

Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach

Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, Lorrie Faith Cranor
Carnegie Mellon University
pkelley@cs.cmu.edu, ljcesca@cmu.edu, jbresee@andrew.cmu.edu, lorrie@cs.cmu.edu

ABSTRACT

Earlier work has shown that consumers cannot effectively find information in privacy policies and that they do not enjoy using them. In our previous research we developed a standardized-table format for privacy policies. We compared this standardized format, and two short variants (one tabular, one text) with the current status quo: full-text natural-language policies and layered policies. We conducted an online user study of 764 participants to test if these three more-intentionally designed, standardized privacy policy formats, assisted by consumer education, can benefit consumers. Our results show that standardized privacy policy presentations can have significant positive effects on accuracy and speed of information finding and on reader enjoyment of privacy policies.

Author Keywords

Privacy, policy, P3P, information design, standardization

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces; K.4.1 Computers and Society: Public Policy Issues-Privacy

General Terms

Human Factors, Design, Security.

INTRODUCTION

Consumer testing has shown privacy policies are unusable. An online survey of over 700 participants that tested policies from six companies in three existing formats found “participants were not able to reliably understand companies’ privacy practices with any of the formats” and “all formats and policies were similarly disliked” [10].

In the United States, Internet privacy remains almost entirely unregulated, which means consumers who wish to find websites with privacy-protective practices must be able to read and understand privacy policies. However, policies are

commonly long, textual explanations of data practices, most frequently written by lawyers to protect companies against legal action.

We used an iterative, user-centered design process to develop a more compelling and informative privacy policy format. We conducted a large online user study to evaluate three variants of our privacy policy format as well as two formats commonly used by large corporate websites today.

In the next section, we detail related work on the drawbacks of current privacy policies and describe other efforts to design better policy formats. We explain each of the five formats we tested, followed by accuracy, comparison, timing, and enjoyability results. We then discuss the implications of this work with some future directions.

RELATED WORK

We discuss work highlighting the problems with current online privacy policies, as well as work aimed at making privacy policies more useful.

Privacy Policies are Unusable

Reading current online privacy policies is challenging and time consuming. It is estimated that if every Internet user read the privacy policies for each site they visited, this lost time would cost about \$781 billion per year [9]. It is admittedly unrealistic to expect users to read and understand the privacy policy of every site they visit. Most policies are written at a level that is suitable for consumers with a college-level education and use specific domain terminology that consumers are frequently unfamiliar with [4, 10]. Rarely is a policy written such that consumers have a clear understanding of where and when their data is collected, how and by whom it will be used, if it will be shared outside of the entity that collected it, and for how long and in what form it will be stored. Even worse, it is unlikely consumers will even read a single policy given a widespread consumer belief that there are no choices when it comes to privacy: consumers believe they do not have the ability to limit or control companies’ use of their information [7].

A Privacy “Nutrition Label”

We previously proposed a privacy “nutrition label” to assist consumer understanding of privacy policies [6]. This work was influenced by studies of the design and consumer acceptance of nutrition labeling programs [3, 1]. This tabular privacy format was designed to enhance user understanding of privacy practices, increase the speed of information

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2010, April 10-15, 2010, Atlanta, Georgia, USA.

Copyright 2010 ACM 978-1-60558-929-9/10/04...\$10.00.

finding, and facilitate policy comparisons.¹ We previously tested this approach in a series of focus groups and a small 24-participant laboratory study. In this paper we describe a much larger online evaluation that compares two variants of this approach with a standardized-text format we developed as well as with two formats currently in use.

Layered Policy Notices

Layered privacy notices, popularized by the law firm Hunton & Williams [12, 13], provide users with a high-level summary of a privacy policy. The design is intended to be a “standardized” format; however, the only standard components are a tabular page layout and mandatory text for the section headers. Other design details and the text of each section are left to the discretion of each company. Additionally, the amount of information to include in a layered notice is left up to each company, with layered notices requiring consumers to click through to the full text policy.

Financial Policy Notices

The Gramm-Leach-Bliley Act (GLBA), passed in 1999, contains the Financial Privacy Rule, which requires that financial institutions disclose their privacy policy “at the time of establishing a consumer relationship...and not less than annually” [14]. Financial institutions must comply with requirements on what they disclose, but their disclosures may be in any format.

In 2004, seven federal agencies launched a multi-phase initiative to “explore the development of paper-based, alternative financial privacy notices...that are easier for consumers to understand and use” [7]. The Kleimann Communication Group (KCG) conducted the first-phase, which tested multiple designs across seven cities and surveyed consumers about financial privacy notices. In their final report the KCG proposed a three-page design for further evaluation [7].

In December 2008, the second phase report was published by Levy and Hastak [8]. This report detailed a 1032-participant mail/interview study that tested four privacy notice formats. Two of the four notices were developed by the KCG, with contextual information and an opt-out form. The KCG table notice displayed financial institutions’ practices in a grid format, whereas their prose notice used a bulleted list. The other two notices were both text-based, with the “current notice” mimicking notices that financial institutions currently use, and the “sample clause” notice generated from GLBA-provided phrases. Levy and Hastak concluded that the KCG table notice performed best. They attributed this improvement to an increased level of comprehension, given the table notice’s “[provision] of a fuller context...the part-to-whole display approach seems to help consumers focus on information sharing as important and differentiating features of financial institutions.” However, on several study questions other notices, notably the sample clause notice, tested best.

