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

Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, Lorrie Faith Cranor

January 12, 2010

CMU-CyLab-09-014

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213
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  

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...
finding, and facilitate policy comparisons. 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 Kleiman 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.

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

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.

Footnote:
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].
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.  

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
search by selecting companies that had websites with over one million views per month\(^3\) 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 (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
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.

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

<table>
<thead>
<tr>
<th>Std.</th>
<th>Std.</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table</td>
<td>Short</td>
<td>Policy</td>
</tr>
<tr>
<td>Participants</td>
<td>188</td>
<td>167</td>
</tr>
</tbody>
</table>

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

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, 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 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.

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.

<table>
<thead>
<tr>
<th></th>
<th>Pol. 1</th>
<th>Pol. 2</th>
<th>Pol. 3</th>
<th>Pol. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Policy Text</td>
<td>2127</td>
<td>6257</td>
<td>4399</td>
<td>2912</td>
</tr>
<tr>
<td>Std. Short Text</td>
<td>175</td>
<td>127</td>
<td>108</td>
<td>90</td>
</tr>
<tr>
<td>Layered Text</td>
<td>409</td>
<td>800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Participant Demographics across conditions

<table>
<thead>
<tr>
<th></th>
<th>number</th>
<th>percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>355</td>
<td>46.5%</td>
</tr>
<tr>
<td>Female</td>
<td>409</td>
<td>53.5%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-28 years old</td>
<td>321</td>
<td>42.0%</td>
</tr>
<tr>
<td>28-40 years old</td>
<td>250</td>
<td>32.7%</td>
</tr>
<tr>
<td>40-55 years old</td>
<td>116</td>
<td>15.2%</td>
</tr>
<tr>
<td>55-70 years old</td>
<td>31</td>
<td>4.1%</td>
</tr>
<tr>
<td>did not disclose</td>
<td>46</td>
<td>6.0%</td>
</tr>
<tr>
<td>Number of Privacy Policies Read in the last 6 months</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never read a privacy policy</td>
<td>189</td>
<td>24.7%</td>
</tr>
<tr>
<td>None in the last six months</td>
<td>130</td>
<td>17.0%</td>
</tr>
<tr>
<td>1 policy</td>
<td>100</td>
<td>13.1%</td>
</tr>
<tr>
<td>2-5 policies</td>
<td>230</td>
<td>30.1%</td>
</tr>
<tr>
<td>5+ policies</td>
<td>101</td>
<td>13.2%</td>
</tr>
<tr>
<td>did not disclose</td>
<td>14</td>
<td>1.8%</td>
</tr>
</tbody>
</table>
timimg 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 varied between conditions since policy content varied between conditions. We present these aggregate results in Figure 4. 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 table 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 tests, see the Appendix.

**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 of 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 stand-
ardized short-text format performed best (81%). Since both
formats presented missing information in an identical man-
ner 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 stan-
dardized 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 standard-
ized 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 par-
ticipant 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 pol-
icy. In the longest of our standardized short-text policies,
the text “Cookie information” is in the middle of a sub-
stantial 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 para-
graphs 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% ac-
curacy, 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 Inter-

**Layered Text**
Layered policies, by design, do not necessarily provide a
complete understanding of a company’s practices. Each com-
pany 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, on average were significantly faster than the full text policy, with the layered format also significantly faster than the full text policy (\( p = 0.025, t(329) = -4.55 \), and standardized short text vs. full text \( t(327) = -6.01 \), were all significant at \( p < 0.05 \)). The layered format was also significantly 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). 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: \( F(4, 756) = 4.26, p < 0.05; 7-9: F(4, 756) = 10.65, p < 0.05 \)). The three comparison Likert questions show a larger preference for the standardized formats over the full policy text. 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). 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: \( F(4, 756) = 4.26, p < 0.05; 7-9: F(4, 756) = 10.65, p < 0.05 \)).

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. Additionally, 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


