Jenny Tang, Hannah Shoemaker, Ada Lerner, and Eleanor Birrell*

Defining Privacy: How Users Interpret Technical Terms in Privacy Policies

Abstract: Recent privacy regulations such as GDPR and CCPA have emphasized the need for transparent, understandable privacy policies. This work investigates the role technical terms play in policy transparency. We identify potentially misunderstood technical terms that appear in privacy policies through a survey of current privacy policies and a pilot user study. We then run a user study on Amazon Mechanical Turk to evaluate whether users can accurately define these technical terms, to identify commonly held misconceptions, and to investigate how the use of technical terms affects users' comfort with privacy policies. We find that technical terms are broadly misunderstood and that particular misconceptions are common. We also find that the use of technical terms affects users' comfort with various privacy policies and their reported likeliness to accept those policies. We conclude that current use of technical terms in privacy policies poses a challenge to policy transparency and user privacy, and that companies should take steps to mitigate this effect.

Keywords: privacy policies, policy transparency

DOI 10.2478/popets-2021-0038
Received 2020-11-30; revised 2021-03-15; accepted 2021-03-16.

1 Introduction

Privacy policies are a cornerstone of current online privacy practices: companies describe their data collection and data use practices in a publicly-available document, and users who proceed to use a service are presumed to have consented to the described practices. However, fifteen years of critiques have conclusively demonstrated that these privacy policies fail to provide adequate privacy protections [2, 17, 23, 37, 39, 41, 42].

Jenny Tang: Wellesley College, E-mail: jtang4@wellesley.edu
Hannah Shoemaker: Pomona College, E-mail: hnsa2018@mymail.pomona.edu
Ada Lerner: Wellesley College, E-mail: alerner@wellesley.edu
***Corresponding Author: Eleanor Birrell:** Pomona College, E-mail: eleanor.birrell@pomona.edu

Recent legal efforts such as the EU's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) are attempts to enhance user privacy. In addition to providing affirmative privacy rights—such as the GDPR's right of access and right of erasure and the CCPA's right to opt-out of sale and right to delete—the new privacy laws impose transparency requirements for disclosures. GDPR requires that information about data collection and use be communicated "in a concise, transparent, intelligible and easily accessible form, using clear and plain language" [12] and the current CCPA regulations require that this information be "written in a manner that provides consumers a meaningful understanding of the information being collected" [28]. However, recent work has found that privacy policies are still written at a high reading level [18], and a large-scale, longitudinal comparison of privacy policies in the EU pre- and post-GDPR found that privacy policies increased in length without demonstrating improvements in sentence structure complexity [22].

In this work, we investigated the comprehensibility of privacy policies from a different perspective: we evaluated how well users understand technical terms that appear in privacy policies, and we explored how users' misunderstandings impact their comfort with various data use policies.

Drawing on both manual and automated surveys of privacy policies for popular sites and apps, we developed a list of 57 technical terms that appear in privacy policies. We ran a qualitative pilot on Amazon Mechanical Turk in which we asked users to define these technical terms in their own words. Based on these responses, we identified 20 technical terms that we suspected to be commonly misunderstood by Internet users; we also included 2 terms that were generally well-understood. We then constructed a quantitative survey in which we tested whether respondents correctly understood each technical term individually; we also compared how comfortable users were with data use practices described using technical terms to how comfortable they were with the same practices described in non-technical, explanatory language.

We found that misconceptions and misunderstandings about technical terms were pervasive. For all twenty technical terms that we hypothesized were commonly misunderstood, fewer than 60% of users could correctly define the term; for nine of the technical terms, less than 30% of users could correctly define the term. Certain misconceptions were common. For example, 28% of users believed that *device fingerprinting* meant the use of a fingerprint as a password for a device, 23% of respondents believed that a *pixel tag* was a way to identify an image, and 21% of respondents believed that a company could read or decrypt their messages if its application used end-to-end encryption.

We also found that the use of technical terms affected how comfortable users were with a privacy policy; depending on the underlying misconceptions, users could be either more comfortable or less comfortable with the policy containing the technical term than with the equivalent policy that used non-technical, explanatory language. For example, users were less comfortable with a policy that listed specific types of metadata that would be collected than with a policy that used the technical term *metadata*. On the other hand, users were more comfortable with a policy that explained that information about them would be stored on their machine or device than with a policy that stated that it used *browser storage* to store information about them. Overall, we found statistically significant differences in users' comfort level between the policies that contained technical terms and the policies that contained explanatory non-technical language for 60% of the commonly-misunderstood technical terms we considered. These results suggest that consent to a policy that contains technical terms might not imply comfort with—or informed consent to—the data use practices described by that policy.

We view the primary contributions of this work as three-fold:

1. We provide the first quantitative evaluation of how well users understand technical terms used in privacy policies.
2. We identify common misconceptions and identify terms that are broadly misunderstood.
3. We demonstrate that the use of technical terms affects users' comfort with particular data use practices.

Based on our results, we believe that further attention must be paid to the use of technical terms in privacy policies and other user-facing documentation, and that dependence on commonly misunderstood technical terms should be minimized wherever possible to enhance transparency and empower informed consent.

2 Related Work

Although this is the first work to specifically examine misconceptions about technical terms used in privacy policies, prior work has previously explored user misconceptions in other contexts and has evaluated how well users understand privacy policies more broadly.

2.1 Technical Term Misconceptions

User studies have consistently found that users misunderstand the technical term *privacy policy*. A 2005 study about the online shopping behavior of American consumers found that 75% falsely believed that the presence of a privacy policy meant that a website would not share their information with other websites or companies [41]. A 2014 survey by the Pew Research Center [38] and a subsequent longitudinal study [42] found that a majority of Internet users continue to hold this misconception. Our user study confirmed that this misconception continues to be prevalent among American Internet users.

Felt et al. [10] explored how well users understand Android permissions, a particular class of technical terms. Overall, they found that users answered 21% of permission comprehension questions correctly, and just 2.6% of respondents answered all three questions correctly. Common misconceptions included that the permission for *full network access* did not allow an app to load ads (39.5% of respondents), that permission to *read phone state* did not allow an app to track you across applications (59.3%), and that permission to *read contact data* did not allow an app to read your call history (60.6%). Overall, they found that many users could not connect the resource-specific technical terms used in permission names to particular risks that would be enabled by those permissions, and that significant work would be required to make the Android permission system widely accessible.

2.2 Privacy Policy Transparency

A large body of prior work has investigated the broader problem of privacy policy transparency and how well users understand privacy policies.

Readability

One line of work performs automated analysis of privacy policy text to evaluate readability using standardized metrics such as Flesch Reading Ease [11]. These studies have consistently found that privacy policies are difficult to comprehend and are often written at a level that surpasses the educational levels of large swaths of the public they're intended to inform [1, 3–5, 8, 9, 13, 14, 16, 18, 21, 24, 25, 34, 37]. Studies that looked at healthcare privacy policies found that none of the policies examined were readable by a majority of English speaking Americans [13] and that on average 80% of the people living in areas surrounding the hospitals whose privacy policies were studied were not at the reading level required by these policies [3]. Other work has studied the readability of financial privacy policies [1, 14, 21]. Recently and most comprehensively, Fabian et al. [9] studied the privacy policies of 50,000 English-speaking websites, finding that these policies were, on the whole, difficult to comprehend, requiring a reader "to have graduated from a high school or some college, having completed about 13 study years or 16 formal educational years."

Writing from the perspective of critical linguistics, Pollach [29, 30] found that the policies studied used linguistic techniques such as euphemisms, passive verbs and nouns, and modality markers such as "may" to obscure the meaning of their policies and verbally mitigate their role in data collection and dissemination. Hochhauser [14] studied specific style and organizational elements of 60 financial privacy policies that might make them difficult for readers to comprehend, finding that the majority of these policies had poor writing styles, were full of uncommon words, and often had difficult-to-read fonts, font sizes, and small margins; they conclude that the policies studied were not reasonably understandable, and would be illegible for some readers.

User Comprehension

An alternative approach to evaluating policy transparency is to evaluate user comprehension through user studies. This line of work has consistently found that privacy policies are difficult for the public to understand [40, 41]. Even when respondents had the proper education level to understand privacy policies, they still had poor comprehension of these policies. Proctor et al. [31] found that the respondents they surveyed were only able to answer approximately 50 percent of comprehension questions about privacy policy practices despite these policies being written at their education level. A 2007 study conducted by Vu et al. [43] found similar results, with participants at the reading level required by the policies displaying poor overall understanding of their contents.

Policy Specificity.

Another line of prior work focuses specifically on studying how factors such as vague wording, lack of context, ambiguous words and phrases, and internal contradictions contribute to a lack of reader comprehension [4, 19, 33, 35, 36]. Kumar [19] studied 23 policies from major telecommunications companies, finding that "vague or unclear language hinders comprehension of company practice" and inhibits users from making informed choices about whether or not to engage in business with a company. Reidenberg et al. [33] studied how ambiguity in key privacy policy terms such as *core services* contributes to privacy policy misunderstandings and misconceptions. They found that non-experts often misinterpreted policy silence on a specific practice as meaning that the policy was unclear on whether or not the practice was permitted. Experts, on the other hand, more frequently interpreted policy silence correctly, as meaning that the practice was permitted.

3 Methodology

To understand how users interpret technical terms that appear in privacy policies, we ran a pilot study to identify commonly misunderstood technical terms followed by a large-scale, quantitative user study to explore how people interpret those terms and how any misunderstandings affect users' level of comfort with data use practices described in privacy policies. We concluded our work by running a small-scale follow-up study to validate the relationship between comfort with and acceptance of privacy policies. For the pilot, the full study, and the follow-up study, we recruited respondents from Amazon Mechanical Turk; participation was restricted to workers located in the United States with at least 50 hits and at least a 95% approval rate. Each study

included an open-ended attention check question. Responses that were nonsensical or unrelated to the question were rejected; participants who successfully completed the attention check question were compensated based on an estimated survey completion time at a prorated rate of $12 dollars per hour. These studies received a waiver from each of the authors' institutional ethics review board (IRB). Participants were informed about our data storage and data use practices in advance, and no personally-identifiable information was collected.