¹The tabular format can be filled in automatically if a site uses a W3C protocol called the Platform for Privacy Preferences (P3P) [15, 2]. P3P specifies a standardized format for machine-readable privacy policies. Previous work evaluated a much more complicated tabular format based on P3P policies [11].

POLICY FORMATS

We tested five privacy policy formats: standardized table, standardized short table, standardized short text, full policy text, and layered text. Three of these formats are standardized and were created by our lab using an iterative design approach. Of these, two are tabular and one is textual. Two explicitly describe absent information and one presents it in the context of the policy. Each of these formats is followed by a list of 16 definitions of privacy terms, consistent across formats. These definitions define the row and column headers in the tables and the text tokens in the standardized short text. They also assist with understanding the terminology used in the survey questions.

Full Policy Text

Natural-language, full-text policies are the de facto standard for presenting privacy policy information online. For this experiment, we selected four policies from well-known companies. We stripped each policy of all formatting, retaining only internal hyperlinks to reference other areas of the policy, if available in the original. We anonymized all identifying branding, including company and product names, affiliates, and contact information.

Standardized Table

The standardized-table format, (Figure 1, left) has ten rows, each representing a data category the company may collect, four columns detailing the ways that data may be used, and two columns representing ways that data may be shared outside the company. This table is filled with four symbols, dark red to represent that your data may be used or collected in that way, light blue to represent that your data will not be used or collected in that way, and two intermediate options labeled “opt in” and “opt out.” This is a variant of the “nutrition label” format discussed above [6], modified based on follow-up design iterations.

Standardized Short Table

The standardized short table (Figure 1, right) is a shortened version of our proposed tabular approach, which removes the data categories (rows) that are never collected by a company. These removed data categories are listed immediately following the table to maintain a holistic understanding of a company’s privacy practices. While the removal of data categories allows the table to fit into a smaller area, it may make comparisons between policies less straightforward.

Standardized Short Text

We created a short, natural-language format (Figure 2) by translating each row in the standardized short table into an English statement, using the column and row headers from the table to form each statement. Similar rows are merged into combined statements for brevity. This format allows us to compare textual and tabular formats directly.

Layered Text

Finally, we tested the layered privacy notice — a summarized, one-screen privacy policy in a tabular format that links to the full natural-language policy [12, 13]. Layered policies

Acme

information we collect	ways we use your information				information sharing	
	provide service and maintain site	marketing	telemarketing	profiling	other companies	public forums
contact information		opt out	opt out			
cookies						
demographic information		opt out	opt out			
financial information						
health information						
preferences		opt out	opt out			
purchasing information		opt out	opt out			
social security number & gov't ID						
your activity on this site		opt out	opt out			
your location						

Access to your information
This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site
Please email our customer service department

acme.com
5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com

Acme

information we collect	ways we use your information				information sharing	
	provide service and maintain site	marketing	telemarketing	profiling	other companies	public forums
contact information		opt out	opt out			
cookies						
demographic information		opt out	opt out			
preferences		opt out	opt out			
purchasing information		opt out	opt out			
your activity on this site		opt out	opt out			

Information not collected or used by this site: social security number & government ID, financial, health, location.

Access to your information
This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site
Please email our customer service department

acme.com
5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com

opt out we will collect and use your information in this way

by default, we will collect and use your information in this way unless you tell us not to by opting out

opt in we will not collect and use your information in this way

by default, we will not collect and use your information in this way unless you allow us to by opting in

Figure 1. An example of a standardized table is shown on the left, and a standardized short table on the right. The comparison highlights the rows deleted to “shorten” this version. These deleted rows are listed directly below the table. While both formats contain the legend (bottom right), it is displayed only on the right here due to space constraints.

have already been deployed by major corporations, making them a viable, real world summary format for privacy policies. These policies were stripped of brand information, but the formatting and styles were retained.

METHODOLOGY

We conducted an online user study in summer 2009 using Amazon’s Mechanical Turk and a tool we developed called Surveyor’s Point. Mechanical Turk offers workers the ability to perform short tasks and get compensated. People can place jobs through Mechanical Turk, specifying the number of workers they are looking for, necessary qualifications, the amount they are willing to pay, and details about the task. Mechanical Turk payments are generally calibrated for the length of the task. For our approximately 15-minute study, we paid \$0.75 on successful completion.

We developed a custom survey-management tool called Surveyor’s Point to facilitate our data collection. Our implementation allows us to show respondents a single question on the screen along with links for switching back and forth between two policies within a single browser window. This allowed us to track the number of users who looked at each policy and the number of times they switched between them. Additionally, Surveyor’s Point allowed us to collect the amount of time that users spent reading the policies, as well as information about whether they clicked through to opt-out forms, to additional policy information links, or from a layered notice through to the full text policy.