### User Accuracy Across Policy Formats and Groups (with question information)

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Answer</th>
<th>Std. Table</th>
<th>Std. Short Table</th>
<th>Std. Short Text</th>
<th>Full Policy Text</th>
<th>Layered Text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simple Tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>1</td>
<td>Does the policy allow Acme to collect information about which pages you visited on this web site?</td>
<td>Yes</td>
<td>82.35</td>
<td>86.25</td>
<td>91.57</td>
<td>80.23</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Yes</td>
<td>87.21</td>
<td>85.06</td>
<td>89.53</td>
<td>92.11</td>
</tr>
<tr>
<td>Group A</td>
<td>2</td>
<td>Acme might want to use your information to improve their website. Does this policy allow them to use your information to do so?</td>
<td>Yes</td>
<td>79.41</td>
<td>77.50</td>
<td>83.13</td>
<td>82.56</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Yes</td>
<td>76.74</td>
<td>77.01</td>
<td>86.05</td>
<td>89.47</td>
</tr>
<tr>
<td>Group A</td>
<td>3</td>
<td>Does the policy allow Acme to collect information about your current location?</td>
<td>No</td>
<td>48.04</td>
<td>23.75</td>
<td>43.37</td>
<td>18.60</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>No</td>
<td>46.51</td>
<td>24.14</td>
<td>53.49</td>
<td>3.95</td>
</tr>
<tr>
<td>Group A</td>
<td>4</td>
<td>Based on the policy will Acme register their secure certificate with VeriSign or some other company?</td>
<td>The policy does not say</td>
<td>88.23</td>
<td>81.25</td>
<td>84.34</td>
<td>52.33</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>The policy does not say</td>
<td>79.07</td>
<td>82.76</td>
<td>87.21</td>
<td>43.42</td>
</tr>
<tr>
<td>Group A</td>
<td>5</td>
<td>Based on the policy may Acme store cookies on your computer?</td>
<td>Yes</td>
<td>89.22</td>
<td>92.50</td>
<td>73.49</td>
<td>91.86</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Yes</td>
<td>89.53</td>
<td>80.46</td>
<td>87.21</td>
<td>96.05</td>
</tr>
<tr>
<td>Group A</td>
<td>6</td>
<td>Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?</td>
<td>Yes</td>
<td>84.31</td>
<td>76.25</td>
<td>69.88</td>
<td>48.84</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>No</td>
<td>73.25</td>
<td>58.62</td>
<td>81.40</td>
<td>28.95</td>
</tr>
<tr>
<td><strong>Complex Tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>7</td>
<td>Does the policy allow Acme to share some of your information on public bulletin boards?</td>
<td>Only if I allow them to</td>
<td>75.50</td>
<td>76.25</td>
<td>59.04</td>
<td>15.12</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>No</td>
<td>61.63</td>
<td>57.47</td>
<td>65.12</td>
<td>25.00</td>
</tr>
<tr>
<td>Group A</td>
<td>8</td>
<td>Does the policy allow Acme to share your home phone number with other companies? Yes, unless I tell them not to</td>
<td>Yes</td>
<td>62.75</td>
<td>68.75</td>
<td>67.47</td>
<td>36.05</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Yes</td>
<td>68.60</td>
<td>60.92</td>
<td>20.43</td>
<td>14.47</td>
</tr>
<tr>
<td>Group A</td>
<td>9</td>
<td>Does the policy allow Acme to use your buying history to design custom functionality targeted at you?</td>
<td>Yes</td>
<td>53.92</td>
<td>58.75</td>
<td>53.01</td>
<td>62.79</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Yes</td>
<td>50.00</td>
<td>58.62</td>
<td>69.77</td>
<td>64.47</td>
</tr>
<tr>
<td>Group A</td>
<td>10</td>
<td>Does the policy allow Acme to share your cookie information with other companies?</td>
<td>No</td>
<td>69.61</td>
<td>67.50</td>
<td>50.60</td>
<td>16.28</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>No</td>
<td>79.07</td>
<td>71.26</td>
<td>74.42</td>
<td>26.32</td>
</tr>
<tr>
<td>Group A</td>
<td>11</td>
<td>Will Acme contact you with advertisements? Yes, unless I tell them not to</td>
<td>Yes, unless I tell them not to</td>
<td>54.90</td>
<td>61.25</td>
<td>55.42</td>
<td>38.37</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Yes, unless I tell them not to</td>
<td>44.19</td>
<td>49.43</td>
<td>51.16</td>
<td>14.47</td>
</tr>
<tr>
<td>Group A</td>
<td>12</td>
<td>Does Acme give you control regarding their sharing of your personal data?</td>
<td>Yes</td>
<td>70.59</td>
<td>68.75</td>
<td>73.49</td>
<td>66.28</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>No</td>
<td>56.98</td>
<td>44.83</td>
<td>37.21</td>
<td>31.58</td>
</tr>
<tr>
<td><strong>Comparison Tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>14</td>
<td>Does either company give you options with regards to cookies?</td>
<td>Only with Acme</td>
<td>58.82</td>
<td>52.50</td>
<td>45.78</td>
<td>33.72</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Only with Bell</td>
<td>63.95</td>
<td>48.28</td>
<td>19.76</td>
<td>15.79</td>
</tr>
<tr>
<td>Group A</td>
<td>15</td>
<td>Does either company collect sensitive information (such as banking or medical records)?</td>
<td>Acme</td>
<td>64.71</td>
<td>47.50</td>
<td>53.01</td>
<td>20.93</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>Neither company</td>
<td>73.26</td>
<td>63.22</td>
<td>80.23</td>
<td>52.63</td>
</tr>
<tr>
<td>Group A</td>
<td>16</td>
<td>By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.</td>
<td>True</td>
<td>59.80</td>
<td>61.25</td>
<td>34.94</td>
<td>19.77</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td></td>
<td>False, both can</td>
<td>56.98</td>
<td>74.71</td>
<td>77.91</td>
<td>46.05</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Tasks</th>
<th>Group A</th>
<th>Group B</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Accuracy</td>
<td>90.00</td>
<td>90.00</td>
<td>-1.39</td>
<td>0.166</td>
</tr>
<tr>
<td>Text</td>
<td>90.00</td>
<td>90.00</td>
<td>-2.89</td>
<td>0.004</td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Accuracy</td>
<td>80.00</td>
<td>80.00</td>
<td>-2.43</td>
<td>0.015</td>
</tr>
<tr>
<td>Text</td>
<td>80.00</td>
<td>80.00</td>
<td>-1.41</td>
<td>0.158</td>
</tr>
<tr>
<td>Simple</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Accuracy</td>
<td>85.00</td>
<td>85.00</td>
<td>-2.56</td>
<td>0.010</td>
</tr>
<tr>
<td>Text</td>
<td>85.00</td>
<td>85.00</td>
<td>-2.75</td>
<td>0.006</td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Accuracy</td>
<td>75.00</td>
<td>75.00</td>
<td>-0.58</td>
<td>0.565</td>
</tr>
<tr>
<td>Text</td>
<td>75.00</td>
<td>75.00</td>
<td>-0.58</td>
<td>0.565</td>
</tr>
</tbody>
</table>

User accuracy across policy formats and groups (with detailed statistics)