. For example, a banking website may indeed collect health information. Mismatched privacy expectations may lead to erroneous privacy decisions. By identifying mismatches and highlighting them in a privacy notice could enable users to make better privacy decisions. Since the number of mismatches regarding data practices may be smaller than the total number of data practices themselves, the notice could be shorter. Shorter notices could imply that users have to spend less time reading them. Further, they would focus on information that they do not already expect or know. However, the key challenge is to identify expectations, mismatches and the impact of mismatches on users. I propose to address this key challenge in this thesis.

4. Related Work

This sections discusses work related to my proposed approach.

4.1. Privacy Expectations

Research in non-privacy domains such as consumer psychology has found that users can have multiple levels or types of expectations [15, 29, 46, 48] and these types can impact constructs such as consumer satisfaction [46] and performance [15]. For example, Miller proposed four expectation types: Ideal, Expected, Minimum Tolerable and Deserved [29]. The Ideal represents what users think performance "can be." The Expected is objective, without an affective dimension, and represents what users think performance "will be." The Deserved has an affective dimension and represents what users feel performance "should be." Lastly, the Minimum Tolerable is what users think the lowest performance "must be." Therefore simply asking users what they "expect" can lead to multiple interpretations. Swan and Trawick found two types of expectations: Predictive and Desired [46]. The Predictive is what users objectively think will happen, and the

Desired is what they subjectively want to happen.

Privacy research has explored the concept of expectation of privacy, including seminal work by Altman [3] and Nissenbaum [31]. Altman shows that people continuously modify their behavior to achieve an expected level of privacy [3], and Nissenbaum discusses how expectation of privacy can change based on context [31]. However, to the best of my knowledge, privacy research has not focused on the fact that users can have multiple levels or types of expectations. In Altman and Nissenbaum's work, there is a single level of expectation (similar to the Deserved level) that can change based on factors such as context. Privacy research differentiates between expected level of privacy and actual level of privacy, for example, Altman differentiates between desired and achieved levels [3]. However, this is not same as differentiating between multiple levels or types of expectations that may exist in users' minds. In this work, I suggest, based on work by Miller [29], that people can have multiple levels of expectation of privacy even when all factors are constant. I distinguish between Expected ("will be") and Deserved ("should be") types in measuring user expectations for data practices. I show that distinguishing between them has important consequences on simplifying website privacy notices.

4.2. Mismatches in Expectations

I study mismatches between users' expectations regarding data practices of services related to the Internet and optionally IoT. Prior work has studied mismatches in other types of expectations [12, 16, 24, 30]. Earp et al. compared the types of privacy protective statements users expected website privacy policies to contain with types of statements in the policies [12]. Milne and Bahl examined differences between consumers' and marketers' expectations regarding use of eight information technologies [30]. Gomez et al. compared data practices of websites with data practices that users found concerning [16]. Liu et al. measured disparity between expected and actual Facebook privacy settings. To measure expectation, these studies either used an expectation type with an affective dimension such as *should* [30] or did not clarify the type of expectation [12, 16, 24]. Other than expectations, researchers have studied constructs such as users' willingness to share information [22, 35].

4.3. User Privacy Preferences

Researchers have studied users' privacy preferences and willingness to share information in different contexts [22, 35]. According to Acquisti et al. [2], privacy preferences and privacy decision making are prone to uncertainty, context-dependent, shaped by heuristics and cognitive biases, malleable and easily influenced by framing. Elicited privacy preferences are therefore often difficult to generalize, and actual behavior often deviates from stated preferences [33].

4.4. Simplified Privacy Notices

To simplify website privacy notices, researchers have developed display formats that are easier for users to understand [9, 21, 28, 49]. A multilayer privacy notice uses a table format to display portions of a policy and provides links to the full policy [9]. Privacy nutrition label approach proposed by Kelley et al. displays data practices in nutrition label format [21]. Based on the data practices contained within a policy, Privee approach proposed by Zimmeck and Bellovin assigns letter grades, A, B or C, to the policy [49]. I do not develop a new display format rather focus on whether highlighting mismatches in expectations can improve shorten notices and improve comprehension.

5. Proposed Approach

This section describes the proposed approach and the technical challenges that must be addressed.

5.1. Privacy Expectations and Mismatches

5.1.1 Privacy Expectations

Based on work in consumer psychology [15, 29, 46, 48], I propose to study two types of privacy expectations. I distinguish between expectation in the sense of "desire" and expectation in the sense of "likelihood" or "expected frequency." These map to Deserved ("should be") and Expected ("will be") types in Miller's conceptual model of expectations [29]. They can be considered subjective and objective expectations as per work on expectations by Swan and Trawick [46]. In the context of Internet services, I will elicit user privacy expectations regarding website data practices. Further, I will optionally elicit user privacy expectations regarding data practices of devices, for example wearable devices, that are part of Internet-of-Things.