In preparation for this study we first performed three smaller pilot tests of our survey framework. We ran our pilot studies with approximately thirty users each, across 2-3 conditions. Our pilot studies helped us to finalize remaining design decisions surrounding the standardized short table, refine our questionnaire, and test the integration of Surveyor’s Point with Mechanical Turk.²

We then conducted our large-scale study and completed the analysis with 764 participants (409 female, 355 male), randomly assigned to five conditions (see Table 1): full policy text, standardized table, standardized short table, standardized short text, and layered text. We dropped 25 additional participants from the study prior to analysis due to incomplete data or for completing the study in an amount of time that indicated inadequate attention to the task (defined as time on task that was two standard deviations lower than the mean). We chose a between-subjects design to remove learning effects and ensure the study could be completed within about 15 minutes. Participants in each condition followed the same protocol; only the policy format differed.

Policies

We selected policies for the study from companies that consumers would conceivably interact with. We narrowed our

²The two systems are linked using a shared key that Surveyor’s Point generates on the completion of our survey, which a participant then enters back into Mechanical Turk. This allows us to link an entry in Mechanical Turk with an entry in Surveyor’s Point and verify the worker completed the survey before payment.

Acme

Acme will collect your contact information. They will use this information for providing you service and maintaining the site and profiling. They will also use this information for marketing and telemarketing unless you opt out. They will share this information with other companies unless you opt out. They will share this information on public forums if you opt in.

Acme will collect your activity on this site, demographic information, your health information, and cookie information. They will use this information for providing you service and maintaining the site and profiling. They will also use this information for marketing and telemarketing unless you opt out. They will not share this information.

Acme will collect your preferences and your purchase information. They will use this information for providing you service and maintaining the site and profiling. They will also use this information for marketing and telemarketing unless you opt out. They will share this information on public forums if you opt in.

Information not collected or used by this site:
financial, SSN or government ID, and location.

Access to your information
This site gives you access to your contact data and some of its other data identified with you

acme.com
5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com

Figure 2. An example of the standardized short-text format.

search by selecting companies that had websites with over one million views per month³ and were P3P enabled. Additionally, we selected two companies with layered policies deployed on their websites. The four policies we selected were Microsoft, IBM, Target, and Disney. We randomly assigned half our participants in each condition to answer questions about anonymized versions of the Target and Disney privacy policies (Group A), and assigned the other half of our participants to answer questions about anonymized versions of the Microsoft and IBM privacy policies (Group B). By having participants answer questions about policies from different companies we are able to gain insights into where our results may be due to features of a specific policy and where they may be generalizable across many policies.

The policies range in length, but are representative of common practices. Table 2 summarizes word counts across the full text, standardized short text, and layered policies.

Study Questions

Our study was designed to include questions across seven blocks, with time-to-task-completion recorded for each task:

1. *Demographics*: We collected standard information about our participants: gender, age, and current occupation.
2. *Internet and Privacy*: We asked participants four questions to better understand their Internet usage and their prior knowledge of privacy.
3. *Simple Tasks*: We showed participants the “Acme” policy and asked six questions pertaining to it. We refer to these information-finding tasks as simple questions as each question can be answered by looking at a specific row or column in the table. The answer options for these questions

³We used data from <http://www.quantcast.com/> to select these websites.

	Std. Table	Std. Short Table	Std. Short Text	Full Policy Text	Layered Text
Participants	188	167	169	162	78

Table 1. Study participants across formats (N=764).

	Pol. 1	Pol. 2	Pol. 3	Pol. 4
Full Policy Text	2127	6257	4399	2912
Std. Short Text	175	127	108	90
Layered Text			409	800

Table 2. Word counts across the three text variants. Note that the definitions that we append to each policy format add an additional 433 words.

(with the exception of question four) were “Yes,” “No,” or “The policy does not say.”

4. *Complex Tasks*: We asked participants six questions pertaining to the Acme policy. We refer to these information-finding tasks as complex questions because each dealt with some interaction between some category of data and either data use or data sharing. The answer options for these were “Yes,” “No,” “Yes, unless I tell them not to,” “Only if I allow them to,” or “The policy does not say.”
5. *Single Policy Likeability*: After completing the simple and complex tasks, we presented a series of 7-point Likert questions for qualitative feedback on the format.
6. *Comparison Tasks*: We showed participants a notice stating that they would now be comparing two policies: the Acme policy, which they had already seen, with the policy for the Bell Group. We asked three information-finding questions and two preference questions that required looking at both policies.
7. *Policy Comparison Likeability*: We asked participants three more Likert questions to collect qualitative feedback on the task of comparing two policies.

Analysis

Table 3 shows the gender and age breakdown of the participants, as well as the number of privacy policies participants reported reading in the previous six months. 56.4% of our participants reported reading at least one policy in the previous six months. Participants reported that they had the following occupations: student (17.3%); science, engineering, IT (16.5%); unemployed (13.2%); business, management, and finance (9.9%); education (7.3%); administrative support (6.7%); service (4.8%); art, writing, and journalism (4.7%); retired (2.4%); medical (2.0%); skilled labor (1.8%); legal (1.3%); and other (9.3%). 2.7% declined to answer.

