Tabor, April

From: fschaub@cs.cmu.edu
Sent: Friday, October 09, 2015 3:15 PM
To: Tabor, April
Cc: ElectronicFilings
Subject: Re: File Request - PrivacyCon > Hyperlink #2 to Upload Confidential Submission

You have received 1 secure file from fschaub@cs.cmu.edu.
Use the secure link below to download.

Please find attached a research paper currently under submission to the ACM Conference on Human Factors in Computing (CHI 2016).

Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online

Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, Ruogu Kang
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 formed expectations to determine with which site they feel comfortable interacting. Mismatches between actual practices and a user's expectations may result in users exposing themselves to unanticipated privacy risks – even if the practices were disclosed in a privacy notice. One approach for mitigating these risks is to highlight elements of privacy notices that users do not likely expect. We present an approach for identifying such mismatches and analyze the results of a study based on this approach. Our findings suggest that focusing on mismatches could help design privacy notice interfaces that significantly reduce user burden.

If accepted, Ashwini Rao would like to present our work at PrivacyCon.

Contact details:
Ashwini Rao (arao@cmu.edu, 703.635.5077)
Florian Schaub (fschaub@cmu.edu, 412.478.2163)
Norman Sadeh (sadeh@cmu.edu, 412.580.2135)
Alessandro Acquisti (acquisti@andrew.cmu.edu)

Confidentiality request:
We request to keep our request for presentation confidential to avoid jeopardizing our anonymized CHI submission. The request for presentation and the draft paper can be made publicly available at the time of the PrivacyCon event in January 2016.

Thank you for your consideration,
Florian Schaub, Ashwini Rao, Alessandro Acquisti, Norman Sadeh, Ruogu Kang

Secure File Downloads:
Available until: **13 October 2015**

Click link to download:

expectations.pdf
474.41 KB

Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online

Ashwini Rao
Carnegie Mellon University
arao@cmu.edu

Florian Schaub
Carnegie Mellon University
fschaub@cs.cmu.edu

Norman Sadeh
Carnegie Mellon University
sadeh@cs.cmu.edu

Alessandro Acquisti
Carnegie Mellon University
acquisti@andrew.cmu.edu

Ruogu Kang
Carnegie Mellon University
ruoguk@andrew.cmu.edu

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 formed expectations to determine with which site they feel comfortable interacting. Mismatches between actual practices and a user's expectations may result in users exposing themselves to unanticipated privacy risks – even if the practices were disclosed in a privacy notice. One approach for mitigating these risks is to highlight elements of privacy notices that users do not likely expect. We present an approach for identifying such mismatches and analyze the results of a study based on this approach. Our findings suggest that focusing on mismatches could help design privacy notice interfaces that significantly reduce user burden.

Author Keywords

Privacy, expectations, contextual integrity, privacy policy.

ACM Classification Keywords

H.1.2. Models and Principles: User/Machine Systems—*Human Factors*; K.4.1. Computers and Society: Public Policy Issues—*Privacy*.

INTRODUCTION

Privacy policies serve as the primary mechanism for notifying users about a website's data practices, such as collection and sharing of personal information. However, website privacy policies, written in natural language, can be long, time consuming to read [24, 15], and difficult to understand for users [36, 32]. They are therefore often ignored by users [7, 33]. Prior work has proposed simplifying website privacy policies through summary notices that display data practices in an easy to understand visual format [8, 18, 44, 39]. 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 [37]. 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 [27, 28]. 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 [28, 23]. 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 [7]. 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 [39, 40] that should warn users about unexpected practices in particular [12]. The challenge, however, lies in identifying unexpected practices. Much work has studied users' privacy preferences in different contexts [31, 19]. However, privacy behavior differs from stated preferences [29] and they are not reliable for identifying mismatches between expectations, in the sense of "stated preferences," and actual site practices.

Main contributions: To advance toward more practical solutions that can impact privacy notice design, we outline a practical 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 mean at least four different things in consumers' context [25], but in the privacy context most work has focused on expectations in the desired sense or preferences [26, 19] or has not clarified the meaning of expectation [11, 14, 20]. We propose to elicit privacy expectations, in the sense of "expected frequencies," rather than privacy preferences and use them to identify mismatches in expectations. By focusing on expectations of what is happening, we avoid problems with unreliable subjective preferences of what should happen. We conducted a study to gain insights into users' mismatched privacy expectations concerning different types of websites. We identified practices that are unexpected by participants. Our 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.

From our results, we derive guidelines on what data practices are likely unexpected and should therefore be emphasized in privacy notices. We discuss the potential of contextualizing and personalizing privacy notices to provide privacy information most relevant to certain groups of users in a given transactional context. Our insights can benefit service providers. They can use our approach to identify data practices that users will likely not expect and that may therefore become cause for privacy concern. Service providers can 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. In addition, our results can inform the design of privacy services and tools, such as browser extensions, that aim to improve privacy transparency online.

BACKGROUND & RELATED WORK

Researchers have studied users' privacy preferences and willingness to share information in different contexts [31, 19]. 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 [29]. Observing privacy behavior is preferable, but it is challenging and resource-intensive to conduct behavioral studies at scale.

Privacy research has also explored the concept of expectations of privacy, including seminal work by Altman [3, 22], Marx [23] and Nissenbaum [27, 28]. For instance, Altman showed that individuals continuously modify their behavior to achieve an expected level of privacy [3], and Nissenbaum discusses how expectation of privacy can change based on context [27]. Privacy research typically differentiates between expected privacy and actual privacy, for example, Altman differentiates between desired and achieved levels of privacy [3].