3.1 Pilot Study

To develop our pilot study, we performed a manual analysis of the Alexa top 10 U.S. websites as of June 2020 and select apps on the Android Play Store. For each study, two authors independently read through the policy and identified technical terms that appeared in that policy. We also added a few terms that we believed to be broadly misunderstood from prior work (e.g., *privacy policy* [41]) or from anecdotal personal experience (e.g., *encryption* and *aggregation*). We then discussed the list as a group and excluded any terms that were not deemed to be technical terms by consensus. This manual process resulted in a preliminary list of 57 technical terms (listed in Appendix A.1).

We validated this dictionary of terms with an automated analysis of a corpus of 3609 English-language privacy policies [22]. Using the porter stemmer, we produced frequency counts for unigrams, bigrams, and trigrams that appeared in the corpus. For 20 of the terms in our preliminary list, that term or the related *root term* appeared in 500 or more unique privacy policies; for 39 of the terms, that term or a related technical term appeared in 100 or more unique privacy policies. All technical terms that appeared in 500 or more policies were included in our pilot study.

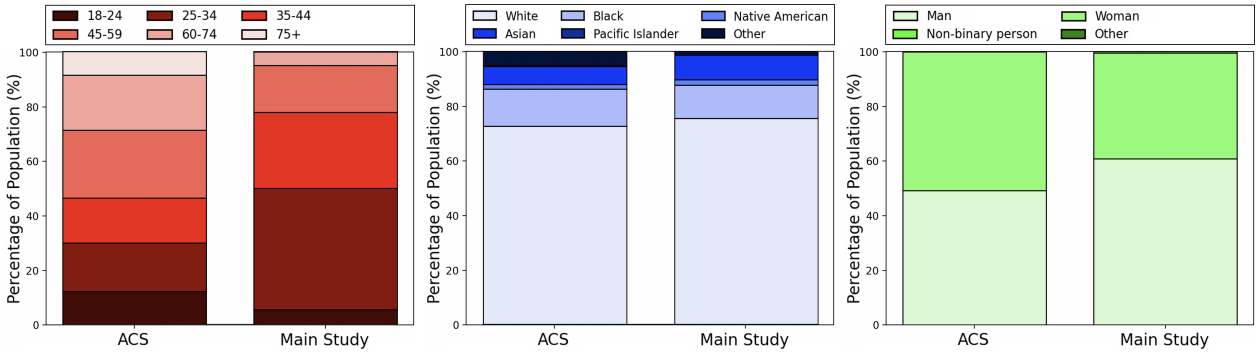
To ensure that the pilot survey was a reasonable length, we divided the 57 technical terms into 11 categories; each pilot participant was randomly shown one question from each category. For each of their randomly assigned technical terms, the respondent was given an example sentence from a privacy policy and asked to define that technical term in their own words. For some technical terms, the respondent was also asked a follow-up question about that technical term (e.g., to list examples of *device attributes*).

We recruited 100 users on Amazon Mechanical Turk to take our pilot survey; 20 responses were rejected because the responses were determined to be automatically generated and unrelated to the survey questions, and we analyzed the remaining 80 responses. Three authors independently qualitatively coded each response for correctness; responses were considered incorrect if 2 of the 3 coders marked them as incorrect. Each coder also noted any significant misconceptions exhibited by qualitative responses.

Overall, 30% of responses correctly defined the given technical term. For 33 of the terms considered (including 10 terms that appeared in 500 or more privacy policies in our corpus), at least 25% of respondents were unable to correctly define the term. In addition, respondents exhibited significant misunderstandings about 13 of the technical terms; for example, eight of the sixteen respondents who were asked to define *pixel tag* instead provided a definition for *pixel* or a definition for *image tag*, and seven of the twelve respondents who were asked to define *device fingerprinting* described the use of fingerprints as a biometric.

Root term	Freq.	Technical Term
Cookie	3079	Session cookie Persistent cookie
Privacy Policy	3029	Privacy policy Data use policy
Personal information	2984	Personal information
Tracking	2222	Tracking
Aggregated	1849	Aggregated information
Personally identifiable information	1565	Personally identifiable information (PII)
Anonymize	1512	Anonymized information
Encryption	1144	Encryption End-to-end encryption
Web beacons	1010	Web beacons
Do Not Track	500	Do Not Track request
Pixel tag	318	Pixel tag
De-identified	294	De-identified information
Keys	230	Keys
Local storage	121	Local storage
Metadata	57	Metadata
Fingerprinting	39	Device fingerprinting
Private browsing	12	Private browsing/ incognito mode
Browser web storage	7	Browser web storage
Cryptographic hash	4	Cryptographic hash

Table 1. The 22 terms selected, with the frequency of each term. The root term is the term as appears in our privacy policy corpus, and the technical term is the term as appears on our studies.



(a) Age **(b)** Race **(c)** Gender

Fig. 1. The demographics of participants in our main user study, compared to the the demographics of the United States as published in the American Community Survey (ACS).

3.2 Main Study

Drawing on the results of our pilot study, we developed a preliminary draft of our main study that investigated how users understand 22 key technical terms (Table 1); this dictionary of terms is comprised of the 10 high-frequency terms[1] that were commonly misdefined in our pilot study together with the 10 additional terms for which pilot study participants exhibited significant misunderstandings. We also included free-response questions for two high-frequency terms that were correctly defined by most users: *personal information* and *personally-identifiable information*.

The first section of the survey asked users to define various technical terms through a sequence of multiple choice questions. Each question presented the term in an example sentence drawn from one of the privacy policies we manually examined. For most terms, the user was asked to select the best response; however, for a few terms the user was asked to select all responses that apply. In all cases, the incorrect options were drawn from incorrect responses submitted for our pilot study. Two additional terms—*personal information* and *personally identifiable information(PII)*—were included as open-answer questions. The order of these questions and the order of the responses were randomized. Two terms were excluded from this section: *session cookie* was excluded in favor of *persistent cookie* and *aggregated information* was excluded for lack of a standardized definition.

In the second section of the survey, the user was asked to rate how comfortable they would be with various possible policies on a five-point Likert scale. Half of the policies used technical terms while the remaining policies used non-technical, explanatory language. The order of the questions and the direction of the Likert scale was randomized. Ten terms were omitted from this section either because they did not relate to the handling of personal data or because there were no significant misconceptions in our pilot study.

Due to the number of questions in the full survey, each respondent was assigned to answer only half of the questions (divided by the set of technical terms they were asked about). Each survey then concluded with questions about the respondent's demographics, technical background, and prior interactions with privacy polices.

We conducted six cognitive interviews (three for each half of the survey). In these interviews, we asked participants to take the survey while sharing their thoughts out loud. Based on their thoughts and reactions to the survey, we eliminated unnecessary questions and modified the wording to eliminate ambiguities. Complete copies of the final set of questions included in the main study are included in Appendix A.2.

After finalizing our survey, we recruited 1159 users on Amazon Mechanical Turk. 359 responses were rejected for entering irrelevant or incoherent responses to the open-ended definition questions; we analyzed the remaining 800 responses. 5.5% of our population were 18-24, 44.4% were 25-34, 27.9% were 35-44, 17.1% were 45-59, 5% were 60-74, and .1% were 75 or older. Racially, 78% of our respondents were white, 12.6% were Black or African American, 2.1% were Native American, 9.3% were Asian, and 1.5% identified as other.[2] 60.8% of our

1 A term was deemed to be high-frequency if the corresponding root term appeared in 1000 or more policies in our corpus.

2 The numbers do not sum to 100% because respondents were allowed to select more than one race.

users identified as male, 38.8% identified as female, and .5% identified as non-binary. A graphical summary of these demographics—along with a comparison to the demographics of the United States as published in the American Community Survey—is depicted in Figure 1.

We observe briefly that the demographics of these user studies are not an exact match to the overall demographics of the United States. Notably, our sample population is younger than the overall U.S. population, a known property of the Amazon Mechanical Turk worker pool [6, 15, 20, 26]. Our sample population is also more male that the overall population, despite the fact that the U.S. Mechanical Turk worker population skews female [6, 26, 27]; this mismatch may indicate selection bias that could limit the generalizability of our results. Nonetheless, prior work has found that Mechanical Turk results for surveys on security and privacy topics generalize well despite the demographic discrepancy, particularly for populations under 50 with at least some college education [32].

3.3 Follow-up Study

As comfort is a somewhat nebulous term, we also conducted a small follow-up survey ($n = 108$) to determine whether users have different attitudes towards accepting policies with technical versus explanatory terms. For this study, we selected the 10 high-frequency technical terms from our main study that describe data use practices (this excluded *privacy policy* and *data use policy*).

In the first section of our follow-up study, we asked users how likely they were to accept different policies on a five-point Likert scale. As in the main study, half the policies used technical language while the other half used explanatory language for the same terms. The second section asked users to choose the best definition or best response for each of the terms encountered in the first section of the survey. The definitions and incorrect responses were drawn from responses given in our pilot study. An "I don't know" option and "Other" option (with a textbox) were also provided for each definition question; however, no respondents selected the "Other" option for any of the questions. Complete copies of the final set of questions included in the main study are included in Appendix A.3.

3.4 Analysis Plan

In an effort to limit the number of Type I errors, we committed to an analysis plan prior to analyzing the responses from our main study.