While this sample population from Mechanical Turk is not a completely representative sample of American Internet users, it is a useful population to study. Our participants appear to read privacy policies more than the general population; however, it is possible that participants, realizing that we were going to ask them to compare privacy policies,

	number	percentage
<i>Total Participants</i>	764	
<i>Gender</i>		
Male	355	46.5%
Female	409	53.5%
<i>Age</i>		
18-28 years old	321	42.0%
28-40 years old	250	32.7%
40-55 years old	116	15.2%
55-70 years old	31	4.1%
did not disclose	46	6.0%
<i>Number of Privacy Policies Read in the last 6 months</i>		
Never read a privacy policy	189	24.7%
None in the last six months	130	17.0%
1 policy	100	13.1%
2-5 policies	230	30.1%
5+ policies	101	13.2%
did not disclose	14	1.8%

Table 3. Participant Demographics across conditions

may have sought to seem more knowledgeable about privacy policies. Nutritional and drug labeling literature reports that standardization efforts assist most those who seek out the information [3]. If participants on Mechanical Turk do read more privacy policies than the general population then we may be refining our label to help the group that will be most likely to leverage privacy policy information.

We began our analysis by marking all answers to questions as correct or incorrect (although, as we will discuss later, in some cases there were varying degrees of correctness). We also computed the time it took participants to answer each question. We performed the following statistical analysis:

1. We performed an ANOVA on the average accuracy scores, totaled for each participant, across conditions. We performed additional t-tests for paired comparisons.
2. We also scored each simple, complex, and comparison task individually for accuracy. We performed factorial logistic regressions across the policy formats.
3. We performed ANOVAs on the log normalized timing information for the above tasks.
4. We performed ANOVAs for the Likert questions.

RESULTS

We describe our big-picture accuracy results, followed by a more in-depth analysis of each policy format, summarize our timing results, and conclude with an analysis of participants' enjoyment of reading privacy policies.

Overall Accuracy Results

Each participant completed 15 information-finding tasks. We scored each participant on a scale from 0 to 15, based on the number of these questions they answered correctly, and averaged those scores across conditions. Note, correct answers

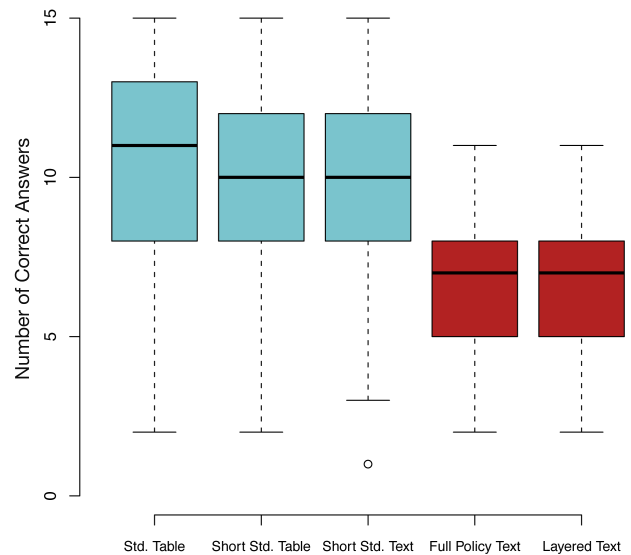


Figure 3. Accuracy results for each of the five policy formats.

varied between conditions since policy content varied between conditions. We present these aggregate results in Figure 3. This summary shows a large divide between the standardized and non-standardized formats (ANOVA significant at $p < 0.05$, $F(4, 1094) = 73.75$). The three standardized formats, scoring 62-69%, are shown in light blue; while the two real-world text policies, scoring 43-46%, are shown in red. The standardized policies significantly outperformed the full-text policy (standard table vs. full text, $t(510) = -14.4$, standardized short table vs. full text $t(490) = 12.9$, and standardized short text vs. full text $t(491) = 14.3$, were all significant at $p < 0.05$). The layered format did not perform significantly differently from the full text policy ($p = 0.83$, $t(314) = -0.21$).

Accuracy Results by Format

We analyzed the results on a per-question basis to gain insights into the strengths and weaknesses of each format. We performed factorial logistic regressions with the standardized table as the base for comparison across formats. The content of the questions and further results are presented in the Appendix. Here we describe the overall performance of each format and highlight specific questions to illustrate features of each format. Where statistically significant results on the user data across formats are discussed in the section below, the values of the statistical tests ranged from $z = 1.97$ to $z = 7.52$, and the level of significance was at least $p < 0.05$. For more detailed information on the statistical results, see our technical report [5].

Standardized Table

All of our standardized formats benefit from structured information presentation, clear labeling of information that is not used or collected, standardized terminology to minimize length and increase the clarity of the text, and definitions of standardized terms. In addition, the standardized table's tabular display presents a holistic view of the policy.

Overall, we did not find significant differences between the standardized formats, however the standardized table significantly outperformed at least one of the other standardized formats for some policies in 10 of our 13 questions (3, 5-8, 10, 12, 14-16). It was only significantly outperformed twice (9, 16).

There were some questions that proved difficult across all of the standardized formats. For example, question 3 asked: “Does the policy allow Acme to collect information about your current location?” This information is not collected in any of the policies. The standardized table displays a blank row explicitly labeled “your location,” yet more than half our participants answered incorrectly. We believe this was due to participants’ misunderstanding location information to be related to the row labeled “contact information.”