However, in other domains, researchers have found that individuals have multiple levels or types of expectations [25, 13, 41, 43] and these types of expectations can impact constructs such as consumer satisfaction [41] and performance [13]. For instance, Miller proposed four expectation types: *Ideal, Expected, Minimum Tolerable,* and *Deserved* [25]. 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."

Based on Miller's work [25], we argue that people likely also have multiple levels of privacy expectations, beyond desired and achieved privacy. Therefore, we conceptually distinguish between the Expected ("will be") and Deserved ("should be") expectation types in measuring user expectations for website data practices, and focus on eliciting the Expected ("will be") type to identify mismatches.

We identify mismatches in user expectations regarding website data practices. We study if users expect that a website *will* collect, share or delete data. Prior work has studied mismatches in other types of expectations [11, 26, 14, 20]. To measure expectation, these studies either used an expectation type in the sense of desired preferences (*should*) [26], or they did not clarify the type of expectation [11, 14, 20]. Earp et al. compared what privacy-protective statements users expected websites' privacy policies to contain with statements in the policies [11]. Milne and Bahl examined differences between consumers' and marketers' expectations regarding use of eight information technologies [26]. Gomez et al. compared data practices of websites with data practices that users found concerning [14]. Liu et al. measured disparity between expected and actual Facebook privacy settings.

METHODOLOGY

Our goal is to identify mismatches between user expectations regarding website data practices and the practices website's disclose in their privacy policy. We defined expectation as what users think a website "will" do or is doing as opposed to what they prefer a website "should" do. We elicited user expectations for different online scenarios that varied in terms of data practices, website type, and other website characteristics, in order to understand the impact of contextual factors on privacy expectations. We also studied how user characteristics influence expectations. To identify unexpected practices, we compared elicited expectations with the data practices described in websites' privacy policies.

In the rest of this section, we describe the study design and parameters. The analysis of privacy policies and the procedure we used to identify and classify mismatched expectations are described in the next section.

Study Design

To assess the impact of different website scenarios on privacy expectations, we conducted an online study involving 16 websites and 240 participants. We opted for a between-subjects design to prevent fatigue and learning effects, in

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
Bankrd.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 1. Websites used in the study (Rank as of 3/10/2015).

which we 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 our study, resulting in a 3x2 design with six conditions. In total, we studied 16 websites, listed in Table 1, across the three website types (7 Health, 7 Finance, 2 Dictionary). Fifteen participants wer 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.

Survey Questionnaire

We designed 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. These website practices were treated as 17 dependent variables.

At the beginning of the survey, we explained the purpose of the study. We framed 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 [6]. We also did not mention privacy or data practices to avoid biasing participants. After explaining the purpose, we asked whether participants had visited or used the assigned website before. We then asked participants to familiarize themselves with the website for 2–3 minutes.

After they interacted with the website, we provided definitions of contact, financial, health and current location information. Next, we provided further contextualized by first showing them a scenario description, e.g.: "*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].*" We then asked 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 were also framed as opinion rather than knowledge questions [6], e.g., "*What is the likelihood that [website name] would collect your information in this scenario?*" Note that we framed the questions as "would collect" in order to capture participants' objective expectations. We provided a 4-point scale {Likely, Somewhat likely, Somewhat unlikely, Unlikely} as the response option. We wanted respondents' "best guess" and did not provide a neutral or not sure option. We did so 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. We asked an open-ended question to understand how they thought the website collected their information without having an account on the website. Then, participants answered questions regarding their expectations if having an account. When inquiring about sharing questions, we also asked 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, we asked participants whether they expected that the website would allow them to delete all, some or none of their data.

In the second part of the survey, we captured different user characteristics in order to study their impact on the participants' privacy expectations. We ordered these questions based on ease of answering, level of threat, and effect on subsequent answers [6]. First, we asked additional questions about their *past experiences* with the assigned website, such as the website's perceived trustworthiness. Participants then provided demographic information (gender, age, education, occupation) and whether they had a background in in computer-related fields, which may indicate a more accurate understanding of online data practices. We also asked for their U.S. state of residence, to assess whether privacy regulation on the state level, e.g., in California, impacts privacy expectations. We further included questions about privacy-protective behavior [30] and their familiarity and knowledge of privacy concepts and privacy-enhancing technologies [17]. We also asked about whether participants had negative online experiences [34] as they may expect data practices to be more privacy invasive. Lastly, we included the 10-item IUIPC scale [21] to assess online privacy concerns.

Study Deployment & Demographics

We received IRB approval for our study. Before deploying the study, we conducted multiple pilot interviews using think-aloud and verbal-probing [42] to assess the comprehensibility of our questions and refined the survey accordingly. We then deployed our questionnaire as an online survey in February 2015. We recruited 240 participants on Amazon Mechanical Turk [5]. Participants had to live in the United States, have at least a 95% approval rate and completed at least 500 tasks. Participants received $3.50 for completing the study. Each participant was randomly assigned to one of the 16 websites. We implemented our survey on SurveyGizmo and ensured that each participant could only take the survey once. To ensure data quality, we screened for participants that completed the study in less than 10 minutes (pilot tests suggested a 30-minute completion time), and checked whether participants consistently answered two questions about prior experience with the assigned website at the the beginning and end of the

survey. All participants passed at least two of three quality criteria.

The 240 participants completed our online survey in 22.5 minutes on average (SD = 12.8, median 18.6). The sample was 42% female and 58% male. The average age was 34.4 years (SD = 10.3, median 32). The majority (85.3%) had at least some college education and 61.6% reported an Associates, Bachelors or Graduate degree. A fifth of the participants (19.5%) had a college degree or work experience in a computer-related field. The top primary occupations were administrative staff (17.5%), service (14.1%), and business/management/financial (12%).