For the questions that evaluated how well users understand technical terms, we simply report user accuracy. For single-response questions, only one of the possible choices was correct, so responses were identified as correct or incorrect depending on whether or not the respondent selected the single correct response. For multi-response questions, each response was coded as demonstrating or not demonstrating the hypothesized misconception; responses that demonstrated the misconception were coded as incorrect and all other responses were coded as correct.[3]

To analyze the Likert-scale questions relating to users' level of comfort with various privacy policies, we first converted the Likert scales to numerical scales, with 1 being "Very uncomfortable" and 5 being "Very comfortable". Because definitions of comfort are likely to be inconsistent between different users, we did not analyze absolute reported comfort levels; instead, we ran paired t-tests to detect within-subject changes in comfort levels between technical terms and explained terms (i.e. comparing a user's comfort for the technical term with the same user's comfort for the explained term). A similar conversion was done with the Likert scales of likelihood to accept a policy in the follow-up study, with 1 being "Very unlikely" and 5 being "Very likely". We also ran paired t-tests for the follow-up study regarding changes in a user's likelihood of accepting a policy.

Because users with technical background are known to be over-represented in the Amazon Mechanical Turk worker population, we also tested whether our result depended on this background variable. We used unpooled two-sample t-tests to detect whether or not the mean level of comfort or number of correct answers were significantly different between groups.[4]

Once all the tests were complete, we used the Benjamini-Hochberg correction for multiple comparison to limit the false discovery rate. Results that are statistically significant using the standard $p < .05$ threshold are denoted by *; results that are statistically significant under this conservative correction are denoted by **.

3 The precise correctness coding is provided in Appendix A.2 and Appendix A.3 above each multiple response question.

4 We chose to use t-tests as they are robust to the normality assumption, particularly with large sample sizes, which were present in this study.

4 Results

We analyzed the the data from our main study in two independent ways. First we analyzed the questions that asked users to define various technical terms in order to determine how accurately users understand those terms and whether there are any common misconceptions. We also tested whether technical background is associated with statistically significant differences in how accurately users define technical terms. These results are described in Section 4.1. Overall, we found that users misinterpret many technical terms used in privacy policies and that particular misconceptions are common.

Second, we evaluated how the use of technical terms affects users' reported comfort level with various possible privacy policies. We also tested whether there were significant differences in comfort level between users who correctly defined the term and those who did not or between users with a technical background and those without. These results are described in Section 4.2. Overall, we found that the use of technical terms significantly affected users' comfort levels.

We also analyzed the results of our follow-up study for correctness and to determine whether there were significant differences in likelihood of accepting a policy with a technical term and one with equivalent explanatory language. These results are described in Section 4.1.3 and Section 4.3.

4.1 Correctness Results

When we analyzed the questions that asked respondents to define various technical terms, we found that these technical terms are commonly misunderstood by American Internet users. Across our main study, respondents answered on average 39.57% of questions correctly, 49.01% of questions incorrectly, and answered "I don't know" for 11.41% of questions. Overall, for 15 of the 20 technical terms in our main study, less than half of survey respondents were able to correctly define the term. Accuracy for individual technical terms is discussed in Section 4.1.1, the effect of technical background on definition accuracy is discussed in Section 4.1.2, and the correctness results from our follow-up study are discussed in Section 4.1.3.

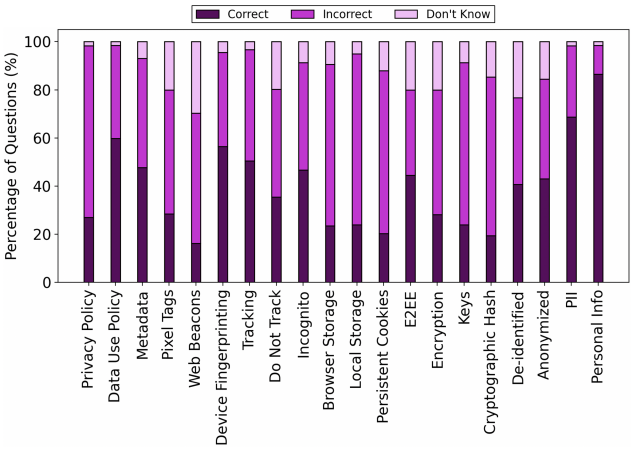
4.1.1 Accuracy for Individual Technical Terms.

We individually analyzed how accurately users were able to define each of the technical terms that appeared in our main study. These results are summarized in Figure 2a.

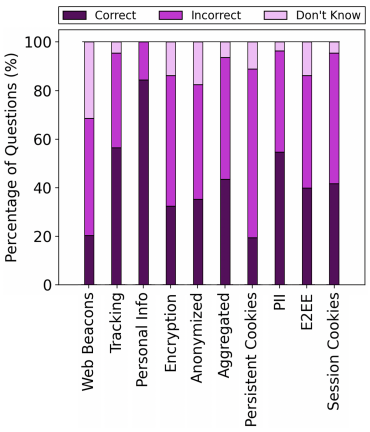
Privacy Policy and Data Use Policy. Privacy policies (sometimes called data use policies) are legal documents that describe how a company collects and uses data about users. Prior work has found that users misunderstand the term *privacy policy*; a majority of users believe that having a privacy policy means that the website will not share information with other websites or companies [38, 41, 42]. Our user study replicated those results: when asked to define the term *privacy policy*, 71% of respondents incorrectly selected answers that describe privacy policies as guaranteeing data protection, confidentiality, or consent. However, when we asked users to define the alternative term *data use policy*, only 39% of users selected incorrect answers. These results confirm that the term *privacy policy* is commonly misunderstood, and that using this term can mislead users about a company's data use practices. To improve privacy, companies should adopt alternative language such as *data use policy*.

Metadata. Metadata is data about data or files, often automatically generated when that data is created or used, but preliminary results from our pilot study suggested that some users associate the term *metadata* with collections of data generally. In our main study, we found that a plurality of users (48%) were able to correctly define *metadata*, and that no single misconception was widely held. On the other hand, many user did not recognize how many different types of data could be considered metadata: only 20% of respondents recognized that size of file, time of message, message receiver, location of a photo, and file uploader are all types of metadata.

Pixel Tags and Web Beacons. Although tracking-enabling technologies such as pixel tags and web beacons are commonly used, preliminary evidence suggested that many users are unsure what these terms mean. Our main study confirmed this result; many respondents simply stated that they did not know what these terms meant (29% for *web beacons* and 20% for *pixel tags*), and few users were able to correctly define these terms (16% and 29% respectively).



(a) Main Study



(b) Follow-up Study

Fig. 2. User accuracy at defining technical terms that appear in privacy policies.

Device Fingerprinting. Device fingerprinting is a method of generating a unique identifier based on browser or device features in order to track a user. After analyzing the qualitative responses from our pilot study, we hypothesized that many users mistakenly believe that *device fingerprinting* refers to the use of fingerprint-based biometrics to unlock a device. However, although our main study did find that 28% of users hold this misconception, the majority of respondents were able to correctly define *device fingerprinting* as "a method of tracking that involves using browser features to uniquely identify a specific device."

Tracking. Tracking refers to the practice of monitoring user behavior across sites they visit. Based on our pilot results, we hypothesized that users would mistakenly associate the term *tracking* with location tracking. However, we were surprised to find that 51% of respondents correctly identified *tracking* as "a way to monitor and record which sites you visit and/or your behavior on those sites". Only 12% of respondents thought *tracking* was "a way to monitor and record your location".

Do Not Track. Do Not Track [7] is a browser signal that indicates to websites that the user does not want to be tracked; however, despite the name there is no legal requirement that websites respect this request. We hypothesized that users would misunderstand *Do Not Track* as a requirement rather than as a request; this hypothesis was confirmed by our main user study in which only 36% of users correctly identified that they can still be tracked by websites after sending a Do Not Track request. We observe, however, that the CCPA now requires websites who do not respect Do Not Track requests to explicitly say so in their privacy policy, so this misconception may be ameliorated in the future.

Private Browsing/Incognito Mode. Private Browsing or incognito mode are privacy features of browsers that do not store local data (e.g., browsing history or cookies) after the end of a session. We initially hypothesized that respondents would be unaware that data from their private browsing sessions can be stored by the websites visited or third parties. We were thus surprised to find that 46.75% of respondents were aware that information about their online behavior while in private browsing mode could be stored by at least one remote principal. However, there were still many users (44.5%) who mistakenly believed that private browsing data is either stored nowhere or just on their device.

Browser Storage and Local Storage. Local storage (sometimes called browser web storage) refers to the storage of files on the user's device, for example in the browser cache. Pilot survey responses suggested that many users interpret *local storage* as a geographic restriction (i.e., that data must be stored in the user's local geographic area or on company premises). Our main study confirmed this hypothesis: 67% of users selected either the option "on machines or servers that are onsite to the company" or "on machines or servers in your local area," indicating that they have a physical definition of the word local. However, we found that *browser web storage* is also not well understood; only 24% of respondents selected the correct definition.

Persistent Cookies. Persistent cookies are cookies that are stored until they expire (or until the user deletes them). Pilot survey responses indicated that users mistakenly believe that persistent cookies never expire. This preliminary result was confirmed by our main study, where only 20.25% of respondents correctly indicated that persistent cookies have an expiration date and where 7.25% of respondents selected the exclusive answer "Never" when asked when persistent cookies expire. However, there is a sizable portion of our survey population (25.5%) that selected an answer indicating that persistent cookies have a short life span (either that they are deleted when you close your browser, when you sign out of the website, when you close the tab, or when you navigate away from the website).

End-to-End Encryption and Encryption. Unlike standard systems that use encryption, end-to-end encryption is a method of securing communications in which only the endpoints (and not any centralized servers) have access to the decryption keys. Based on anecdotal evidence, we hypothesized that users would believe that point-to-point encryption (as opposed to end-to-end encryption) is insecure. However, we also hypothesized that users do not in fact understand what protections are offered only by end-to-end encryption. Our results validated our first hypothesis. 50.25% of respondents indicated that they believed *end-to-end encryption* to be secure while only 17% believed other sorts of *encryption* to be secure. Contrary to what we had initially predicted, 70.25% of respondents were able to correctly identify a protection offered by end-to-end encryption and 44.5% of respondents were able to correctly identify both protections associated with end-to-end encryption. It is evident that users have a better understanding of end-to-end encryption protections than we had initially predicted.