Participants performed much better on the other five simple tasks, achieving accuracy results from 73% to 90% across those questions. For the complex questions we saw accuracy results drop; however, the standardized table still fared much better than the full policy text. Question 15, which concerns “sensitive information,” found overall accuracy between 72% and 84% for people who had already correctly answered question 6 (concerning only medical information). Participants in the standardized-table condition scored correctly more than three times as frequently as did those in the full-policy-text condition on this question that involved multiple data types and comparing two different policies.

Standardized Short Table

Unlike the standardized table, the standardized short table does not show blank rows for uncollected data categories. Instead the standardized short table lists these categories in a textual notice below the table. The standardized short table showed highly similar overall results to the standardized table, though it did perform significantly better for one question (16) and significantly worse for four (3, 6, 14, 15). Importantly, this format still performed significantly better than the full policy text in overall accuracy.

This table variant was created to reduce the size of the table, under the assumption that a less space-consuming table could be an advantage over the large table. However, we were concerned that removing rows would make policies more difficult to compare. The full table performs significantly better in two (14, 15) of our three comparison questions; the standardized short table only in one (16).

The text notice underneath the table that describes the absence of information may require further testing. In question 6, which asks about the collection of medical information, the standardized short-table format performed poorly (59%) when medical information was absent; however, the standardized short-text format performed best (81%). Since both formats presented missing information in an identical manner using the same font size, we expect that the difference is due to users being less likely to notice the text when it is under the table than when it is presented in a font size larger than most of the rest of the text on the page.

Standardized Short Text

The standardized short text is a direct translation of the standardized short table into text. Rows are grouped by similarity and transformed into sentences. This format did not perform significantly differently from the standardized table overall, but was significantly outperformed in eight questions (5-8, 10, 12, 14, 16), and performed better than the standardized table in two questions (9,16). Similar to the other standardized formats, it performed significantly better than the full policy text overall. The standardized short-text format is the simplest format we tested. It is compact and requires a participant to understand no symbols, colors, or tables.

One drawback of the standardized short-text format is that the length of the text grows with the complexity of the policy. In the longest of our standardized short-text policies, the text “Cookie information” is in the middle of a substantial block of text. Only 73% of participants assigned to this policy answered question 5 (“may Acme store cookies on your computer?”) correctly (compared to 80-96% across other standardized conditions). At only 175 words, this text seems quite short and participants may not use the search functionality of their browser to find the word “cookie.” This is speculative; however, and a follow up study with the paragraphs rearranged may lead us to better understand whether blind spots exist in this format.

We saw evidence of participants in the standardized short-text condition misreading text in other questions as well. Question 8 asked: “Does the policy allow Acme to share your home phone number with other companies?” Looking at the standardized short-text condition responses for a policy where the answer was “Yes,” we see that only 20% answered correctly, while 47% answered “Yes, unless I tell them not to,” implying they believed an option existed where it did not. We suspect this comes from misreading the text, as an option was mentioned for another type of data later in that same paragraph. For question 14 we see a similar pattern. The standardized short text received only 20% accuracy, with 49% of respondents incorrectly answering that neither company gave options regarding cookies.

While the standardized short-text format does perform well, drilling down into these questions shows that it may not scale well, with complex options resulting in longer paragraphs with confusing details.

Full Policy Text

As discussed above, the full-policy-text format had worse overall accuracy scores than the standardized formats. We have seen that this format, the de-facto standard on the Internet today, has led participants in our study to search multiple pages of text for the absence of a single point of information; uses terms, descriptions, and definitions that may be hard to find or confusing to consumers; and leads to searching multiple sections to find answers which in other formats prove to be much more attainable.

Question 4 asked: “Based on the policy will Acme register their secure certificate with VeriSign or some other com-

Average Timing Information (in seconds)

Question #	Std. Table		Std. Short Table		Std. Short Text		Full Policy Text		Layered Text		F-Statistic (dof)
	avg.	σ	avg.	σ	avg.	σ	avg.	σ	avg.	σ	
1-6	236	205	210	103	237	174	367	248	317	406	15.994 (4,756)
7-12	176	194	135	73	163	122	249	358	186	210	8.751 (4,756)
13-17	158	125	148	97	169	122	236	227	187	157	5.094 (4,756)
Full Study	912	572	852	407	938	515	1267	810	1089	768	11.273 (4,756)

Table 4. Average time per condition in seconds for questions 1-6 (simple), 7-12 (complex), and 13-17 (comparison), as well as total. While there were significant differences across formats, overall significant differences between the standardized formats were not observed.

pany?” Participants in the three standardized conditions correctly answered, “The policy does not say,” with 79-88% accuracy. However accuracy dropped to 31-52% for the full-policy-text and layered-text conditions. Neither policy mentioned Verisign or any other certificate registrar, nor did either policy contain the word “certificate.” We attributed this difficulty to the full-policy-text format forcing users to scan for the absence of information over several pages of text.

Worse yet, finding information by looking for specific terms proved difficult. Question 6 asked if medical information was collected, yet in one policy only 49% of participants correctly identified that medical information was collected even though the policy referenced “counseling from pharmacists,” an “online prescription refill service,” and “prescription medications.”