Scenario Parameters

We defined multiple scenarios that varied in key parameters, namely data practices and website characteristics. We hypothesized that these parameters may influence privacy expectations and mismatches.

Data Practices of Interest

We decided to focus on data practices concerning *collection, sharing and deletion of personal information* as prior research has shown that users are especially concerned about surreptitious collection, unauthorized disclosure and wrongful retention of personal information [37]. We considered the collection and sharing of four categories of privacy-sensitive information [1, 16, 19]: *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).

We further distinguished 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. In general, users may not be aware of implicit or automated data collection, e.g., of IP addresses and cookies. Websites may use IPs, email addresses and other information to acquire additional data about individuals, such as purchase history or interests, from social media services and data brokers [35].

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. We 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, we studied 17 data practices summarized in Table 2.

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. Summary of data practices.

Website Characteristics

To understand whether mismatched privacy expectations vary based on context, we considered three website characteristics: website type, popularity and ownership. *Website type* may influence what information users expect a website to collect [27]. We selected 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, we included banking, credit card and online payment websites. In the health category, we included 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, we included 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 [10]. Therefore, we studied 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. Our sample population was limited to the United States, and in the post-Snowden era, people may expect government websites to be more privacy invasive than private websites. Hence, we studied whether user expectations varied between government and privately-owned health and financial websites.

Identifying Mismatched Expectations

To identify mismatched expectations and therefore unexpected data practices, we aim to compare participants' expectations concerning a specific data practice with the results of the privacy policy. The information about a given website data practice extracted from the privacy the website's privacy policy, may be Yes, No, Unclear or Not addressed. We elicited an objective "will" expectation from

study participants. They rated their expectation of whether a website *will* engage in a specific data practice on a 4-point scale (Unlikely–1, Somewhat unlikely–2, Somewhat likely–3, Likely–4). These ratings can be interpreted as indications of a positive (Yes) or a negative (No) expectation that can be compared to the policy analysis results. Comparing a website's data practices and users' expectations this way, results in eight potential combinations, as shown in Table 3: For Yes–Yes and No–No, users' expectations match the websites' practices. Yes–No and No–Yes combinations constitute explicit 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 this study are 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 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.

STUDY RESULTS

To identify unexpected practices – those that did not match participants' privacy expectations – we first analyzed the privacy policies of the websites used in our study and then compared them to expectations elicited from study participants.

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

Table 3. 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 (?).

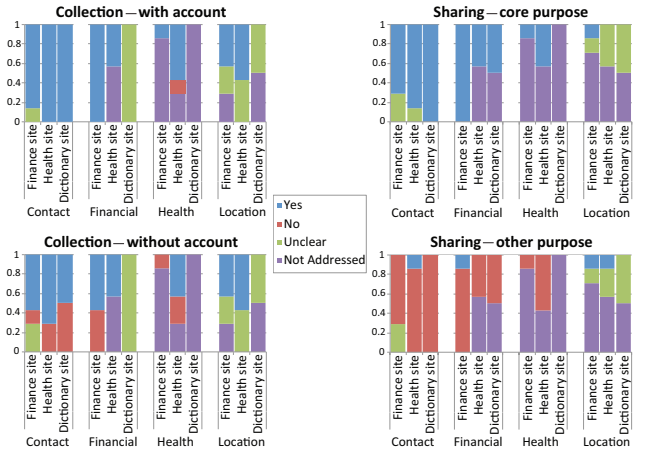


Figure 1. Collection and sharing data practices of the 16 websites used in our study, based on the analysis of the websites' privacy policies.

Website Privacy Policy Analysis

Two annotators, one with legal and another with privacy expertise, independently read each of the 16 privacy policies (cf. Table 1) and extracted the relevant collection, sharing and deletion data practices described earlier. Disagreements were resolved afterward. Following an annotation approach similar to Reidenberg et al. [36], annotators coded collection and sharing practices as *yes, no, unclear* or *not addressed*, in order to take ambiguity in the policy language (*unclear*) or silence on a specific practice (*not addressed*) into account. Collection and sharing practices were analyzed with regard to contact, financial, health and current location information, as well as for two collection contexts (with/without user account) and for two sharing purposes (core/other). Deletion practices were annotated as *full deletion* (websites allows deletion of all user data), *partial deletion* (deletion of only some data), *no deletion*, *unclear*, or *not addressed*.

Figure 1 gives an overview of data practices extracted from the privacy policies of the 16 websites (financial 7, health 7 and dictionary 2) used in our study. It shows the percentage of collection and sharing data practices that are clear, unclear or not addressed in the policies. We find that policies in all three website categories are mostly clear about practices concerning the collection or sharing of contact information, i.e., they make explicit statements about whether they collect or not

collect contact information and make clear statements about sharing (dominantly yes for core purposes; no for other purposes).

Not surprisingly, finance websites make explicit statements about collection and sharing of financial information. Note that credit card and online payment finance websites collect financial information even from non-registered users, for e.g. when users buy products, but banking websites do not do so. About half of the health websites' privacy policies also make explicit statements concerning financial information, however, the other half is silent on whether they collect or share financial information. Interestingly, the dictionary websites make statements that leave it unclear if they may collect financial information, but are either explicit or silent on sharing of financial information. Dictionary sites mention processing payments or posting transactions, but not explicit collection of financial information.

All dictionary websites and all but one of the financial websites do not address collection or sharing of health information. One of the finance websites, Bankofamerica is explicit about collecting health information from registered users and sharing it with third parties for core purposes. It does so from its insurance related affiliates, which may not be obvious to users. However, all but two of the health websites are explicit about whether they collect health information. Both the health clinic websites do not address collection of health information in their website privacy policy, but contain links to additional policies, which may disclose their collection practices. Health websites are less explicit about sharing of health information compared to collection of health information.