5.1.2 Identifying Mismatches

To identify mismatches in privacy expectations, I compare expectations elicited from users with data practices of Internet and Internet-of-Things services. To identify data practices of Internet websites, I extract information from website privacy policies. Given the low adoption rate of privacy policies written in standardized machine readable formats such as P3P [10], recent work has focused on extracting information from natural language text privacy policies using machine learning, natural language processing and crowdsourcing techniques [23, 42, 49]. However, fully automated efforts [49] to interpret privacy policies face the challenge of ambiguous and missing policy statements. One approach to mitigate this challenge is to use semi-automated techniques that involve humans in the loop [42]. To improve reliability of extracted information, researchers have proposed using multiple methods [40]. For example, Reidenberg et al. use a combination of crowdworkers and experts to annotate policies. They suggest using answers from crowdworkers if there is at least 80% agreement among crowdworkers if not use answers from domain experts.

In this work, I use annotations from experts to extract information from privacy policies. As the semi-automated methods [42] improve, we can scale-up the number of policies from which we can extract information. If privacy policies are available for devices in Internet-of-Things, I will use a similar approach to extract data practices of such devices. If privacy policies are not available, I propose to consult experts with knowledge of such devices to understand the devices' data practices. To extract answers, I adapt a questionnaire from a prior study by Riedenberg et al. [40] For each collection and sharing data practice, the answers can be Yes, No, Unclear or Not addressed. The "Yes" option implies that the website engages in the data practice. The "No" option implies that the website does not engage in the data practice. The "Unclear" option indicates that the website's policy contains statements that are unclear about the data practice. For example, a policy may state that the website collects IP address, but does not clarify whether IP address will be used to identify users' current location. Therefore the policy is unclear about collection of current location information. The "Not addressed" option indicates that the website's policy does not contain statements that address whether or not the website engages in the data practice. For deletion data practice, extracted answers indicate whether the website allows deletion of all data ("Full deletion"), allows deletion of some data ("Partial deletion"), is unclear about deletion ("Unclear") or does not address ("Not addressed") deletion of information.

5.1.3 Types of Mismatches

To identify mismatched expectations and therefore unexpected data practices, I compare participants' expectations concerning a specific data practice with the actual data practice. In general, answers extracted for a given data practice may be Yes, No, Unclear or Not addressed. To compare users' Likelihood expectations with actual data practices, expectations elicited from users can be interpreted as indications of a positive (Yes) or a negative (No) expectation. This results in eight potential combinations, as shown in Table 1: For Yes–Yes and No–No, users' expectations match the websites' practices. Yes–No and No–Yes combinations constitute explicit

		User:	Yes	No
Website:	Yes		✓	X
	No		X	✓
	Unclear		?	?
	Not addressed		?	?

Table 1. *Overview of matched and mismatched expectations. Match (✓) or mismatch (X) between a website's data practice and a user's expectation. If the website's policy is unclear or silent on a practice, it cannot be determined if it matches user expectations (?).*

mismatches. For Unclear–Yes, Unclear–No, Not addressed–Yes and Not addressed–No, it is not clear whether expectations are mismatched because the website's policy is unclear or silent on the particular data practice.

It is worth taking a closer look at the types of mismatches. Although, both Yes–No and No–Yes are mismatches, they may impact users' perception of privacy violations differently. In the case of Yes–No, the website will collect or share information, but users optimistically expect it not to. Due to lack of awareness that the website shares information, users may decide to use the website. By doing so, they give up data that they do not want to share resulting in violation of their data privacy. Although the website discloses its data practice in its policy, from a user viewpoint, the practice could be considered surreptitious unless users are appropriately and explicitly made aware of the practice. When found out, such data practices may damage a company's reputation.

In contrast, in the case of No–Yes, a website will not engage in a collection or sharing practice, but users pessimistically expect it to. As a result, users may have reservations of using the website or some features, which may affect their utility but not their privacy. In such cases, websites should aim to make users aware of the privacy-protective practices to assuage pessimistic expectations.

The number of unclear website data practices can be high, for example ~40% of collection data practices in a study I conducted were unclear. Hence, it is important to analyze the impact of unclear data practices. Consider the Unclear–Yes case. If the website is really collecting information, then it would be a Yes–Yes match. If the website is not collecting information, then it would be a No–Yes mismatch. The same applies to Unclear–No. As discussed, a Yes–No mismatch, could potentially violate user privacy. Hence, for analysis, we could treat Unclear as a likely Yes. We could use a similar approach for Not addressed–Yes and Not addressed–No.

We can similarly analyze mismatches in case of deletion data practice by considering two types of Yes values, Yes–full and Yes–Partial, separately. We could also simplify the analysis by combining the two Yes values. In case of deletion, users may use a website if they think that the website allows deletion whereas for collection and sharing they may not use the website. Hence, in case of deletion, the implications of No–Yes and Yes–No mismatches are reversed.