Moving on to complex tasks, question 7 asked: “Does the policy allow Acme to share some of your information on public bulletin boards?” In a tabular format, this question required the participants to find the column for public sharing, and see if any type of data would be allowed. Across the standardized formats accuracy ranged from 59% to 76%. Participants given the full-policy-text format have strikingly low results for this question (16-34%), regardless of the policy they were assigned. Many incorrectly reported the policy did not specify whether the information would be shared on public bulletin boards, indicating they were unable to find the section of the policy that discussed this.

The first content comparison task, Question 14 asked: “Does either company give you options with regards to cookies?” For the full policy text, 55% of the participants reviewing one set of policies believed that both companies provided options regarding cookies. This means that they incorrectly answered that the Acme policy had options regarding cookies (it did not). Searching for “cookie” in that text brings up a section entitled “Use of Cookies,” under which the fourth paragraph reads: “You have the ability to accept or decline cookies. Most Web browsers automatically accept cookies, but you can usually modify your browser setting to decline cookies if you prefer...” Although this sounds like an option regarding the use of cookies, it is not one that Acme provides, rather it is a function of most web browsers. The text in this case, as in many other instances, was confusing.

Layered Text

Layered policies, by design, do not necessarily provide a complete understanding of a company’s practices. Each company decides what information is most relevant to include. Furthermore, companies may use the same language that exists in the full policy text — language that was problematic in the full-policy-text condition. The layered-text format did not strongly differentiate itself from the full text policy in any of the detailed per-question results that we examined. Given that the participants in this condition had access to the full text policy we expected this, though only 25 of 78 participants ever clicked through to the full policy text.

Timing Results

We examined completion times for the simple, complex, and comparison tasks, as presented in Table 4. Note, time for comparison tasks includes both information-finding tasks and preference questions. We tested statistical significance using ANOVA on the log-normalized time information across policy formats. For each of these three groups of questions, as well as the overall study completion time there were statistically significant differences across policy formats ($p < 0.0001$ for questions 1-6, 7-12, 13-17, and overall). The standardized formats significantly outperformed the full policy text in overall time (standard table vs. full text, $t(348) = 5.36$, standardized short table vs. full text $t(327) = -6.01$, and standardized short text vs. full text $t(329) = -4.55$, were all significant at $p < 0.05$). The layered format was also significantly faster than the full text policy ($p = 0.025$, $t(238) = 2.25$). The standardized formats, on average were between 26-32% faster than the full text policy, and 22% faster than the layered text policy.

Enjoyability Results

For the most qualitative of our measures, we asked the participants how they felt about looking at privacy policies. We asked six 7-point Likert scale questions after they completed the single-policy tasks and three more after they completed the policy-comparison tasks. The results are summarized in Table 5. The Likert scale ranged from “Strongly Disagree” (1) to “Strongly Agree” (7), where higher scores indicate more user enjoyment or perceived usefulness of the format. While there were significant differences for nearly all the Likert questions, we will not go into the details of each question, but average across the two groups of questions.

Question Number	Std. Table	Std. Short Table	Std. Short Text	Full Policy Text	Layered Text
1-6*	4.16	4.06	4.13	4.00	4.14
7-9*	4.84	4.63	4.47	3.83	4.52

Table 5. Mean enjoyability scores on 7-point Likert scale for single-policy questions (1-6), and comparison questions (7-9). The Likert scale ranged from “Strongly Disagree” (1) to “Strongly Agree” (7). While participants feel neutral with a single policy, the range widens when comparing policies. Rows marked with an asterisk represent statistically significant enjoyability differences between conditions (1-6: $F(4, 756) = 4.25, p < 0.05$; 7-9: $F(4, 756) = 10.65, p < 0.05$).

For the single-policy tasks, participants across the board reported that they felt “confident in my understanding of what I read of Acme’s privacy policy.” The question with the most significant strength in the single-policy tasks was the final question: “If all policies looked just like this I would be more likely to read them,” with the three standardized policies scoring higher than the full policy text.

The three comparison Likert questions show a larger preference for the standardized formats over the full policy text. The questions asked whether comparing two policies was “an enjoyable experience,” was “easy to do,” and if participants “would be more likely to compare privacy policies” if they were presented in the format they saw. The gap between the full policy text and the standardized formats widens from about half a point when looking at a single policy to as much as one and a quarter points after making comparisons.

While the layered text notice performed quite similarly to the full policy text in accuracy measures, we see a very different result in participants’ feelings about using layered notices. The likert scores for layered policies were not significantly different than the standardized-table format (1-6: $t(756) = -1.57, p = 0.115$; and 7-9 $t(756) = -1.48, p = 0.138$).

The comments provided by participants at the end of the study provide insights into their enjoyment. Participants who saw the full policy text described privacy policies as “torture to read and understand” and likened them to “Japanese Stereo Instructions.” On the other hand, participants in the standardized-format conditions were more complimentary: “This layout for privacy policies is MUCH more consumer friendly. I hope this becomes the industry standard.”

DISCUSSION

Our large-scale online study showed that policy formats do have significant impact on users’ ability to both quickly and accurately find information and on users’ attitudes regarding the experience of using privacy policies.