Keys. Cryptographic keys are randomly generated bit-strings used to encrypt and decrypt messages and files. However, the initial responses from our pilot study suggested that users conflate cryptographic keys with passwords. Our main study confirmed that the misconception that keys are the same as passwords is relatively commonly held (18%) compared with the proportion of users who correctly define the term *keys* (24%). We also found that users confuse keys with authentication tokens: 39% of users selected the option "information that allows you to access your account without logging in every time" when asked to define *keys*. Together, these results suggest a widespread misconception that keys are primarily involved in authenticating users rather than with encrypting data.

Cryptographic Hash. Cryptographic hashes are message digests produced by deterministic hash functions, often used to store validation information for passwords or other values. Pilot survey responses led us to hypothesize that users conflate cryptographic hashes with encryption. This hypothesis was validated by the results of our full survey: only 20% of respondents selected the correct definition. When asked to select the most accurate meaning of the policy, "we store a cryptographic hash of your password", the plurality of users (32%) incorrectly selected, "we store an encrypted copy of your password."

De-identified, Anonymized, and Aggregated. De-identification, anonymization, and aggregation are different ways of processing data. We asked users whether *de-identified*, *anonymized*, and *aggregated* information can ever be tied to an individual user. We initially hypothesized that respondents would interpret *de-identified* and *anonymized* as synonymous. This hypothesis was generally supported by our findings: 64.25% of respondents provided the same answer for both questions. Moreover, significant minorities (40% and 43%) were aware that de-identified information (resp. anonymized information) can be tied to individuals under some circumstances. By comparison, 56.75% of respondents believed that aggregated data can sometimes be tied to an individual, despite the lack of standardized definition for this term.

PII and Personal Information. *Personally-identifiable information* (PII) (resp. *personal information*) are legal categories of data defined by the GDPR (resp. CCPA). *Personal information* is defined more expansively than *PII*, but our pilot study results show that users do not necessarily differentiate between what types of data collected count as PII and personal information. We hypothesized that the types of examples users would give of PII and of personal information would be similar. We coded users that correctly gave two or more correct examples of either PII or personal information as correct. As more types of data are considered personal information than PII, 69% of users were correct about PII and 86.5% about personal information; we did not observe qualitative differences between the examples given for PII and the examples given for personal information.

4.1.2 Effect of Technical Background

Because users with technical background are known to be over-represented on Amazon Mechanical Turk, we tested whether there were differences in accuracy between users with technical background and users without. We found that users with technical backgrounds defined 41.97% of questions correctly on average whereas users without technical backgrounds defined 38.49% of the technical terms correctly; this difference was statistically significant ($p = 0.034$). We also found that users with a technical background were more confident about defining technical terms; on average, respondents with a technical background selected "I don't know" for just 4.00% of the questions, compared to 14.77% of respondents without a technical background ($p < .001$). These results suggest that while users with technical background are likely more familiar with and more comfortable with technical terms, this general technical background does not necessarily imply completely accurate knowledge of the specific technical terms that appear in privacy policies.

4.1.3 Correctness for Follow-Up Survey

In our follow-up study, users were asked to rate how likely they would be to accept a policy (rather than how comfortable they were with a policy), and the questions about defining technical terms were asked after the questions about accepting various privacy policies (in contrast with the main study, which asked the definition questions before asking respondents to rate the example policies). Overall, the results of the follow-up study—depicted in Figure 2b—were consistent with the results of the main study.

Eight of the ten definition questions asked in the follow-up study were identical to questions asked in the main study. Users in the follow-up study answered 42.82% of these questions correctly on average compared to 39.57% for users in the main study. Respondents in the the follow-up study scored slightly lower on 5 of the shared questions and slightly higher on 3 of them. For five of the eight shared questions, the percentage of users who correctly defined the term differed by at most 5% and for the other three questions that proportion differed by less than 15%.

These results suggest that user accuracy at defining terms is robust to the ordering of questions in our study.

4.1.4 Limitations

These results about accuracy clearly depend on the precise selection of terms included in this study; the fact that respondents on average were able to correctly answer fewer than 40% of questions accurately does not necessarily imply users are unfamiliar with all technical terms, merely that the terms included in this study (many of which also appear frequently in real-world privacy policies) are broadly misunderstood by many users.

Because incorrect options to these multiple choice questions were drawn from incorrect real user responses, some of the questions may contain ambiguities that would fail to distinguish between users with partial knowledge of a term and users with no knowledge of the term. It is therefore possible that users with some technical background do have more knowledge about the technical terms included in this study, but that our analysis simply failed to detect this partial expertise.

4.2 Comfort Levels

In addition to evaluating how accurately users can define technical terms, we also evaluated the effect of technical terms on user comfort by asking users to rate their comfort level with a variety of different privacy policies; some of these privacy policies included technical terms while others were equivalent policies using non-technical, explanatory language. For this section, we focused on 12 technical terms that commonly appear in privacy policies; the other 10 terms were omitted from this analysis either because they do not relate to the handling of personal data or because there were no significant misconceptions in our pilot study.

4.2.1 Overall Effect of Technical Terms on Comfort

For most of our technical terms, we found significantly different levels of comfort between the policies that used technical terms and the equivalent policies that used non-technical language. In all cases, the direction of the effect was consistent with our hypothesis about what misconception(s) users might hold about that technical term. A summary of these results is given in Table 2; the corresponding boxplots are shown in Figure 3 and

Term	H	Comfort (Tech.)	Comfort (Expl.)	p-value
Metadata	↓	**2.71**	**2.36**	$< 0.001^{**}$
Pixel Tags	↓	**2.74**	**2.31**	$< 0.001^{**}$
Web Beacons	↓	**2.61**	**2.31**	$< 0.001^{**}$
Device Fingerprinting	↓	**2.77**	**2.70**	**0.243**
Browser Storage	↑	**2.75**	**3.05**	$< 0.001^{**}$
Local Storage	↑	**2.93**	**3.05**	0.024^{*}
Persistent Cookies	↑	**2.57**	**2.78**	$< 0.001^{**}$
Session Cookies	↑	**2.97**	**2.96**	**0.875**
E2EE	↑	**3.90**	**3.84**	**0.263**
Encryption	↑	**2.84**	**3.45**	$< 0.001^{**}$
PII	–	**2.34**	**2.34**	**0.958**
Personal Info	–	**2.46**	**2.34**	0.016^{**}

Table 2. Mean comfort for technical terms compared with the explained versions of those terms. The arrows indicate the hypothesized direction of change in comfort between policies with the technical term and policies with the explanatory language. p-values show the results from paired t-tests (H_0 : No difference in comfort for users encountering a policy with a technical term or explanatory term). * denotes p-values under the 0.05 cutoff but above the threshold from multiple comparison corrections. ** denotes p-vales under the cutoff after corrections have been applied.

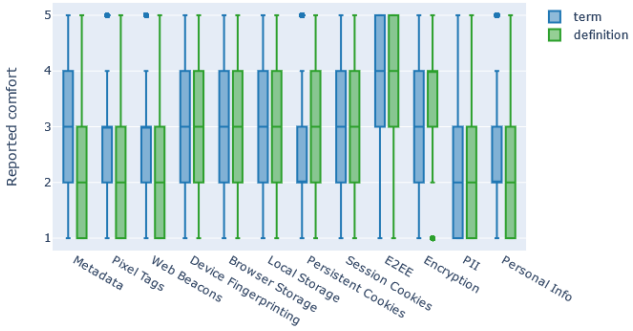


Fig. 3. Boxplots of reported comfort levels for technical terms against comfort levels for the same term but explained/defined

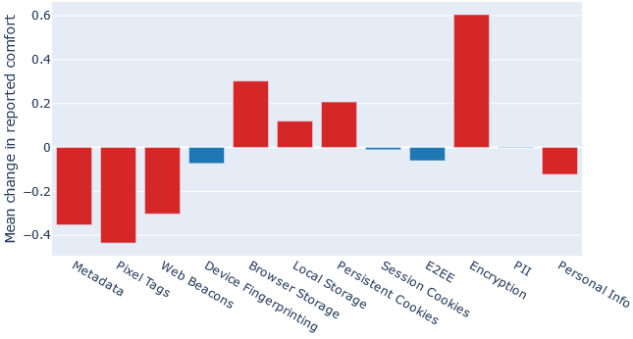


Fig. 4. Difference between the mean reported comfort for the explanatory policy and the equivalent policy that used the technical term. Negative values indicate lower comfort with the explanatory policy than with the technical policy. Red bars indicate changes in comfort that are statistically significant ($p < 0.05$)

the mean difference in reported comfort levels are summarized in Figure 4.[5]

For certain technical terms, the common user misconceptions failed to appreciate the full scope of data use practices. For these terms, we found that users were significantly less comfortable with a non-technical, explanatory policy than with an equivalent policy that uses the technical term. For example, users reported significantly lower levels of comfort with the equivalent explanatory policy than with the policy that used the term *metadata* ($p < .001$). This result is consistent with our observation that while many users are able to correctly define *metadata*, most users (80%) do not realize how many different types of data are considered metadata. Similarly, for the technical terms related to tracking that were unknown to most users—*pixel tags* and *web beacons*—reported comfort decreased significantly when given the non-technical, explanatory policy compared the policies that used these technical terms ($p < .001$).

Surprisingly, there was no significant difference in the reported comfort level for the term *device fingerprinting* and the equivalent explanatory policy even though 28% of respondents held the misconception that device fingerprinting referred to the use of fingerprint-based biometrics to unlock a device. We suspect this is due to the fact that a majority of respondents were able to correctly define this term; we explore this hypothesis further in Section 4.2.2.