Similar to mismatches in Likelihood expectations ("will") , I analyze mismatches in Desired expectations ("should"). Users may answer Yes or No to whether a website *should* engage in a data practice. Considering "should" expectations in addition to "will" expectations, adds an additional dimension to the assessment of the implications stemming from matched or mismatched expectations.

For instance, consider when a user's "will" expectation matches the website's data practices (Yes–Yes). When combined with the "should" expectation type, only Yes–Yes–Yes is a perfect match, whereas Yes–Yes–No is a mismatch, i.e., users may expect the practice but prefer it to be different. For example, for data collection, a Yes–Yes–No indicates that a user is correctly aware that a website will collect information, but feels that it should not. The user may continue to use

the website due to lack of awareness of other websites that do not collect information. It may also imply market failure due to monopoly or due to all websites in the website category being equally privacy invasive. An example of such market failure can be search engine websites; although users may know that Google's search website collects certain data, they may continue to use Google for convenience and utility reasons.

Similarly, in case of a mismatch due to a website engaging in unexpected practices, the "should" expectation type may change the meaning of the mismatch. For example, when a Yes–No mismatch is combined with a "should" expectation. In a Yes–No–No mismatch, users both incorrectly think that a website will not engage in a data practice and feel that it should not. They may decide to use the website and lose data privacy. For Yes–No–Yes, users want the website to engage in a practice, but do not expect it to at the moment. For instance, users may want a website to provide personalized services based on their data. In this scenario, users may decide not to use the website and lose utility, but not data privacy.

The examples discussed above demonstrate the importance and potential of distinguishing and capturing the meaning of different expectation types in privacy research. In the case of website privacy notices, by distinguishing between expectation types, we may be able to better identify user needs and display appropriate information. For example, in case of a Yes–Yes–No mismatch, a privacy tool could display alternative websites with more privacy-friendly practices. In case of a Yes–No–Yes mismatch, such a tool could display whether an opt-in option for personalization is available.

Lastly, in addition to the semantics of mismatches, I investigate the impact of mismatches on users. Some mismatches may surprise users, but not really concern them. When designing simplified notices, we could display only the subset of mismatches that are concerning to users. This could further reduce the amount of information that users have to process while making privacy decisions.

5.2. Mismatches in Likelihood Expectation Type

To understand users' Likelihood privacy expectations ("will" or "likelihood") and mismatches that correspond to such expectations, I use empirical data from a quantitative user study. I study important website data practices related to collection, sharing and deletion of data on websites. I vary the scenarios under which data is collected and shared. To understand the impact of contextual factors on user expectations and mismatches, I consider several website characteristics. To understand how expectations and mismatches vary by users, I study several user characteristics. I use expert annotations to extract data practices from website privacy policies. I compare expectations elicited from users with data practices disclosed in the policies to identify mismatches.

5.2.1 Study Details

Data Practices and Scenarios: I focus on data practices concerning *collection, sharing and deletion of personal information* as prior research shows that users are especially concerned about surreptitious collection, unauthorized disclosure and wrongful retention of personal information [41]. I consider the collection and sharing of four categories of privacy-sensitive information [1, 19, 22]: *contact information* (e.g., email or postal address), *financial information* (e.g., bank account information, credit card details, or credit history), *health information* (e.g., medical history or health insurance information), and *current location* (e.g., from where a user is accessing the website).

I further distinguish between scenarios in which users have or do not have an *account with the website.* Websites typically collect data when users create an account, often explicitly provided by the user, thus registered users may be more aware of a website's data practices. They can collect data without an account, for example, via surveys or subscription to newsletters. In these cases, users actively input data and are aware of collection. In general, users may not be aware

Action	Scenario	Information type
Collection	With account	Contact
		Financial
		Health
		Current location
	Without account	Contact
		Financial
		Health
		Current location
Sharing	For core purpose	Contact
		Financial
		Health
		Current location
	For other purpose	Contact
		Financial
		Health
		Current location
Deletion	–	Personal data

Table 2. *Data practices and website scenarios used for studying Likelihood expectations.*

of implicit or automated data collection, e.g., of IP addresses and cookies. Websites may use IP addresses, email addresses and other information to acquire additional data about individuals, such as purchase history or interests, from social media services and data brokers [38]. Profiles may contain hundreds of attributes including name, address, purchases, health interests and credit summary, and are available based on IP address, email address etc. [38]

Similarly, information sharing with third parties, while abundant, is less visible to users. Websites assume to have the users' permission because they are using the website and therefore implicitly consent to its privacy policy. I distinguish between third party sharing for *core purposes*, such as sharing a user's information to provide the requested service (e.g., payment processing or providing contact information to a delivery service), and sharing for unrelated *other purposes*, such as advertising or marketing. In all, I study 17 data practices summarized in Table 2.