The three standardized formats that were designed by researchers with usability in mind performed significantly better across a variety of measures than the full-text and layered-text policies that currently exist online today. The large amount of text in full-text policies and the necessity to drill down through a layered policy to the full policy to understand specific practices lengthens the amount of time and effort required to understand a policy. Addition-

ally, more complex questions about data practices frequently require reading multiple sections of these text policies and understanding the way different clauses interact, which is not an easy task.

Our earlier work [6] showed that the standardized table performed much better than text policies; however, it was unclear whether the improvement came from the tabular format or the standardization. We have shown here that it is not solely the table-based format, but holistic standardization that leads to success. Our standardized short-text policy left no room for erroneous, wavering, or unclear text, serving as a concise textual alternative to tabular formats.

While the standardized short text policy we developed was successful for most tasks, it is not as easy to scan as a table. Indeed, one participant suggested policies could be improved if they were set up “like a chart so you can scan it visually for answers instead of having to take the time to read it.”

In addition, the standardized short-text format may not scale as gracefully as the standardized tables. The standardized short-text policy did perform significantly worse than the standardized table for some policies. This is evident in the information-collection tasks where users had difficulty finding certain types of information in the short text, especially if it was in the middle of a block of text. Because of the way we generate the text, complex policies are longer than simple policies; however, complexity is often privacy protecting and should not be cognitively penalized. The short-text policy could grow to up to ten paragraphs for complex policies, which is a concern for information finding.

The standardized short policy text did perform well with information that was not collected, used, or shared, even in comparison to the standardized short table with which it shares an identical text notice for this information. We believe that the notice about unused information stood out more in the text policy than the short table. In the text policy this notice was larger than the other text. In the short table the colorful table is more likely to attract users’ attention than the text below it.

One area where the full-text policies did perform as well as the other formats was on user enjoyment of the single-policy tasks. This may be partially attributed to users’ pre-existing familiarity with similar formats. However, this dropped when users reached the comparison tasks, which we expected to be a difficulty with long text policies. From our earlier work, we observed that when asked to compare the enjoyment of reading policies between the standardized-table format and the full policy text, we noted steep improvements in enjoyment of the table format [6]. With this study’s between-subjects design, we were not able to measure such effects, although the free response comments provide some evidence.

Enjoyability results for the layered policies were significantly better than for the full-text policies, even though accuracy scores were not significantly different. Layered policies also took participants less time to use, on average, than full-text policies, although they still took significantly longer than the standardized formats. Some questions could not be answered

correctly from reviewing the layered policy without clicking through to the full policy. However, in this study only 25 of the 79 layered-format condition participants ever clicked through the layered policy to access the full policy. Those who accessed the full policy at least once took an average of 6.6 minutes longer to answer the study questions than those in the layered-format condition who never accessed the full policy. Surprisingly, there were not significant differences in accuracy between layered-format participants who never viewed the full policy and those who did; both groups answered just under half the questions correctly.

The standardized formats performed the best overall, across the variety of the metrics we looked at. The accuracy, comparison, and speed results eclipse the results of the text formats in use today.

The standardized table and standardized short table overall performed very similarly. While there are five cases where the full table outperforms the short table, and only one in the other direction, these differences are frequently small. One concern in the design stage was that removing rows from the table would make comparisons a more cognitively difficult task. This may be evidenced from the significant performance differences in questions 14 and 15; however, the differences in number of rows in the policies we selected were not extreme, never differing by more than one row. It is not clear how great the differences in the types of data collected between real-world policies actually are.

While the accuracy with our standardized formats is better than guessing, there is still room for further study and improvement. Complex information-finding tasks and policy-comparison tasks proved difficult. Future work should continue to concentrate on not just how to present policy information, but also on how to facilitate comparisons. Levy and Hastak recommend continuing to provide better education and context to help consumers make better decisions [8]. While our attached list of definitions is a start, framing the policy with contextual information and presenting comparisons in more useful ways would be productive directions to take future research in usable privacy policies.

ACKNOWLEDGMENTS

This research was supported by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from ARO; Microsoft through the Carnegie Mellon Center for Computational Thinking; and NSF grants CNS-0627513, CNS-0831428, and DGE-0903659. The design team was led by Patrick Gage Kelley and included Joanna Bresee, Aleecia McDonald, Robert Reeder, Sungjoon Steve Won, and Lorrie Cranor. Thanks to Cristian Bravo-Lillo, Robert McGuire, Daniel Rhim, Norman Sadeh, and Janice Tsai. We thank our shepherd, Clare-Marie Karat, for helping us improve this paper.