In other cases, commonly held misconceptions caused users to fail to appreciate security and privacy that was offered by a company because those properties were obscured behind commonly misunderstood technical terms. For example, for the two technical terms that dealt with storage—*browser storage* and *local storage*—we observed that users were significantly more comfortable with the equivalent explanatory policy than with the policies that contained the technical terms. This is consistent with our observation in Section 4.1 that many users mistakenly believe that *browser storage* or *local storage* allow their information to be stored on remote machines under the control of other principals. We also found that users were significantly more comfortable with the policy that used the term *local storage* than

5 Note, however, that the summary of differences between means should only be interpreted as indicating the direction and significance of the effect; we make no claims about the magnitude of these changes due to the ordinal nature of our survey data.

Term	H	User Incorrect	User Correct	User Didn't Know	p-value (Corr. vs. Incorr.)	p-value (Corr. vs. DK.)
Metadata	↓	**2.87**	**2.59**	**2.50**	**0.017****	**.656**
Pixel Tags	↓	**2.78**	**2.80**	**2.58**	**0.872**	**.116**
Web Beacons	↓	**2.71**	**2.71**	**2.38**	**0.997**	**.053**
Device Fingerprinting	↓	**3.10**	**2.58**	**2.17**	**< 0.001****	**.156**
Browser Storage	↑	**3.03**	**2.84**	**2.44**	**0.145**	**.033***
Local Storage	↑	**2.94**	**2.98**	**2.65**	**0.752**	**.134**
Persistent Cookies	↑	**2.49**	**3.07**	**2.19**	**< 0.001****	**< .001****
E2EE	↑	**3.78**	**4.09**	**3.7**	**0.009****	**.005****
Encryption	↑	**2.76**	**2.98**	**2.75**	**0.079**	**.121**
PII	–	**2.58**	**2.22**	**2.87**	**0.008****	**.261**

Table 3. Analysis of mean user comfort broken down by whether the user correctly defined the relevant technical term. Arrows depict the hypothesized direction of change in comfort between users who incorrectly define the term and users who correctly define the term; in all cases, users who didn't know the answer were expected to report lower comfort levels. p-values show the results of two-sample t-tests (H_0 : No difference in comfort for users who correctly defined, incorrectly defined, or did not know a term, with policies containing the term). * denotes p-values under the 0.05 cutoff but above the threshold from multiple comparison corrections. ** denotes p-vales under the cutoff after corrections have been applied

the policy that used the term *browser storage* (mean comfort 2.93 vs. 2.75, $p < .001$), despite the fact that privacy policies appear to use these terms interchangeably.

Similarly, while most users were familiar with the term *cookie*, many users misunderstood the term *persistent cookie* to mean that the cookie would never expire. We therefore observed significantly lower levels of comfort with that technical term than with the equivalent explanatory policy ($p < .001$). By contrast, there was no significant change in comfort levels between the term *session cookies* and its equivalent explanation.

We also observed a similar effect for technical terms relating to encryption. Users reported significantly lower comfort levels for the policy containing the technical term *encryption* than for the equivalent explanatory policy ($p < .001$), likely because recent news coverage about end-to-end encryption has given rise to the mistaken belief that non-end-to-end encryption is not secure. We also found that users reported significantly lower comfort levels for the policy containing the technical term *end-to-end encryption* than for the equivalent explanatory policy, perhaps due to the fact than many users were not aware of the key features of end-to-end encryption (that third parties cannot decrypt the contents of messages). Users were significantly more comfortable with end-to-end encryption than with other sorts of encryption (mean comfort 3.90 vs. 2.84, $p < .001$), despite being generally unfamiliar with the precise definitions or the security properties enforced by these cryptographic techniques.

4.2.2 Effect of Term Comprehension

We also evaluated whether there were significant differences in comfort level between users who correctly defined a term, users who incorrectly defined it, and users who did not know.

For each term, we divided the answers into "Correct", "Incorrect", and "Don't know".[6] We then performed unpooled two-sample t-tests to determine whether there were statistically significant differences in comfort level between these populations. These results are summarized in Table 3.

For five of the technical terms—*metadata*, *device fingerprinting*, *persistent cookies*, *end-to-end encryption*, and *personally-identifiable information (PII)*—we detected statistically significant differences between respondents who correctly defined the term and respondents who selected an incorrect definition. For *metadata* and *fingerprinting*, users who correctly defined the term were less comfortable with a policy that contained it than users who incorrectly defined the term; this was consistent with the results of Section 4.2.1, which found that overall, users were less comfortable with the explanatory language than with the technical term. In both cases, this was consistent with our hypothesis that users do not understand the full privacy implications of these terms. For *persistent cookies*, users who cor-

6 As in Section 4.1, response to multi-answer questions were coded as correct or incorrect depending on whether they exhibited a particular misconception.

Term	H	Accept (Tech.)	Accept (Expl.)	p-value
Web Beacons	↓	2.66	2.43	0.069
Tracking	↓	2.82	2.64	0.105
Personal Info	↓	2.74	2.5	0.030*
Encryption	↑	2.94	3.66	< 0.001**
Anonymized Info	↑	3.21	3.93	< 0.001**
Aggregated Info	↑	3.17	3.04	0.158
Persistent Cookies	↑	2.91	2.81	0.416
PII	–	2.65	2.5	0.142
E2EE	–	4.08	4.04	0.747
Session Cookies	–	3.17	3.17	1.0

Table 4. Mean likelihood to accept a policy with technical terms compared with likelihood to accept a policy with equivalent explanatory language. Arrows denote the hypothesized change in acceptance rates between the technical term and the explanatory language (based on change in comfort observed in the main study). * denotes p-values under the 0.05 cutoff but above the threshold from multiple comparison corrections. ** denotes p-vales under the cutoff after corrections have been applied

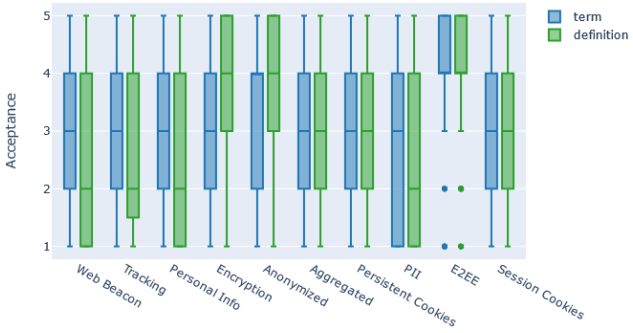


Fig. 5. Boxplots of reported acceptance likelihood of policies containing technical terms against comfort levels for the same term but explained/defined

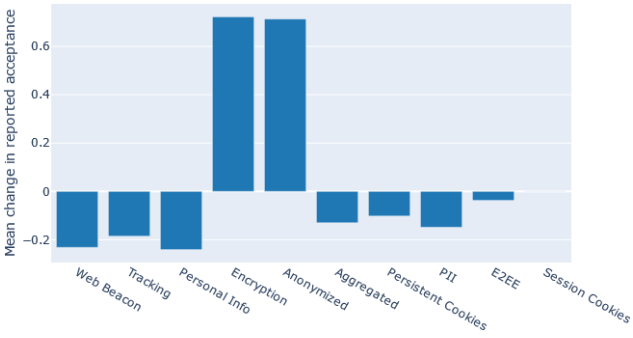


Fig. 6. Differences between the mean reported likelihood of acceptance between the explanatory policies and the equivalent policies that used the technical terms. Negative values indicate lower likelihood of acceptance for the explanatory policy than for the technical policy.

rectly defined the term were more comfortable than users who incorrectly defined the term; this is again consistent with the results of Section 4.2.1 and with our hypothesis (that users don't realize that persistent cookies can be deleted or expire). For *device fingerprinting* and *personally-identifiable information (PII)*, there were significant differences between users who correctly defined the term and users who incorrectly defined the term despite the fact that we did not observe differences in reported comfort level between the policies that contained these technical terms and equivalent explanatory policies. Surprisingly, in all other cases there were no significant differences in comfort between users who correctly defined the term and users who defined it incorrectly.

Although we also expected a significant difference in comfort levels between users who correctly defined a term and those who did not know what a term meant, we only detected a significant effect for two terms, *persistent cookies* and *end-to-end encryption*, after corrections for multiple comparisons. However, in most cases the number of "I Don't Know" responses was relatively small, likely precluding statistically significant results.

4.2.3 Effect of Technical Background

Due to the technical nature of the terms used in this study, we hypothesized that users with a technical background would be more comfortable with policies that used technical terms than users without a technical background (H_A). Our results from a two-sample t-test show that there is a difference between these groups, and the difference in means validates this alternative hypothesis. (H_0 : No difference in comfort with a policy for users with and users without a technical background.) Moreover, users with a technical background are more than twice as likely to rate their level of comfort as "Very comfortable" than users without a technical background ($p < .001$). This higher level of comfort might be due to those with technical backgrounds feeling more comfortable with the terms due to higher familiarity or to a (possibly incorrect) belief that they understand the meaning and purpose of those terms due to their background.

4.3 Acceptance Rates

To determine whether the observed changes in comfort might translate into changes in behavior, we ran a follow-up study in which we asked whether users would

accept policies with the same language as our main study; to ensure that no priming effects occurred due to the definitions questions, these questions were placed before the definition questions in the survey order and users were not allowed to go back.

We compared acceptance likelihoods of policies with technical language against those with equivalent explanatory language using paired t-tests. For terms that appear in both the main study and the follow-up study, we hypothesized that we would observe the same effects on acceptance likelihood as we had seen on comfort (H_0 : No difference in acceptance likelihood for users encountering a policy with a technical term and one with an explanatory term). The complete results of this analysis, along with the hypothesized effect for each term, are shown in Table 4. Boxplots summarizing these findings are shown in Figure 5 and the mean differences in acceptance likelihood are summarized in Figure 6.