Website Characteristics: To understand whether mismatched privacy expectations vary based on context, I consider three website characteristics: website type, popularity and ownership. *Website type* may influence what information users expect a website to collect [31]. I study three website categories: finance, health and dictionary. Users may expect finance and health websites to collect sensitive information (health or financial data, respectively). In contrast, users may not expect dictionary websites to collect sensitive information. In the financial category, I include banking, credit card and online payment websites. In the health category, I include pharmacy, health clinic and health reference websites.

Users' expectations may be influenced by their offline interactions with entities affiliated with a website, such as visiting a bank branch or a clinic. Hence, I include websites with *offline interactions* as well as online-only websites in the health and financial categories; dictionary websites were online-only.

Interestingly, popular financial websites have been shown to have more privacy-invasive data practices than less popular ones [11]. Therefore, I study websites of comparable utility but

varying in *popularity*, as determined by their traffic rankings [4].

For a given website type, *government or private ownership* may influence user expectations. For example in the United States, in the post-Snowden era, people may expect government websites to be more privacy invasive than private websites. Hence, I study whether user expectations vary between government and privately-owned health and financial websites. Table 3 summarizes the website characteristics that I consider in the study.

User Characteristics: I explore the impact of the following user characteristics on user expectations: demographic, past privacy protective behavior, familiarity with privacy related concepts and tools, privacy knowledge, negative online experience, online privacy concern, and experience with website.

I explore the impact of the following standard demographic factors: gender, age, education level and occupation. In addition, I include background in computer related fields because such users may have more accurate understanding of website data practices.

I hypothesize that privacy knowledge would impact user expectations. For example, a user who understands how IP address works may have different expectation about collection of location information than a user who does not. Familiarity may also impact user expectations. For example, a user may be familiar with secure data transmission protocol (HTTPS) but not know the technical details. I use a validated privacy scale [20] that measures knowledge and familiarity of the following concepts and tools: private browsing or incognito mode, cookies, Tor, Virtual Private Network (VPN), IP address, HTTPS and proxy server.

I hypothesize that users who take steps to protect their online privacy may have more accurate understanding of website data practices. I include behaviors about protecting personal information [34]: checked that a website was secure before providing personal information; asked public or private sector organizations why they needed personal information; and read privacy policies and notifications before providing personal information. I also measure online privacy protective behaviors [20] such as cleared browser cookies and history, used a proxy server and encrypted web communications.

I hypothesize that users who have had negative online experiences may expect data practices to be more privacy invasive, for example, they may expect collection of information to be more likely. I use a scale that measures negative online experiences such as a company leaked user's personal information, user's important personal information such as Social Security Number or credit information was stolen and user had been a victim of an online scam and lost money.

I hypothesize that users with higher online privacy concern may expect data practices to be more privacy invasive. I measure online privacy concern using 10 items from IUIPC scale [25].

Lastly, I hypothesize that past experience with the website may impact user expectations. For example, users who had used a website may know whether a website allows deletion of data. I explore the impact of the following: user has an account on the website or on a similar website; amount of recent use; familiarity with website; and perception of trustworthiness of website. Table 3 provides summary of user characteristics used in the study.

Survey Questionnaire: I propose to use a questionnaire to measure user expectations for eight collection data practices (4 info. types collected with/without account), eight sharing data practices (4 info. types shared for core/other purposes), and one deletion data practice.

At the beginning of the survey, I will explain the purpose of the study. I will frame the purpose of the study as understanding user opinions about websites rather than their knowledge of data practices, to avoid self-presentation issues associated with knowledge questions [7]. Threat of self-presentation can increase the chance of users looking-up or discussing answers in self-administered studies [7]. I will also not mention privacy or data practices to avoid biasing participants. After explaining the purpose, I will ask whether participants have visited or used the assigned website before. I will ask participants to familiarize themselves with the website for 2–3 minutes.

Website characteristic	
Type	Finance
	Health
	Dictionary
Popularity	More
	Less
Ownership	Private
	Government
User characteristic	
Demographic: age, gender, education, occupation computer background, state of residence	
Privacy protective behavior	
Familiarity with privacy concepts and tools	
Knowledge of privacy concepts and tools	
Negative online experience	
Online privacy concern	
Experience with website: amount of recent use, has account, familiarity, trust	

Table 3. *Website and user characteristics used for studying Likelihood expectations.*