About half of financial and health websites are clear about collection of current location information, but none of the dictionary sites are clear. Almost all website privacy policies are unclear or silent on whether they share location information with third parties. Only one finance website explicitly states that it shares user location for core and other purposes. Only one health website explicitly states that it shares user location for other purposes, but it is unclear whether it shares it for core purposes.

Financial websites are more explicit about deletion data practices compared to health and dictionary websites. Nearly 71% (5) financial websites clearly disclose their practice in contrast to 50% (1) of dictionary and 28% (2) of health websites that do so. However, nearly half of the financial websites (3) do not allow any deletion of data and two allow partial deletion. In contrast, when clear about the practice, health websites (2) and dictionary websites (1) allow full deletion.

The privacy policy analysis shows that some data practices are common across different website types, whereas others are category-specific or even vary within a category. This suggests that if users use website characteristics to anchor their privacy expectations, these heuristics may lead to mismatches between their expectations and a website's stated data practices.

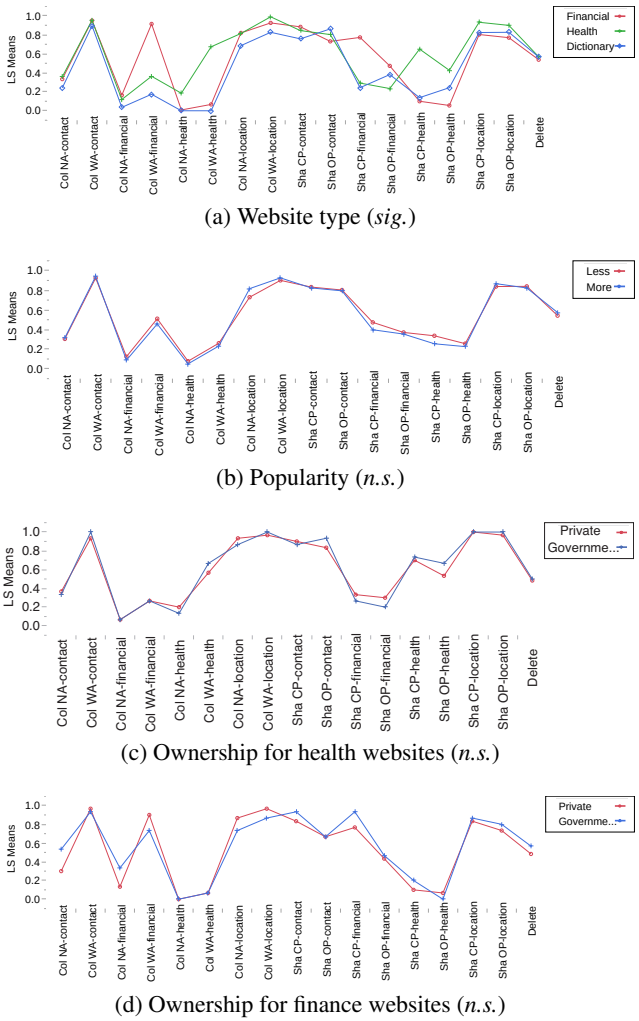
Impact of Website Characteristics

We find that a website's type has a significant impact on user expectations. This implies that what data practices users expect a website to engage in is influenced by what type of website it is. We did not find significant differences for popularity or ownership, suggesting they play no or a lesser role in shaping privacy expectations. For example, users expect different data practices from a finance website than from a health website, but have similar expectations for two finance websites, even if one of them is more popular than the other (e.g., in our dataset bankofamerica.com's popularity rank is 33; woodlandbank.com is ranked 915,921), or whether they are privately owned or government-operated. We describe our analysis in more detail in the following.

We used a mixed-model ANOVA to analyzed the impact of website type and popularity on user expectations. We considered website type (health, finance, dictionary) and popularity (high, low) as nominal between-subjects independent variables. We considered participant expectations concerning the 17 data practices as continuous repeated measures dependent variables (DV), which, as a group, measured users' overall expectation. We verified that the group of DVs has an approximate normal distribution with a normal-quantile plot of a linear combination of the individual DV scores. A Shapiro-Wilk W test showed only moderate departure from normality (W=0.988, p=0.041).

Results showed that interaction of website type and data practices was significant ($F(32.438)=12.819$, $p<0.0001$), see Figure 2a for an interaction plot. An interaction effect suggests that website type impacts what data practices users expect. Compare the impact of financial website type on users' expectations concerning collection of financial and health information from registered users (*COL WA-financial*), *COL WA-health*). Higher Least Square Means value implies that users are more likely to expect a data practice. Users expect financial websites to collect financial (high *LS Means*), but not health data (low *LS Means*). Figures further shows interactions of website popularity and ownership, which were not significant. Note that only the health and finance categories contained government-operated websites, dictionary websites are therefore not shown in the ownership plot.

We also studied the impact of website type on individual data practices. The distribution of values of individual data practices was non-normal. We treated them as two-level nominal variables and used a χ^2 statistical test. Figure 3a shows what information types participants expect websites to collect from registered users. If *LS Means*>0.5, users are likely to expect the data practice. Type of website has a significant impact for expectations of collection of financial ($\chi^2(2,N=240)=87.7$, $p<0.0001$, $R^2=0.302$) and health information ($\chi^2(2,N=240)=105.826$, $p<0.0001$, $R^2=0.3935$), but not for collection of contact and current location information. Whereas users expect the collection of contact and location information regardless of website type, the website type shapes their expectations concerning the collection of financial and health information. As shown in Figure 3c, participants are unlikely to expect websites to collect contact, financial and health data from users without



(a) Website type (*sig.*)

(b) Popularity (*n.s.*)