We found that for all terms that appeared in both our main and follow-up study, the acceptance likelihood on a 5-point Likert scale was slightly higher than the reported comfort, a pattern that likely reflects the fact that users may accept policies even if they are uncomfortable with the data use practices described. The direction of the observed changes in acceptance likelihood were generally consistent with the changes in comfort observed in the main study, although the p-values were higher (likely due to the smaller sample size of the follow-up study). Of the terms that appeared in both studies, the only inconsistent trend was for *persistent cookies*, and the difference between acceptance likelihood for the policy with that technical term and the equivalent policy with explanatory language was not statistically significant ($p = .416$). These results affirm the findings of our main study: that the use of technical terms in privacy policies is likely to affect user comfort (and likelihood of accepting a policy) in ways that negatively impact user privacy.

5 Conclusion

In this work, we identified technical terms used in privacy policies, and we ran a set of user studies on Amazon Mechanical Turk to evaluate how well users understand these privacy policies, to identify common misconceptions, and to evaluate how these misunderstandings and misconceptions affect users comfort with policies that contain technical terms. We found that technical terms are commonly misunderstood, and that particular misconceptions are widespread. We also found that how comfortable users are with various policies varies depending on whether the language of the policy includes technical terms, and that for most terms this discrepancy holds whether or not the user is able to correctly define that particular technical term. We validated these results with a follow-up study that found the same trends for how likely users are to accept a privacy policy.

Our results suggest that the current use of technical terms in privacy policies is a barrier to informed consent. While some use of technical terms is likely necessary to maintain legal enforceability, the use of such terms needs to be carefully balanced with the need for data use transparency and informed consent. Possible approaches might include replacing frequently misunderstood terms with alternate language, the addition of explanatory definitions to privacy policies—either in dedicated sections or in combination with interactive elements such as hover—and/or augmentation of privacy policies with summaries, annotations, visualization tools, or other technologies. Further work will be required to evaluate the effect of these various solutions on user comprehension and to determine the best path toward creating clear, accessible policies for the full community of users.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Annie Anton, Julia Earp, Qingfeng He, William Stufflebeam, Davide Bolchini, and Carlos Jensen. Financial privacy policies and the need for standardization. *IEEE Security & Privacy*, 2:36–45, 03 2004.

[2] Manon Arcand, Jacques Nantel, Mathieu Arles-Dufour, and Anne Vincent. The impact of reading a web site's privacy statement on perceived control over privacy and perceived trust. *Online Information Review*, 2007.

[3] Peter Breese and William Burman. Readability of notice of privacy forms used by major health care institutions. *JAMA : the journal of the American Medical Association*, 293:1593–4, 05 2005.

[4] Rochelle Cadogan. An imbalance of power: The readability of internet privacy policies. *Journal of Business & Economics Research (JBER)*, 2, 02 2011.

[5] Gitanjali Das, Cynthia Cheung, Camille Nebeker, Matthew Bietz, and Cinnamon Bloss. Privacy policies for apps targeted toward youth: Descriptive analysis of readability. *JMIR Mhealth Uhealth*, 6(1), Jan 2018.

[6] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. Demographics and dynamics of mechanical turk workers. WSDM '18, page 135–143, 2018.

[7] W3c do not track standard. http://www.w3.org/TR/2015/WD-tracking-compliance-20150714/.

[8] Tatiana Ermakova, Benjamin Fabian, and Eleonora Babina. Readability of privacy policies of healthcare websites. 03 2015.

[9] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. Large-scale readability analysis of privacy policies. In *Proceedings of the International Conference on Web Intelligence*, page 18–25, New York, NY, USA, 2017. Association for Computing Machinery.

[10] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, New York, NY, USA, 2012. Association for Computing Machinery.

[11] Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[12] General data protection regulation (GDPR), 2016.

[13] Mark Graber, Donna D'Alessandro, and Jill Johnson-West. Reading level of privacy policies on internet health web sites. *The Journal of family practice*, 51:642–5, 08 2002.

[14] Mark Hochhauser. Lost in the fine print: Readability of financial privacy notices. 06 2001.

[15] Panagiotis G Ipeirotis. Demographics of mechanical turk. 2010.

[16] Musa Jafar and Amjad Abdullat. Exploratory analysis of the readability of information privacy statement of the primary social networks. *Journal of Business & Economics Research (JBER)*, 7, 02 2011.

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

[18] Barbara Krumay and Jennifer Klar. Readability of privacy policies. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 388–399. Springer, 2020.

[19] Priya Kumar. Privacy policies and their lack of clear disclosure regarding the life cycle of user information. In *AAAI Fall Symposia*, 2016.

[20] Kevin E. Levay, Jeremy Freese, and James N. Druckman. The demographic and political composition of mechanical turk samples. *SAGE Open*, 6(1):2158244016636433, 2016.

[21] Stephen D. Lewis, Robert G. Colvard, and C. N. Adams. A comparison of the readability of privacy statements of banks, credit counseling companies, and check cashing companies. *Journal of Organizational Culture, Communications and Conflict*, 12(2):87–93, 2008.

[22] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. The privacy policy landscape after the gdpr. *Proceedings on Privacy Enhancing Technologies*, 2020(1):47–64, 2020.

[23] Aleecia M Mcdonald, Robert W Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. A comparative study of online privacy policies and formats. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 37–55. Springer, 2009.

[24] Gabriele Meiselwitz. Readability assessment of policies and procedures of social networking sites. In *Online Communities and Social Computing*, pages 67–75. Springer Berlin Heidelberg, 2013.

[25] George Milne, Mary Culnan, and Henry Greene. A longitudinal assessment of online privacy notice readability. *Journal of Public Policy & Marketing - J PUBLIC POLICY MARKETING*, 25:238–249, 09 2006.

[26] Aaron Moss and Leib Litman. Demographics of people on amazon mechanical turk. 06 2020.

[27] Aaron J. Moss, Cheskie Rosenzweig, Jonathan Robinson, and Leib Litman. Demographic stability on mechanical turk despite covid-19. *Trends in Cognitive Science*, 24(9), 06 2020.

[28] California Office of the Attorney General. California consumer privacy act regulations: Final text of regulations. https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf.

[29] Irene Pollach. A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. *Journal of Business Ethics*, 62:221–235, 12 2005.

[30] Irene Pollach. What's wrong with online privacy policies? *Commun. ACM*, 50:103–108, 09 2007.

[31] Robert Proctor, Athar Ali, and Kim-Phuong Vu. Examining usability of web privacy policies. *Int. J. Hum. Comput. Interaction*, 24:307–328, 03 2008.

[32] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1326–1343. IEEE, 2019.

[33] Joel Reidenberg, Travis Breaux, Lorrie Cranor, and Brian French. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30, 08 2015.

[34] Julie Robillard, Tanya L. Feng, Arlo B. Sporn, Jen-Ai Lai, Cody Lo, Monica Ta, and Roland Nadler. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. *Internet Interventions*, 17, 2019.

[35] Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, and Helen Nissenbaum. Analyzing privacy policies using contextual integrity annotations. 2018.

[36] Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, and Helen Nissenbaum. Going against the (appropriate) flow: A contextual integrity approach to privacy policy analysis. In *AAAI 2019*, 2019.

[37] Ravi Inder Singh, Manasa Sumeeth, and James Miller. A user-centric evaluation of the readability of privacy policies in popular web sites. *Information Systems Frontiers*, 13(4):501–514, 2011.

[38] Aaron Smith. What internet users know about technology and the web. 11 2014.

[39] H Jeff Smith, Tamara Dinev, and Heng Xu. Information privacy research: an interdisciplinary review. *MIS quarterly*,

pages 989–1015, 2011.

[40] Daniel Solove. Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7):1880–1903, 2013.

[41] Joseph Turow, Lauren Feldman, and Kimberly Meltzer. Open to exploitation: America's shoppers online and offline. *Departmental Papers (ASC)*, page 35, 2005.

[42] Joseph Turow, Michael Hennessy, and Nora Draper. Persistent misperceptions: Americans' misplaced confidence in privacy policies, 2003–2015. *Journal of Broadcasting & Electronic Media*, 62:461–478, 07 2018.

[43] Kim-Phuong L. Vu, Vanessa Chambers, Fredrick P. Garcia, Beth Creekmur, John Sulaitis, Deborah Nelson, Russell Pierce, and Robert W. Proctor. How users read and comprehend privacy policies. In *Human Interface and the Management of Information. Interacting in Information Environments*, pages 802–811. Springer Berlin Heidelberg, 2007.

A Survey Questions

A.1 Pilot Study

Due to space constraints, the full text of the questions used in our pilot study has been eliminated from the submitted version of this paper. However, the list of technical terms included in the study is provided below. For each term, respondents were given an example sentence drawn from a privacy policy—e.g., "We use various technologies to collect and store information, such as **local storage**"—and asked to define the technical term in their own words. For some terms, there was also an additional follow-up question.

1. Browser web storage
2. Local storage
3. Correct
4. Deactivate
5. Aggregated
6. De-identified
7. Infer
8. Anonymize
9. Metadata
10. Server logs
11. Usage data
12. Log data
13. Encryption
14. End-to-end encryption
15. Cryptographically hashed
16. Authentication tokens
17. Keys
18. Push tokens
19. Third parties
20. Third-party payment processor
21. Third party network
22. Data transfer
23. Affiliates
24. Tracking
25. Pixel tag
26. Web beacons
27. Private browsing
28. Do Not Track
29. Personally identifiable information
30. Personal information
31. Sensitive personal information
32. Public information
33. Location-related information
34. Ad identifiers
35. Payment information
36. Targeted ads
37. Online behavioral advertising
38. Ad tag
39. Cookies
40. Session cookies
41. Permanent cookies
42. Fingerprinting
43. Account information
44. Unique identifiers
45. IP address
46. Advertising identifier
47. Device identifier
48. Common account identifier
49. Media Access Control (MAC) address
50. Referrer URL
51. Device attributes
52. Device operations
53. Device signals
54. API/SDK
55. ISP
56. Privacy policy
57. Core service

A.2 Main Study

For questions with correct answers, the correct response will be indicated using bold font. Unless otherwise indicated, respondents selected one response per question. For multiple response questions, correct answers are indicated in bold; the definition used to code responses as correct or incorrect is specified above each multiple response question.