After they interact with the website, I will provide definitions of contact, financial, health and current location information. Next, I provide further contextualization by first showing them a scenario description such as "*Imagine that you are browsing [website name] website. You do not have a user account on [website name], that is, you have not registered or created an account on [website name].*" I will then ask them about their expectations concerning whether and how the website engages in data collection and data sharing, and its policy on data deletion. These questions will be framed as opinion rather than knowledge questions [7], e.g., "*What is the likelihood that [website name] would collect your information in this scenario?*" Note that I will frame the questions as "would collect" in order to capture participants' Likelihood expectations. I will provide a 4-point scale {Likely, Somewhat likely, Somewhat unlikely, Unlikely} without a neutral or not sure option to capture respondents' "best guess." This is because users often do not read privacy policies and decide about data practices of a website based on incomplete information, that is, their best guess. I will ask an open-ended question to understand how they thought the website collected their information without having an account on the website. I will ask participants questions regarding their expectations if they have an account. When inquiring about sharing questions, I will also ask participants to describe how they interpreted core purposes, other purposes, and with whom the website may share information to better understand their rationale. Concerning the data deletion practice, I will ask participants whether they expect that the website would allow them to delete all, some or none of their data. In the second part of the survey, I will capture different user characteristics described in Table 3 in order to study their impact on the participants' privacy expectations. I will order the questions based on ease of answering, level of threat, and effect on subsequent answers [7].

5.3. Comparing Mismatches from Likelihood and Desired Expectation Types

In addition to understanding users' Likelihood privacy expectations ("will" or "likelihood"), I propose to study users' Desired privacy expectations ("should" or "desire"). I will do a three-way comparison between Likelihood expectations, Desired expectations and actual practices, and identify mismatches between the three pairs. Mismatches between Likelihood expectations and actual practices can result in users' making incorrect privacy-related decisions. Mismatches between Desired expectations and actual practices indicate whether gaps between what people want from services and what services actual provide them. Mismatches between Likelihood expectations and Desired expectations could help us decide whether we should display mismatch resulting from Likelihood or Desired expectation. For example, users whose Likelihood expectation matches the actual practice, but Desired expectation does not are more sophisticated. For this group, we could display the mismatch associated with Desired expectation.

I hypothesize that the mismatches between Likelihood expectations and actual practices will be different from mismatches between Desired expectations and actual practices. I hypothesize that there will be differences in factors that impact Likelihood and Desired expectations and mismatches resulting from the two types of expectations. I will analyze whether the number of mismatches between Likelihood expectations and actual practices are fewer than mismatches between Desired expectations and actual practices. If that is the case, users' Likelihood expectations are more realistic than Desired expectations.

5.3.1 Study Details

To do a three-way comparison between Likelihood expectations, Desired expectations and actual practices, I propose to conduct a quantitative user study similar to the study I discussed in Section 5.2 for understanding Likelihood expectations. Additional requirements for the new study are as follows. To understand both users' Likelihood and Desired expectations, I will elicit both expectation types from a given participant. To elicit Desired expectation, I will ask questions such as "*Do you think that [website name] should be allowed to collect your information in this scenario?*"

Based on the results from the first study, I may also refine the new study as follows. I may consider additional data practices. For example, in addition to collection, sharing and deletion practices, I may consider websites tracking practices [38]. Further, I may consider additional scenarios. For example, I may consider additional purposes for sharing such as marketing. I may include additional website categories such as social networking and news categories. Lastly, I may explore additional user characteristics, for example, self-efficacy of privacy protection behavior [47] and questions regarding users' security behavior intentions [13].

Case Study for Internet-of-Things: In addition to studying mismatches in privacy expectations regarding website data practices, I propose an optional study to understand mismatches in expectations regarding data practices of devices in the Internet-of-Things (IoT). Compared to the Internet, ubiquitous computing environment associated with IoT raises new privacy issues [43]. For example, sensors can be invisible to users and collect information without users' knowledge or consent. In contrast, on websites users interact with websites and are relatively more aware of collection of information. I will select and study a class of devices, for example, video cameras or wearables. Similar to the study on website data practices, I will identify data practices (e.g. types of data collected by the device, collection and sharing scenarios etc.) of the device. I will extract data practices from privacy policies if available. Otherwise, I will consults experts who are knowledgeable of data practices of such devices. I will conduct a qualitative study using an appropriate purposeful sampling strategy. Instead of an online survey, I will conduct an in-situ study. I will compare how mismatches in expectations vary between websites and IoT devices, and discuss implications for simplifying privacy notices in the IoT environment.

5.4. Impact of Mismatches on Users

Identifying mismatches is the first step toward designing simplified privacy notices. However, not all mismatches may be important from a user perspective. Hence, it may be possible to simplify privacy notices further by displaying only a subset of mismatches. I hypothesize that the users' level of surprise and concern will vary among different mismatches. As I discussed in Section 5.1.3, mismatches can be of different types, for example, "Yes–No" and "No–Yes," and they may impact users' privacy to different extents. I hypothesize that users' will find "Yes–No" mismatches more privacy invasive than "No–Yes" mismatches. Further, they find it more important to know about "Yes–No" mismatches than "No–Yes" mismatches. As discussed in Section 5.1.3, we could treat a data practice that is"Unclear" as "Yes" and assume that a website engages in the data practice. However, we may not need to do that for all cases. I hypothesize that users would want to treat "Unclear" as "Yes" only in case of "Unclear–No" mismatch. Further, users want to treat "Unclear–No" as "Yes–No" only for sensitive data types such as financial and health information. Lastly, I hypothesize that users find mismatches in Likelihood expectations more important than mismatches in Desired expectations. Hence, privacy notices should focus on mismatches in Likelihood expectations than mismatches in Desired expectations.