REFERENCES

1. S. Balasubramanian and C. Cole. Consumers' search and use of nutrition information: The challenge and promise of the nutrition labeling and education act. In *Journal of Marketing*, 2002.
2. L. F. Cranor. *Web Privacy with P3P*. O'Reilly and Associates, Sebastopol, CA, 2002.
3. A. Drichoutis, P. Lazaridis, and R. Nayga. Consumers' use of nutritional labels. In *Academy Marketing Science Review*, 2006.
4. C. Jensen and C. Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478, Vienna, Austria, 2004.
5. P. Kelley, L. Cesca, J. Bresee, and L. Cranor. Standardizing privacy notices: An online study of the nutrition label approach. Technical Report CMU-CyLab-09-014, Carnegie Mellon University, November 2009.
6. P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A "Nutrition Label" for Privacy. In *Proceedings of the 2009 Symposium On Usable Privacy and Security (SOUPS)*, 2009.
7. Kleimann Communication Group Inc. Evolution of a prototype financial privacy notice., February 2006. http://www.ftc.gov/privacy/privacy_initiatives/ftcfinalreport060228.pdf.
8. A. Levy and M. Hastak. Consumer comprehension of financial privacy notices: A report on the results of the quantitative testing, 2008. http://www.ftc.gov/privacy/privacy_initiatives/Levy-Hastak-Report.pdf.
9. A. McDonald and L. Cranor. The cost of reading privacy policies. In *Proceedings of the Technology Policy Research Conference*, September 26–28 2008.
10. A. M. McDonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. A comparative study of online privacy policies and formats. In *Proceedings of 2009 Workshop on Privacy Enhancing Technologies*. ACM, 2009.
11. R. Reeder, L. Cranor, P. Kelley, and A. McDonald. A user study of the expandable grid applied to p3p privacy policy visualization. In *Workshop on Privacy in the Electronic Society*, 2008.
12. The Center for Information Policy Leadership. Multi-Layered Notices Explained, 2004. http://www.hunton.com/files/tbl_s47Details/FileUpload265/1303/CIPLAPEC_Notices.White.Paper.pdf.
13. The Center for Information Policy Leadership. Ten steps to develop a multilayered privacy notice, 2005. http://www.hunton.com/files/tbl_s47Details/FileUpload265/1405/Ten_Steps.whitepaper.pdf.
14. United States Code. 6803. Disclosure of institution privacy policy, 2008. <http://www.ftc.gov/privacy/glbact/glbsub1.htm#6803>.
15. World Wide Web Consortium. The platform for privacy preferences 1.1 (p3p1.1) specification, 2006. <http://www.w3.org/TR/P3P11/>.

User Accuracy Across Policy Formats and Groups

APPENDIX

Simple Tasks	#	Question	Answer	Std.	Short	Std.	Full	Layered
				Table	Table	Text	Policy Text	Text
Complex Tasks	1	Does the policy allow Acme to collect information about which pages you visited on this web site?	Yes	82.35	86.25	91.57	80.23	84.62
	2	Acme might want to use your information to improve their website. Does this policy allow them to use your information to do so?	Yes	79.41	77.50	83.13	82.56	64.10
	3	Does the policy allow Acme to collect information about your current location?	No	48.04	23.75	43.37	18.60	15.38
	4	Based on the policy will Acme register their secure certificate with VeriSign or some other company?	The policy does not say	88.23	81.25	84.34	52.33	30.77
	5	Based on the policy may Acme store cookies on your computer?	Yes	89.22	92.50	73.49	91.86	88.46
	6	Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?	Yes	84.31	76.25	69.88	48.84	33.33
	7	Does the policy allow Acme to share some of your information on public bulletin boards?	Only if I allow them to	75.50	76.25	59.04	15.12	38.46
	8	Does the policy allow Acme to share your home phone number with other companies?	Yes, unless I tell them not to	62.75	68.75	67.47	36.05	14.10
	9	Does the policy allow Acme to use your buying history to design custom functionality targeted at you?	Yes	68.60	60.92	20.43	14.47	65.38
	10	Does the policy allow Acme to share your cookie information with other companies?	Yes	53.92	58.75	53.01	62.79	44.87
	11	Will Acme contact you with advertisements?	No	69.61	67.50	50.60	16.28	44.87
	12	Does Acme give you control regarding their sharing of your personal data?	Yes, unless I tell them not to	79.07	71.26	74.42	26.32	44.87
	13	Does Acme give you control regarding their sharing of	Yes	54.90	61.25	55.42	38.37	39.74
	14	Does Acme give you control regarding their sharing of	Yes	44.19	49.43	51.16	14.47	39.74
	15	Does either company give you options with regards to cookies?	No	70.59	68.75	73.49	66.28	24.36
	16	Does either company collect sensitive information (such as banking or medical records)?	Only with Acme	56.98	44.83	37.21	31.58	24.36
17	Does either company give you options with regards to cookies?	Only with Bell	58.82	52.50	45.78	33.72	42.31	
18	Does either company collect sensitive information (such as banking or medical records)?	Acme	63.95	48.28	19.76	15.79	42.31	
19	By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.	Neither company	64.71	47.50	53.01	20.93	53.85	
20	By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.	True	73.26	63.22	80.23	52.63	53.85	
21	By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.	False, both can	59.80	61.25	34.94	19.77	24.36	
22	By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.	False, both can	56.98	74.71	77.91	46.05	24.36	

Percentage of participants who answered each question correctly, by policy format and viewed policy group. Group A represents participants who saw Policies 1 and 2, Group B, participants who saw Policies 3 and 4. Percentages in bold indicate statistical differences ($p < 0.05$) for formats compared against the standardized table for that policy. For this analysis two separate logistic regressions were performed, a 1x4 for Group A, and 1x5 for Group B. Differences between companies are not compared. Questions are listed exactly as asked, with the corresponding correct answers for each company.