The survey questions are as follows:

1. Which of these best describes the main purpose of a **privacy policy**? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   - It explains how your data will be protected
   - **It is a legal document that says how users' data will be collected and used**
   - It explains how the company keeps confidential the information it collects on users
   - It says that the company will not share users' data with other sites or companies without permission
   - I don't know

2. Which of these best describes the main purpose of a **data use policy**? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   - It explains how your data will be protected
   - **It is a legal document that says how users' data will be collected and used**
   - It explains how the company keeps confidential the information it collects on users
   - It says that the company will not share users' data with other sites or companies without permission
   - I don't know

3. Imagine that a company's data use policy states that, "We store your **keys** for up to 30 days." What is the most accurate definition of **keys** in this statement? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   - Passwords
   - **Information necessary to read/decode messages**
   - Files stored on the Internet
   - Information that allows you to access your account without logging in every time
   - I don't know

4. Imagine that a company's data use policy states that, "We use **device fingerprinting** on our website." What is the most accurate definition of **device fingerprinting** in this statement? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   - Use of a fingerprint as a password for a device
   - **A method of tracking that involves using browser features to uniquely identify a specific device**
   - Providing information about a device to law enforcement to aid in solving a crime
   - Recording phone calls from a specific device
   - I don't know

5. If you send a **Do Not Track request**, are websites able to track you? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   - **Yes**
   - No
   - I don't know

6. Imagine that a company's data use policy states that, "We use **browser web storage** to store information about you." Where does this mean that information about you can be stored? Please select all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   For this question, respondents were able to select more than one answer. However, "I don't know" was an exclusive response. In order for a response be counted as correct in our analysis, respondents had to select the right answer (designated below in bold) and none of the other answers.
   - **On your machine or device**
   - On machines or servers that are onsite to the company
   - On machines or servers in your local area
   - On any machine or server
   - I don't know

7. Imagine that a company's data use policy states that, "We use **local storage** to store information about you." Where does this mean that information about you can be stored? Please select all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   For this question, respondents were able to select more than one answer. However, "I don't know" was an exclusive response. In order for a response be counted as correct in our analysis, respondents had to select the right answer (designated below in bold) and none of the other answers.
   - **On your machine or device**
   - On machines or servers that are onsite to the company

  - On machines or servers in your local area
  - On any machine or server
  - I don't know

8. Imagine that a company's data use policy states that, "We store **metadata** from your social media posts." What is the most accurate meaning of **metadata** in this statement? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   - Combined data from different sources
   - A better way of using data
   - A set of data or information
   - A large collection of data
   - **Data about data or files (e.g., time, size, location)**
   - I don't know

9. Which of the following might be examples of **metadata**? Select all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
   For this question, respondents were able to select more than one answer. However, "I don't know" was an exclusive response.
   - **Size of a file**
   - **Time a message was sent**
   - **Who an email or message was sent to**
   - **Location where a photo was taken**
   - **Who uploaded a shared file**
   - I don't know

10. Approximately what percent of companies do you think respect **Do Not Track requests**? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents selected a percentage value on a sliding scale. Answers to this question were not analyzed as there is no universally accepted figure regarding the percentage of companies that respect Do Not Track requests.

11. Imagine that a company says "We store a **cryptographic hash** of your password." What is the most accurate meaning of **cryptographic hash** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - We store an encrypted copy of your password
    - **We store information that allows us to check your password, but we do store not your password in plain text**
    - We store information that allows us to check your password, but no one will be able to learn your password from this information even if we get hacked
    - We store your password with extra security
    - I don't know

12. Imagine that a company says "We use **pixel tags** and other similar technologies." What is the most accurate definition of **pixel tag** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - A way to identify an image
    - A small piece of color in an image
    - **A way to monitor which sites users visit using a small image**
    - A way to identify which people or objects are in an image
    - I don't know

13. Imagine that a company says "We use **web beacons** and other similar technologies." What is the most accurate definition of **web beacons** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **A way to monitor which sites users visit using a small image**
    - A tool that allows companies to access data on a user's device
    - A tool that transports data between servers
    - Identifies data that belongs to a specific user
    - I don't know

14. When are **persistent cookies** deleted? Check all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents were able to select more than one answer. However, "I don't know" and "Never" were exclusive responses. In order for a response be counted as correct in our analysis, respondents had to select "They have an expiration date, and they are deleted then."
    - **You can manually delete them, and they are deleted then**
    - **They have an expiration date, and they are deleted then**
    - When you close your browser

- When you sign out of the website
- Until you close the tab or navigate away from the website
- Never
- I don't know

15. Does **private browsing/incognito mode** make you anonymous on the internet? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - Yes
    - **No**
    - I don't know

16. When using **private browsing/incognito mode**, where is data from your browsing session stored? Please select all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents were able to select more than one answer. However, "I don't know" and "Nowhere" were exclusive responses. In order for a response be counted as correct in our analysis, respondents had to at least one of the the right answers (designated below in bold).
    - On your machine or device
    - **On machines or servers belonging to the websites browsed**
    - **On machines or servers belonging to third parties**
    - Nowhere
    - I don't know

17. Which of the following are accurate statements about **end-to-end encryption**? Select all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents were able to select more than one answer. However, "I don't know" and "None of the above" were exclusive responses. In our analysis, we broke this question into 2 parts in order to study misconceptions about end-to-end as well as point-to-point encryption. In order for a response to be counted as correct in our end-to-end encryption analysis, respondents had to not check the answer choices "A company can read/decrypt my messages if their application uses end-to-end encryption" and "The government can get access to my messages through a legal request if an application uses end-to-end encryption." In order for a response to be counted as correct in our point-to-point encryption analysis, respondents had to either select "Applications that use end-to-end encryption are secure" and "Applications that use encryption (but not end-to-end encryption) are secure" or select neither of those two options.
    - **Applications that use end-to-end encryption are secure**
    - **Applications that use encryption (but not end-to-end encryption) are secure**
    - A company can read/decrypt my messages if their application uses end-to-end encryption
    - **A company can read/decrypt my messages if their application uses encryption (but not end-to-end encryption)**
    - The government can get access to my messages through a legal request if an application uses end-to-end encryption
    - **The government can get access to my messages through a legal request if an application uses encryption (but not end-to-end encryption)**
    - None of the above
    - I don't know

18. Can **de-identified information** ever be tied to you as an individual? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **Yes**
    - No
    - I don't know

19. Can **anonymized information** ever be tied to you as an individual? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **Yes**
    - No
    - I don't know

20. Can **aggregated information** ever be tied to you as an individual? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    Responses to this question were not considered in our correctness analysis as there is no standardized definition of the term "aggregated information."
    - Yes
    - **No**
    - I don't know

21. Imagine that a company says "We use standard technologies to perform **tracking**." What is the most accurate definition of **tracking** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **A way to monitor and record which sites you visit and/or your behavior on those sites**
    - A way to monitor and record your behavior on this company's site
    - A way to monitor and record your location
    - I don't know

For the following section of questions, respondents chose their response from a five-point Likert scale with the following options: "Very comfortable", "Somewhat comfortable", "Neither comfortable nor uncomfortable", "Somewhat uncomfortable", "Very uncomfortable". These options are listed for the first question.

22. How comfortable would you be using a product or service if the data use policy says "We use browser web storage to store information about you"?
    - Very comfortable
    - Somewhat comfortable
    - Neither comfortable nor uncomfortable
    - Somewhat uncomfortable
    - Very uncomfortable

23. How comfortable would you be using a product or service if the data use policy says "We use local storage to store information about you"?

24. How comfortable would you be using a product or service if the data use policy says "We use cloud storage to store information about you"?

25. How comfortable would you be using a product or service if the data use policy says "We store information about you on your machine or device"?

26. How comfortable would you be using a product or service if the data use policy says "We store information about you on our machines and servers"?

27. How comfortable would you be using a product or service if the data use policy says "We use device fingerprinting to identify you"?

28. How comfortable would you be using a product or service if the data use policy says "We use unique features of your browser such as your IP address, screen resolution, and operating system and language to identify you"?

29. How comfortable would you be using a product or service if the data use policy says "We collect Personally Identifiable Information (PII) about you to provide our Core Services"?

30. How comfortable would you be using a product or service if the data use policy says "We collect information such as your full name, address, and email to provide our Core Services"?

31. How comfortable would you be using a product or service if the data use policy says "We collect Personal Information about you to provide our Core Services"?

32. How comfortable would you be using a product or service if the data use policy says "We collect Sensitive Personal Information about you to provide our Core Services"?

33. How comfortable would you be using a product or service if the data use policy says "We collect information such as your race and ethnicity, political and religious views, or information about your health to provide our Core Services"?

34. How comfortable would you be using a product or service if the data use policy says "We collect Public Information about you to provide our Core Services"?

35. How comfortable would you be using a product or service if the data use policy says "We collect information such as your username, profile picture, or social media posts to provide our Core Services"?

36. How comfortable would you be using a product or service if the data use policy says "We collect Account Information about you to provide our Core Services"?

37. How comfortable would you be using a product or service if the data use policy says "We collect information such as your username or email address to

provide our Core Services”?

38. How comfortable would you be using a product or service if the data use policy says “We do not recognize or respond to browser-initiated Do Not Track signals”?

39. How comfortable would you be using a product or service if the data use policy says “We do not share the content of your messages, but we do share some metadata with third parties”?

40. How comfortable would you be using a product or service if the data use policy says “We do not share the content of your messages, but we do share information such as the length of your messages, the times of your messages were sent, and the people you sent the messages to with third parties”?