5.4.1 Study Details

To understand the impact of mismatches on users, I propose to do a qualitative study. I will select an appropriate purposeful sampling strategy to recruit participants and interview them in an in-lab setting. I will elicit participant responses using scenario descriptions of mismatches identified from studies on Likelihood and Desired expectations described in Section 5.3. I may optionally do a quantitative online survey with larger number of participants to confirm findings from the qualitative study.

5.5. Implications for Privacy Notices

To understand the practical implications of studying expectations and mismatches, I will design and test privacy notices based on the results from the studies on understanding mismatches in user expectations. I hypothesize that showing only data practices that are not expected is more effective than showing all data practices. Because expectations and mismatches vary based on different factors, I will test notices tailored based on such factors. I hypothesize that showing only data practices relevant for a given context, for example website type, is more effective than showing all data practices under consideration. I hypothesize that personalizing notices by showing data practices based on user characteristics is more effective than notices that are not personalized. Because different mismatches can have different impact on users, I will test notices where data practices that are ranked more important are displayed more prominently. For example, a notice could display a subset of data practices that are deemed important and hide others. I hypothesize that highlighting data practices based on ranking is more effective than showing all data practices without using ranking.

5.5.1 Study Details

To evaluate privacy notices that highlight mismatches in privacy expectations, I propose to conduct a quantitative study. I will evaluate the effectiveness of privacy notice to communicate websites data practices in a succinct manner. I will use a browser plug-in developed as part of the Usable privacy Policy project[1] to display notices while participants browse websites. At the start of the study participants will be asked to download and install the browser plug-in. Participants will be randomly assigned to different groups. Based on the group, privacy notice will display a different set of data practices. In the control group, the notice will not display any data practice.

[1] www.usableprivacy.org

Website	Type	Subtype	Context	Rank
Webmd.com	Health	Reference	Private	107
Medhelp.org	Health	Reference	Private	2,135
Medlineplus.gov	Health	Reference	Government	558,671
Walgreens.com	Health	Pharmacy	Private	315
Bartelldrugs.com	Health	Pharmacy	Private	54,737
Mayoclinic.org	Health	Clinic	Private	297
Clevelandclinic.org	Health	Clinic	Private	2,629
Americanexpress.com	Finance	Credit	Private	76
Discover.com	Finance	Credit	Private	324
Bankofamerica.com	Finance	Bank	Private	33
Woodlandbank.com	Finance	Bank	Private	915,921
Banknd.nd.gov	Finance	Bank	Government	5,267
Paypal.com	Finance	Payment	Private	21
V.me	Finance	Payment	Private	27,289
Merriam-webster.com	Dictionary	–	Private	266
Wordnik.com	Dictionary	–	Private	8,412

Table 4. *Websites used in the study (Rank as of 3/10/2015).*

In the treatment group, a notice will display all data practices under consideration or a subset of data practices based on contextualization, personalization or ranking of mismatches. I will deploy the study as an online study. After participants have interacted with a website and its privacy notice, I will ask them questions to elicit their understanding of the website's data practices.

6. Completed Work

6.1. Mismatches in Likelihood Expectation Type

To understand mismatches in Likelihood expectation type, I have conducted an online study involving 16 websites and 240 participants [39]. I recruited 240 participants from Amazon Mechanical Turk [5] crowdsourcing website. I opted for a between-subjects design to prevent fatigue and learning effects. I asked participants to answer questions about one website randomly assigned to them. Website type (health, finance, dictionary) and popularity (low, high) were the main independent variables in the study, resulting in a 3x2 design with six conditions. In total, I studied 16 websites, listed in Table 4, across the three website types (7 Health, 7 Finance, 2 Dictionary). Fifteen participants were assigned to each website, resulting in the following number of participants per condition: 60 in Health-Low, 45 in Health-High, 60 in Finance-Low, 45 in Finance-High, 15 in Dictionary-Low, and 15 in Dictionary-High.

From the study, I have identified several mismatches in Likelihood privacy expectations regarding important collection, sharing and deletion data practices of websites. Mismatches identified include both "Yes–No" and "No–Yes" mismatches. I have identified that website type, which is a website characteristic, impacts users' expectations. It impacts expectations for financial and health information, but not contact and location information. I have also identified several user characteristics such as age and privacy concern that impact users' expectations. Results from the study confirm that user expectations vary based on contextual factors and by user characteristics. Results also show that the number of mismatches can be much smaller than the total number of data practices. Hence, highlighting or displaying only mismatches in a privacy notice can lead to shorter notices.

7. Research Plan

This section describes the work plan for this thesis.