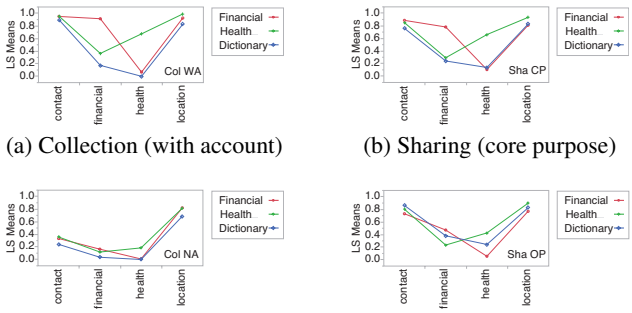
(c) Ownership for health websites (*n.s.*)

(d) Ownership for finance websites (*n.s.*)

Figure 2. Interaction of website characteristics and user expectations for the 17 data practices. Higher Least Square Means value implies users expect data practice to be more likely (Col: Collection, Sha: Sharing, WA: With Account, NA: No Account, CP: Core Purpose, OP: Other Purpose).

an account. Concerning expectations of data sharing, Figure 3b shows that participants likely expect websites, regardless of type, to share contact and current location information for core purposes. However, website type has a significant interaction effect for expectations of sharing financial ($\chi^2(2,N=240)=59.175$, $p<0.0001$, $R^2=0.1868$) or health information ($\chi^2(2,N=240)=77.935$, $p<0.0001$, $R^2=0.2642$) for core purposes. Figure 3d shows expectations of websites sharing for other purposes, In this case, regardless of website type, users expect websites to share contact and location data, but not financial and health data. Lastly, we did not find significant interactions of website type with participants expectations concerning websites' data deletion practices.

When user expectations can vary based on website type, mismatches that are specific to a website type are possible. For example, users do not expect financial websites to collect or share health information. However, one financial website in our study, Bankofamerica, collects health information when



(a) Collection (with account)

(b) Sharing (core purpose)

(c) Collection (no account)

(d) Sharing (other purpose)

Figure 3. Interaction of website type and expectations for specific data practices. Website type significantly interacts with user expectations for financial and health information. Higher Least Square Means value implies users are more likely to expect a data practice.

users have an account and shares it for core purposes, which violates user expectations.

Impact of User Characteristics

We analyzed the effect of multiple user characteristics on participants' data practice expectations. We find that privacy knowledge, privacy concern, age, trust in website, and recent have a significant impact on participants' expectations for certain data practices. Other user characteristics elicited in the survey had no statistically significant impact.

For analysis, we considered user characteristics as naturally-occurring, continuous IVs. The DVs were the user expectations for the 17 data practices. Distributions of the individual DVs were non-normal. Therefore, we considered them as two-level nominal variables (Yes, No) and built nominal logistic regression model for each DV. We assessed internal consistency of summated scale responses using Cronbach's α. For responses to online privacy concern, privacy concept familiarity, privacy knowledge, privacy protective behavior and negative online experience scales, reliability estimates were 0.88, 0.91, 0.63, 0.78, 0.68 respectively. For building regression models, we standardized IV values. To avoid biasing the model due to collinearity of IVs, we computed bivariate non-parametric Spearman rank correlations between IVs and subsequently excluded IVs that had moderate or higher correlation (>0.5). Our analysis of initial regression models showed that among demographic variables only age accounted for a significant amount of variance, therefore other demographics were removed to improve reliability of regression models. As a result, each of the 17 final regression models contained six IVs: privacy knowledge, privacy concern, negative online experience, age, trust in website and recent use. Table 4 lists the user characteristics (IV) and regression models in which the IV was statistically significant in predicting user expectation (DV).

Privacy Knowledge: An individual's privacy knowledge impacts user expectations. Specifically, privacy knowledge can impact if a user expects a website to collect health information from unregistered users. An individual who scores one unit higher on the privacy knowledge scale is two times more

likely to not expect that a website will collect health information.

Privacy Concern: Individuals with higher online privacy concern (IUIPC [21]) expect data practices to be more privacy invasive. Specifically, individuals with one unit increase in online privacy concern are twice as likely to expect that a website will collect current location information. They are ∼1.6 times more likely to expect that a website will share contact and current location information for core purposes.

Age: Individuals' age impacts expectations regarding deletion. With one year increase in age, they are ∼1.8 times more likely to expect that a website will not allow deletion of user data.

Trust in Website: User perception of a website's trustworthiness impacts expectations regarding sharing and deletion data practices. With one unit increase in trust, individuals ∼1.7 times more likely to expect that a website will not share health and financial information for other purposes. They are 1.5 times more likely to expect that a website will share location information for core purposes. Lastly, individuals are twice as likely to expect the website to allow deletion of user data.

Recent Use: Participants self-reported use of the website in the last 30 days impacts expectations regarding three data practices. With one unit increase in usage, individuals are 1.6 times more likely to expect that a website will not collect current location information from registered users. Individuals are 1.5 times more likely to expect that the website will not share contact information for core purposes. Lastly, individuals are 1.6 times more likely to expect that website will not allow deletion.

Since expectations vary based on user characteristics, mismatches can also vary. For example, with increase in age and recent use, users correctly expect websites not to permit deletion of user data. Hence, the likelihood of mismatch is higher in case of younger users as well as relatively new users of a website.

Overall Matched and Mismatched Expectations

As shown in Figure 4, expected and unexpected data practices varied for different information types, and collection and sharing scenarios. We analyzed mismatches when websites explicitly disclosed their data practices as well as when websites were unclear or did not address the data practices. When data practices were explicit, there were three important mismatches. Collection of contact information without an account was mainly a Yes–No mismatch, that is, users did not expect websites to collect information but websites did. Similarly, collection of financial information without an account was a Yes–No mismatch. Sharing of contact information for other purposes was also a mismatch, but a No–Yes mismatch, that is, users pessimistically and incorrectly thought that websites would share information. For the remaining data practices, either users expectations predominately matched website practices or the level of match was equal to the level of mismatch.