41. How comfortable would you be using a product or service if the data use policy says “We collect information about your device attributes to provide our Core Services”?

42. How comfortable would you be using a product or service if the data use policy says “We collect information about your app and file names and types to provide our Core Services”?

43. How comfortable would you be using a product or service if the data use policy says “We collect information about applications you have installed on your device to provide our Core Services”?

44. How comfortable would you be using a product or service if the data use policy says, “We use a form of encryption to secure your messages that ensures that hackers and eavesdroppers cannot read your messages. We only decrypt and share the contents of your messages if we receive a legal request from the government”?

45. How comfortable would you be using a product or service if the data use policy says, “We do not offer end-to-end encryption, but we use encryption to secure your messages”?

46. How comfortable would you be using a product or service if the data use policy says, “We use a form of encryption to secure your messages that guarantees that only you (and whoever you are talking to) can decode the content of your messages”?

47. How comfortable would you be using a product or service if the data use policy says, “We use end-to-end encryption to secure your messages”?

48. How comfortable would you be using a product or service if the data use policy says “We only share de-identified information with third parties”?

49. How comfortable would you be using a product or service if the data use policy says “We only share anonymized information with third parties”?

50. How comfortable would you be using a product or service if the data use policy says “We only share aggregated information with third parties”?

51. How comfortable would you be using a product or service if the data use policy says “We use web beacons to collect information”?

52. How comfortable would you be using a product or service if the data use policy says “We use pixel tags to collect information”?

53. How comfortable would you be using a product or service if the data use policy says “We monitor user behaviors across different websites using invisible images and other website elements”?

54. How comfortable would you be using a product or service if the data use policy says “We use session cookies to store information about you”?

55. How comfortable would you be using a product or service if the data use policy says “We use persistent cookies to store information about you”?

56. How comfortable would you be using a product or service if the data use policy says “We store information about you in small files that remain on your hard drive until deleted or expired”?

57. How comfortable would you be using a product or service if the data use policy says “We store information about you in small files that remain on your hard drive until you log out”?

Users were shown one of the two following questions.

58. Imagine that a company says "We share your **personally identifiable information (PII)** with our business partners." Please list three types of data that could be shared under this policy. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - *Users were provided with a textbox to enter their response*

59. Imagine that a company says "We share your **personal information** with our business partners." Please list three types of data that could be shared under this policy. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - *Users were provided with a textbox to enter their response*

A.3 Follow-up Study

For the following section of questions, respondents chose their response from a five-point Likert scale with the following options: “Very likely“, “Somewhat likely“, “Neither likely nor unlikely“, “Somewhat unlikely“, “Very unlikely“. These options are listed for the first question.

1. How likely would you be to accept a data use policy that says, “We use web beacons to collect information”?
    - Very likely
    - Somewhat likely
    - Neither likely nor unlikely
    - Somewhat unlikely
    - Very unlikely

2. How likely would you be to accept a data use policy that says, “We monitor user behaviors across different websites using invisible images and other website elements”?

3. How likely would you be to accept a data use policy that says, “We use session cookies to store information about you”?

4. How likely would you be to accept a data use policy that says, “We store information about you in small files that remain on your hard drive until you log out”?

5. How likely would you be to accept a data use policy that says, “We use persistent cookies to store information about you”?

6. How likely would you be to accept a data use policy that says, “We store information about you in small files that remain on your hard drive until deleted or expired”?

7. How likely would you be to accept a data use policy that says, “We do not offer end-to-end encryption, but we use encryption to secure your messages”?

8. How likely would you be to accept a data use policy that says, “We use a form of encryption to secure your messages that ensures that hackers and eavesdroppers cannot read your messages. We only decrypt and share the contents of your messages if we receive a legal request from the government”?

9. How likely would you be to accept a data use policy that says, “We use end-to-end encryption to secure your messages”?

10. How likely would you be to accept a data use policy that says, “We use a form of encryption to secure your messages that guarantees that only you (and whoever you are talking to) can decode the content of your messages”?

11. How comfortable would you be using a product or service if the data use policy says “We do not recognize or respond to browser-initiated Do Not Track signals”?

12. How likely would you be to accept a data use policy that says, “We only share aggregated information with third parties”?

13. How likely would you be to accept a data use policy that says, “We only share combined summary information with third parties”?

14. How likely would you be to accept a data use policy that says, “We only share anonymized information with third parties”?

15. How likely would you be to accept a data use policy that says, “We remove any data that can be linked to you as an individual from all information shared

with third parties"?

16. How likely would you be to accept a data use policy that says, "We collect Personal Information about you to provide our Core Services"?

17. How likely would you be to accept a data use policy that says, "We collect Personally Identifiable Information (PII) about you to provide our Core Services"?

18. How likely would you be to accept a data use policy that says, "We collect information such as your full name, address, and email to provide our Core Services"?

19. How likely would you be to accept a data use policy that says, "We perform tracking to provide our Core Services"?

20. How likely would you be to accept a data use policy that says, "We collect information about the sites you visit and your behavior on those sites to provide our Core Services"?

For the following questions, the correct response will be indicated using bold font. Unless otherwise indicated, respondents selected one response per question. For multiple response questions, the answers that are right are indicated in bold; the definition used to code responses as correct or incorrect is specified above each multiple response question.

21. Imagine that a company says "We use **web beacons** and other similar technologies." What is the most accurate definition of **web beacons** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **A way to monitor which sites users visit using a small image**
    - A tool that allows companies to access data on a user's device
    - A tool that transports data between servers
    - Identifies data that belongs to a specific user
    - I don't know
    - Other *(users were provided with a textbox to enter their response)*

22. When are **persistent cookies** deleted? Check all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents were able to select more than one answer. However, "I don't know" and "Never" were exclusive responses. In order for a response be counted as correct in our analysis, respondents had to select "They have an expiration date, and they are deleted then."
    - **You can manually delete them, and they are deleted then**
    - **They have an expiration date, and they are deleted then**
    - When you close your browser
    - When you sign out of the website
    - Until you close the tab or navigate away from the website
    - Never
    - I don't know

23. When are **session cookies** deleted? Check all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents were able to select more than one answer. However, "I don't know" and "Never" were exclusive responses. In order for a response be counted as correct in our analysis, respondents had to not select any of the three responses associated with misconceptions: "They have an expiration date, and they are deleted then.", "Until you close the tab or navigate away from the website", or "Never".
    - **You can manually delete them, and they are deleted then**
    - They have an expiration date, and they are deleted then
    - **When you close your browser**
    - **When you sign out of the website**
    - Until you close the tab or navigate away from the website
    - Never
    - I don't know

24. Which of the following are accurate statements about **end-to-end encryption**? Select all that apply. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    For this question, respondents were able to select more than one answer. However, "I don't know" and

"None of the above" were exclusive responses. In our analysis, we broke this question into 2 parts in order to study misconceptions about end-to-end and point-to-point encryption. In order for a response to be counted as correct in our end-to-end encryption analysis, respondents had to not check the answer choices "A company can read/decrypt my messages if their application uses end-to-end encryption" and "The government can get access to my messages through a legal request if an application uses end-to-end encryption." In order for a response to be counted as correct in our point-to-point encryption analysis, respondents had to either select "Applications that use end-to-end encryption are secure" and "Applications that use encryption (but not end-to-end encryption) are secure" or select neither of those two options.

- **Applications that use end-to-end encryption are secure**
- **Applications that use encryption (but not end-to-end encryption) are secure**
- A company can read/decrypt my messages if their application uses end-to-end encryption
- **A company can read/decrypt my messages if their application uses encryption (but not end-to-end encryption)**
- The government can get access to my messages through a legal request if an application uses end-to-end encryption
- **The government can get access to my messages through a legal request if an application uses encryption (but not end-to-end encryption)**
- None of the above
- I don't know

25. Imagine that a company says "We only share **aggregated information** with third parties." What is the most accurate definition of **aggregated information** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **Combined summary information**
    - Any collected information
    - Fully representative data
    - Data that cannot be linked to an individual
    - I don't know
    - Other *(users were provided with a textbox to enter their response)*

26. Can **anonymized information** ever be tied to you as an individual? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **Yes**
    - No
    - I don't know

27. Imagine that a company says "We use standard technologies to perform **tracking**." What is the most accurate definition of **tracking** in this context? Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - **A way to monitor and record which sites you visit and/or your behavior on those sites**
    - A way to monitor and record your behavior on this company's site
    - A way to monitor and record your location
    - I don't know
    - Other *(users were provided with a textbox to enter their response)*

28. Imagine that a company says "We share your **personally identifiable information (PII)** with our business partners." Please list three types of data that could be shared under this policy. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - *Users were provided with a textbox to enter their response*

29. Imagine that a company says "We share your **personal information** with our business partners." Please list three types of data that could be shared under this policy. Please respond using only your prior knowledge; do NOT consult search engines such as Google.
    - *Users were provided with a textbox to enter their response*

A.4 Demographic Questions

The following demographic questions were included at the end of the main survey and the follow-up survey. Demographic questions were not asked to the respondents of the pilot survey. Unless otherwise specified, respondents picked one answer to each question.

1. Do you work in the tech industry or have you studied computer science or a related field?
   - Yes
   - No

2. Have you read a privacy policy in the past 3 months?
   - Yes
   - No
   - I don't know

3. How often do you read privacy policies?
   - Always
   - Usually
   - About half the time
   - Rarely
   - Never

4. When you read privacy policies, which best describes the way that you read them?
   For this question, respondents were able to select more than one answer.
   - Read thoroughly
   - Skim
   - Look for key words
   - Look at section headers
   - Use third-party tools

5. What is your current age?

6. What is your gender?

7. Choose one or more races that you consider yourself to be.

8. Do you consider yourself to be Hispanic?

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

10. Please indicate the answer that includes your entire household income in (previous year) before taxes.

11. In which state do you currently reside?