7.1. Work Plan

The following table lists the tasks that make up the work plan for the thesis, including their current status, and the estimated duration in months. The total estimated time for completion is 16 months.

Table 5. *Thesis work plan*

Task	Status	Est. Time (months)
Understanding mismatches in Likelihood expectations	done	
Comparing mismatches in Likelihood and Desired expectations	to be done	5
Studying impact of mismatches on users	to be done	3
Validating privacy notices based on mismatches	to be done	5
Write dissertation	to be done	3

7.2. Discussion

My completed work provides evidence that the number of data practices that do not match user expectations can be smaller than the total number of data practices. Further, the results show that website type and several user characteristics are significantly predictive of users' Likelihood expectations. These results have practical implications for design of privacy notices. Highlighting mismatches can lead to shorter notices. It may also be possible to generate contextualized and personalized privacy notices.

- *Comprehension of simplified privacy notices:* Shorter notices by themselves may not improve comprehension. For example, it may improve comprehension among users who are privacy conscious, but not others. However, nudging users to become more privacy conscious is beyond the scope of this work. Users may make incorrect inferences from missing or hidden information in notices. It may be possible to address this with improved instructions and additional training. The risk of confusion and incorrect inferences is more likely the first time a user encounters a simplified privacy notice format. However, it may become easier to comprehend the notice with repeated use. If the study discussed in Section 5.5 does not show improvement in comprehension, I will try the following. First, I will investigate whether a short training phase will help improve comprehension. Second, I will run a longitudinal study to test whether repeated use improves comprehension.
- *Tailoring privacy notices:* Initial results showed that several user characteristics could significantly predict user expectations, predictive power of such factors was not very high. To tailor notices, we need sufficient predictive power. Low predictive power could lead to erroneous notices. In this thesis, I will use study whether, notices that are contextualized and personalized based on mismatches in users' expectations can be shorter and more comprehensible. However, whether we can develop models that have sufficiently high predictive power is not the focus of this work. I will, however, as part of this thesis, test whether more powerful statistical techniques such as factor analysis and clustering can improve predictive power.
- *Links to Software Engineering:* Privacy is often considered a non-functional requirement. In requirements engineering, which is one aspect of software development life cycle, one has to elicit requirements from stakeholders. If one is eliciting privacy requirements that deal with users' expectations, then it is important to understand the type of expectation being elicited. In other words, it is important to distinguish between different types of privacy expectations. Otherwise, the elicited requirements may not match users actual

requirements leading to development of software that will ultimately not meet users' requirements. As part of my thesis, I introduce the novel concept of different types of expectations. Although, I explore only two types of expectations, Likelihood and Desired, other types of expectations e.g. Minimally Acceptable may be relevant for software requirements. In this thesis, I outline approach to identify mismatches in expectations. A similar approach could be employed to identify mismatches in requirements of stakeholders and functionality of existing software tools and techniques.

References

[1] Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. EC '99*, pages 1–8. ACM, 1999. ISBN 1-58113-176-3.

[2] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, January 2015. ISSN 0036-8075, 1095-9203.

[3] Irwin Altman. The environment and social behavior: Privacy, personal space, territory, and crowding. 1975.

[4] Amazon. Alexa website rankings. http://www.alexa.com, 2015.

[5] Amazon. Mechanical turk. https://www.mturk.com/, 2015.

[6] Bettina Berendt, Oliver Günther, and Sarah Spiekermann. Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, 2005.

[7] Norman M. Bradburn, Seymour Sudman, and Brian Wansink. *Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires*. John Wiley & Sons, 2004.

[8] F.H. Cate. The Limits of Notice and Choice. *IEEE Security & Privacy*, 8(2):59–62, March 2010. ISSN 1540-7993.

[9] Center for Information Policy Leadership. Ten steps to develop a multilayered privacy policy, 2006.

[10] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. Use of a p3p user agent by early adopters. In *Proc. WPES 2002*, pages 1–10. ACM, 2002.

[11] Lorrie Faith Cranor, Kelly Idouchi, Pedro Giovanni Leon, Manya Sleeper, and Blase Ur. Are they actually any different? comparing thousands of financial institutions privacy practices. In *Proc. WEIS 2013*, 2013.

[12] Julia B. Earp, Annie I. Antón, Lynda Aiman-Smith, and William H. Stufflebeam. Examining internet privacy policies within the context of user privacy values. *Transactions on Engineering Management,*, 52(2):227–237, 2005.

[13] Serge Egelman and Eyal Peer. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2873–2882. ACM, 2015.

[14] Federal Trade Commission. Internet of things: Privacy & security in a connected world. FTC staff report, January 2015.

[15] Mary C. Gilly, William L. Cron, and Thomas E. Barry. The expectations-performance comparison process: An investigation of expectation types. In *Proc. Conference on Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, pages 10–16, 1983.