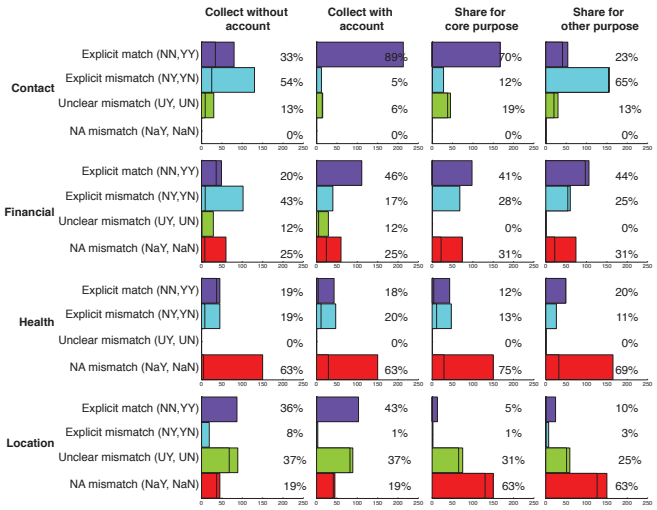


Figure 4. Matches and mismatches in user expectations. Explicit match or mismatch occurs when websites are clear about their data practice. When practice is unclear or not addressed, mismatch is not evident.

For deletion data practice, 32% of users expected websites to allow full deletion, but only 19% allow it. Similarly, 48% expected partial deletion and 12% of permit it. However, about 20% of the participants thought that websites would not allow deletion of any data and 19% of the websites do not allow deletion of any data. User expectations were similar across the three types of websites. There is a mismatch in user expectations regarding deletion, and they seem to expect websites to allow deletion more than websites actually do.

As we discussed earlier, the number of data practices that are unclear or not addressed in a privacy policy can be high. As seen from Figure 4, websites mostly do not address data practices regarding health information. In contrast, they are mostly unclear or do not address data practices regarding location information. Considering Yes–No mismatches to be more privacy invasive, let us assume that a website engages in a data practice when its disclosure is unclear or not addressed. For health information practices, this results in mainly Yes–No mismatches for all scenarios. However, for location information practices, it results in No–Yes mismatches.

DISCUSSION

We identified data practices that do not match user expectations. Our results show that the number of mismatches can be substantial depending on the data practice, and that mismatched expectations vary significantly based on the type of website, as well as user characteristics, such as privacy concern, knowledge, and age. Below, we discuss potential limitations of our study, followed by implications of our results.

Limitations

We conducted an online study to elicit user expectations, and we could benefit from additional in-lab studies conducted under more controlled conditions. We compared user expectations with websites' data practices, as disclosed in websites'

User characteristic (IV)	User expectation (DV)	Model			IV		
		R^2	$\chi^2(6, N{=}240)$	p	Odds(No)	$\chi^2(1, N{=}240)$	p
Privacy knowledge	Collect health info without account	0.10	14.52	0.024	2.09	7.60	0.0058
Privacy concern	Collect location info with account	0.13	13.80	0.0319	0.49	7.22	0.0072
	Share contact info for core purpose	0.09	18.47	0.0052	0.64	5.94	0.0148
	Share location info for core purpose	0.08	15.34	0.0177	0.58	7.67	0.0056
Age	Allow deletion	0.13	30.53	<0.0001	1.77	10.88	0.0010
Trust in website	Share location info for core purpose	0.08	15.34	0.0177	0.65	4.44	0.0352
	Share financial info for other purpose	0.07	21.33	0.0016	1.80	16.82	<0.0001
	Share health info for other purpose	0.05	14.54	0.0241	1.68	1124	0.0008
	Allow deletion	0.13	30.53	<0.0001	0.53	13.64	0.0002
Recent use	Collect location info with account	0.13	13.80	0.0319	1.56	4.01	0.0451
	Share contact info for core purpose	0.09	18.47	0.0052	1.50	6.67	0.0098
	Allow deletion	0.13	30.53	<0.0001	1.56	7.83	0.0051

Table 4. Regression models in which specific user characteristics (IV) significantly impact user expectations (DV). *Odds(No)* indicates, for one unit increase in the IV value, the increase in likelihood that a user will not expect a website to engage in that data practice (*Odds(Yes)*=1 / *Odds(No)*).

privacy policies; how a website actually handles personal information of their users could potentially be different, but this is very difficult to assess in practice.

We recruited participants from Amazon Mechanical Turk. Compared to the general population, they may have more computer knowledge and exposure to privacy-related surveys. Our participants were limited to the United States, and it would be interesting to study expectations of users in other countries or cultures.

We studied collection, sharing and deletion data practices. We asked participants ($n = 240$) if they wanted to know about other data practices; nearly half did not (47.5%). Among the rest, the top three requests were as follows: Participants (14%) wanted additional details about sharing. They wanted to know with whom – partners, affiliates and third-parties – their data was being shared. They wanted to know about data security (12%) and how long their data was retained (7%). We plan to extend our research to cover these and other data practices of interest in the future.

We further plan to study more website categories. Eliciting user expectations for categories with broad or multiple purposes, for example search or social networking categories, could be challenging. For example, users may use Google.com for searching, shopping, directions etc. We are further interested in studying the impact of additional expectation types, such as the "should" (Ideal) expectation type.

Highlighting Unexpected Practices

Our goal is to develop simplified presentations of website privacy notices that can help users understand website data practices. As our results suggest, the number of mismatches is small compared to all of a website's data practices. Thus, focusing on information about likely unexpected data practices could reduce the amount of information that a user has to process. Shorter, user-facing privacy notices [39] could emphasize unexpected practices in addition to a comprehensive privacy policy. As we discussed earlier, different types of mismatches (Yes–No vs. No–Yes) could have different consequences on user privacy, and solutions that highlight mismatches need to consider that.

Existing solutions for simplifying privacy notices, for example nutrition labels [18], although an improvement over privacy policies, are themselves too complex. By using models of people's privacy expectations, we could selectively highlight or display those elements of privacy labels likely to be most relevant to a user. Our results suggest that only a fraction of privacy nutrition labels would need to be shown to properly inform users. However, the effectiveness of such highlighting has to be tested with end users.

Organizations could obtain a competitive advantage by making their website's data practices and privacy policies easier to understand. In the past, organizations such as Google have tried to organize information within their policy, along dimensions that are important to people, with the intent of making information easier to access. Mismatches in expectations are important, and highlighting them can aid in such efforts. 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.

Although organizations could themselves generate simplified notices, the low adoption of simplified and standardized notices mechanisms [9], such as P3P for making privacy policies machine-readable, indicates that they may not do so. An alternative approach is for a third-party to highlight unexpected data practices based on mismatched expectations. For example, a browser plug-in could generate and display a simplified notice. For instance, using color, a browser plug-in could highlight snippets of text from the natural language privacy notice corresponding to mismatched data practices. Currently third-party browser plug-ins, such as Ghostery[1] and Privacy Badger,[2] generate and display information regarding online tracking practices. Similarly, a third-party browser plug-in could display information regarding unexpected data practices. Plug-ins could use just-in-time notifications or

[1] www.ghostery.com

[2] www.eff.org/privacybadger

static icons that users can click to gain more information. At installation time, the plug-in could gather user characteristics such as privacy knowledge, concerns and demographics. In order to scale up, we could extract data practices disclosed in policies using techniques that combine crowdsourcing, machine learning and natural language processing [38, 44].

Generating Simplified Notices

We could potentially simplify privacy notices by highlighting data practices that do not match user expectations. For example, consider Bankofamerica privacy policy, which is one of the 16 policies in our study. A full website privacy notice has to include information about all the 17 data practices that we studied. However, for six data practices, user expectations match the website's data practices. Hence, if the notice displays only mismatches, it has to highlight 11 data practices, which is 35% less information. We could further simplify the notice by prioritizing the impact of mismatches. For example, if we determine that Yes–No mismatches are more concerning to users than No–Yes mismatches, the notice could highlight five Yes–No mismatches among the 11 mismatches, which results in 70% less information.

Our results indicate that the data practices users expect, as well as respective mismatched expectations, vary significantly by website type. For example, users expect health websites to collect health information, but not finance websites. Therefore, website type could serve as a simple and practical feature to contextualize privacy notices in order to highlight those practices unexpected for the respective website type. Third party tools or plug-ins could further predict based on website type, which data practices my be unexpected and emphasize or warn about them. Practices that are likely expected for websites of a given type, do not require warnings. For example, in case of the Bankofamerica banking website, the plug-in can signal a mismatch as the website collects health information. However, the plug-in need not signal a mismatch in case of a health website that collects health information.

User expectations and mismatches vary based on user characteristics. Hence, we could personalize privacy notices based on user characteristics. For example, younger users are significantly more likely to expect a website to allow deletion of user data. Hence, when the website does not allow deletion, the likelihood of a mismatch is higher in case of younger users. Thus, privacy decisions support tools could highlight a mismatch for younger users only.

Semantics and Impact of Mismatches

We discussed mismatches concerning "will" expectations, corresponding to Miller's "Expected" expectation type [25]. We can extend our analysis to additionally include "should" expectations, which are more subjective, as they describe expectations of what would be "Ideal" [25], and are therefore closer to preferences of desired privacy. Users may answer Yes or No to whether a website *should* engage in a data practice. Considering "should" expectations in addition to "will" expectations, would add 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, we need to consider which mismatches matter to 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.

CONCLUSION

We identified mismatches in user expectations regarding online data practices. Further, we identified factors that impact such mismatches. We believe that emphasizing such mismatches in privacy notices could help users make better privacy decisions. Further, given the small number of mismatches compared to the overall number of data practices, it could be possible to generate simplified user-facing privacy notices that contain much less information than full privacy policies. Based on the factors that impact mismatches,

we identified future research opportunities for contextualizing and personalizing privacy notices and privacy tools to ameliorate the effect of mismatched expectations

ACKNOWLEDGMENTS

This research has been partially funded by the National Science Foundation under grants CNS 10-1012763 and CNS 13-30596.