[16] Joshua Gomez, Travis Pinnick, and Ashkan Soltani. Know privacy. Technical report, UC Berkeley School of Information, 2009. http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf.

[17] Robin M. Hogarth. *Judgement and Choice: The Psychology of Decision*. John Wiley & Sons, 1987.

[18] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proc. CHI 2004*, pages 471–478. ACM, 2004. ISBN 1-58113-702-8.

[19] Adam N. Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B. Paine Schofield. Privacy, Trust, and Self-Disclosure Online. *HumanComputer Interaction*, 25(1):1–24, February 2010. ISSN 0737-0024.

[20] Ruogu Kang, Nathaniel Fruchter, Laura Dabbish, and Sara Kiesler. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Proc. SOUPS 2015*. USENIX, 2015.

[21] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A nutrition label for privacy. In *Proc. SOUPS 2009*. ACM, 2009.

[22] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What matters to users? factors that affect users' willingness to share information with online advertisers. In *Proc. SOUPS 2013*. ACM, 2013.

[23] Fei Liu, Rohan Ramanath, Norman Sadeh, and Noah A. Smith. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proc. COLING 2014*, pages 884–894, 2014.

[24] Yabing Liu, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. Analyzing facebook privacy settings: User expectations vs. reality. In *Proc. IMC 2011*, pages 61–70. ACM, 2011.

[25] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[26] GT Marx. Murky conceptual waters: The public and the private. *Ethics and Information technology*, pages 157–169, 2001.

[27] Aleecia M. McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *ISJLP*, 4, 2008.

[28] Aleecia M. McDonald, Robert W. Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. A comparative study of online privacy policies and formats. In *Proc. PETS 2009*, pages 37–55. Springer, 2009.

[29] John A. Miller. Studying satisfaction, modifying models, eliciting expectations, posing problems, and making meaningful measurements. *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, pages 72–91, 1977.

[30] George R. Milne and Shalini Bahl. Are there differences between consumers' and marketers' privacy expectations? a segment and technology level analysis. *Public Policy & Marketing*, 29(1), 2010.

[31] Helen Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79:119, 2004.

[32] Helen Nissenbaum. *Privacy in Context - Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009. ISBN 978-0804752367.

[33] 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. ISSN 1745-6606.

[34] Office of the Australian Information Commissioner. Community attitudes to privacy survey, 2013.

[35] Judith S. Olson, Jonathan Grudin, and Eric Horvitz. A study of preferences for sharing and privacy. In *Proc. CHI 2005*, pages 1985–1988. ACM, 2005.

[36] Irene Pollach. What's wrong with online privacy policies? *Commun. ACM*, 50(9):103–108, September 2007. ISSN 0001-0782.

[37] President's Concil of Advisors on Science and Technology. Big data and privacy: A technological perspective. Report to the President, Executive Office of the President, May 2014.

[38] Ashwini Rao, Florian Schaub, and Norman Sadeh. What do they know about me? contents and concerns of online behavioral profiles. In *Proc. PASSAT 2014*. ASE, 2014.

[39] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Under Review*, 2015.

[40] Joel Reidenberg, Aleecia M. McDonald, Florian Schaub, Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Faith Cranor, Fei Liu, Amanda Grannis, James T. Graves, et al. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30, 2015.

[41] 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, 2015.

[42] Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Aleecia M. McDonald, Joel R. Reidenberg, Noah A. Smith, Fei Liu, N. Cameron Russell, Florian Schaub, et al. The usable privacy policy project. Technical report, CMU-ISR-13-119, Carnegie Mellon University, 2013.

[43] Florian Schaub. *Dynamic Privacy Adaptation in Ubiquitous Computing*. PhD thesis, Ulm University, 2014.

[44] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Proc. SOUPS '15*, pages 1–17. USENIX, 2015. ISBN 978-1-931971-249. URL https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub.

[45] Florian Schaub, Bastian Könings, and Michael Weber. Context-adaptive privacy: Leveraging context awareness to support privacy decision making. *IEEE Pervasive Computing*, 14(1):34–43, 2015.

[46] John E. Swan and I. Frederick Trawick. Satisfaction related to predictive vs. desired expectations. *Refining Concepts and Measures of Consumer Satisfaction and Complaining Behavior*, pages 7–12, 1980.

[47] Donghee Yvette Wohn, Jacob Solomon, Dan Sarkar, and Kami E. Vaniea. Factors related to privacy concerns and protection behaviors regarding behavioral advertising. In *Proc. CHI 2015*, pages 1965–1970. ACM, 2015.

[48] Valarie A. Zeithaml, Leonard L. Berry, and Arantharanthan Parasuraman. The nature and determinants of customer expectations of service. *Academy of Marketing Science*, 21(1):1–12, 1993.

[49] Sebastian Zimmeck and Steven M. Bellovin. Privee: An architecture for automatically analyzing web privacy policies. In *Proc. USENIX Security 2014*, 2014.