REFERENCES

1. Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. 1999. Privacy in e-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. EC '99*. ACM, 1–8.
2. Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (Jan. 2015), 509–514.
3. Irwin Altman. 1975. The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding. (1975).
4. Amazon. 2015a. Alexa Website Rankings. http://www.alexa.com. (2015).
5. Amazon. 2015b. Mechanical Turk. https://www.mturk.com/. (2015).
6. Norman M. Bradburn, Seymour Sudman, and Brian Wansink. 2004. *Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires*. John Wiley & Sons.
7. F.H. Cate. 2010. The Limits of Notice and Choice. *IEEE Security & Privacy* 8, 2 (March 2010), 59–62.
8. Center for Information Policy Leadership. 2006. Ten Steps to Develop a Multilayered Privacy Policy. (2006).
9. Lorrie Faith Cranor. 2012. Necessary but Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice. *Journal on Telecommunications and High Technology Law* 10 (2012), 273.
10. Lorrie Faith Cranor, Kelly Idouchi, Pedro Giovanni Leon, Manya Sleeper, and Blase Ur. 2013. Are they actually any different? Comparing thousands of financial institutions privacy practices. In *Proc. WEIS 2013*.
11. Julia B. Earp, Annie I. Antón, Lynda Aiman-Smith, and William H. Stufflebeam. 2005. Examining Internet Privacy Policies within the Context of User Privacy Values. *Transactions on Engineering Management,* 52, 2 (2005), 227–237.
12. Federal Trade Commission. 2015. *Internet of Things: Privacy & Security in a Connected World*. FTC staff report.
13. Mary C. Gilly, William L. Cron, and Thomas E. Barry. 1983. The Expectations-Performance Comparison Process: An Investigation of Expectation Types. In *Proc. Conf. Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*. 10–16.
14. Joshua Gomez, Travis Pinnick, and Ashkan Soltani. 2009. *Know Privacy*. Technical Report. UC Berkeley School of Information. http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf.
15. Carlos Jensen and Colin Potts. 2004. Privacy Policies As Decision-Making Tools: An Evaluation of Online Privacy Notices. In *Proc. CHI '04*. ACM, 471–478.
16. Adam N. Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B. Paine Schofield. 2010. Privacy, Trust, and Self-Disclosure Online. *HumanComputer Interaction* 25, 1 (Feb. 2010), 1–24.
17. Ruogu Kang, Nathaniel Fruchter, Laura Dabbish, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Proc. SOUPS '15*. USENIX.
18. Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A Nutrition Label for Privacy. In *Proc. SOUPS '09*. ACM.
19. Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. 2013. What Matters to Users? Factors that Affect Users' Willingness to Share Information with Online Advertisers. In *Proc. SOUPS '13*. ACM.
20. Yabing Liu, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing Facebook Privacy Settings: User Expectations vs. Reality. In *Proc. IMC '11*. ACM, 61–70.
21. Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (2004), 336–355.
22. Stephen T. Margulis. 2003. On the Status and Contribution of Westin's and Altman's Theories of Privacy. *Journal of Social Issues* 59, 2 (June 2003), 411–429.
23. GT Marx. 2001. Murky conceptual waters: The public and the private. *Ethics and Information technology* (2001), 157–169.
24. Aleecia M. McDonald and Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. *ISJLP* 4 (2008).
25. John A. Miller. 1977. Studying Satisfaction, Modifying Models, Eliciting Expectations, Posing Problems, and Making Meaningful Measurements. *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction* (1977), 72–91.
26. George R. Milne and Shalini Bahl. 2010. Are there Differences between Consumers' and Marketers' Privacy Expectations? A Segment and Technology Level Analysis. *Public Policy & Marketing* 29, 1 (2010).
27. Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79 (2004), 119.

28. Helen Nissenbaum. 2009. *Privacy in Context - Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
29. Patricia A. Norberg, Daniel R. Horne, and David A. Horne. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs* 41, 1 (2007), 100–126.
30. Office of the Australian Information Commissioner. 2013. Community Attitudes to Privacy Survey. (2013).
31. Judith S. Olson, Jonathan Grudin, and Eric Horvitz. 2005. A Study of Preferences for Sharing and Privacy. In *Proc. CHI '05*. ACM, 1985–1988.
32. Irene Pollach. 2007. What's Wrong with Online Privacy Policies? *Commun. ACM* 50, 9 (Sept. 2007), 103–108.
33. President's Concil of Advisors on Science and Technology. 2014. *Big Data and Privacy: A Technological Perspective*. Report to the President. Executive Office of the President.
34. Lee Rainie, Sarah Kiesler, Ruogu Kang, and Mary Madden. 2013. Anonymity, Privacy, and Security Online. http://pewinternet.org/Reports/2013/Anonymity-online.aspx, *PEW Research Center* (September 2013).
35. Ashwini Rao, Florian Schaub, and Norman Sadeh. 2014. What do they Know about me? Contents and Concerns of Online Behavioral Profiles. In *Proc. PASSAT '14*. ASE.
36. Joel Reidenberg, Aleecia M. McDonald, Florian Schaub, Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Faith Cranor, Fei Liu, Amanda Grannis, James T. Graves, and others. 2015a. Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding. *Berkeley Technology Law Journal* 30, 1 (2015), 39–88.
37. Joel R. Reidenberg, N. Cameron Russell, Alexander J. Callen, Sophia Qasir, and Thomas B. Norton. 2015b. Privacy Harms and the Effectiveness of the Notice and Choice Framework. *ISJLP* 11 (2015).
38. 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, and others. 2013. *The Usable Privacy Policy Project*. Technical Report. CMU-ISR-13-119, Carnegie Mellon University.
39. Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015a. A Design Space for Effective Privacy Notices. In *Proc. SOUPS '15*. USENIX, 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub
40. Florian Schaub, Bastian Könings, and Michael Weber. 2015b. Context-Adaptive Privacy: Leveraging Context Awareness to Support Privacy Decision Making. *IEEE Pervasive Computing* 14, 1 (2015), 34–43.
41. John E. Swan and I. Frederick Trawick. 1980. Satisfaction Related to Predictive vs. Desired Expectations. *Refining Concepts and Measures of Consumer Satisfaction and Complaining Behavior* (1980), 7–12.
42. Gordon B. Willis. 2004. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications.
43. Valarie A. Zeithaml, Leonard L. Berry, and Arantharanthan Parasuraman. 1993. The Nature and Determinants of Customer Expectations of Service. *Academy of Marketing Science* 21, 1 (1993), 1–12.
44. Sebastian Zimmeck and Steven M. Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *Proc. USENIX Security '14*.