

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 789 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” [9].

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 com-

monly 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 [8]. It is admittedly unrealistic to expect users to read and understand the privacy policy of every site they visit as people can rarely find information in even a single policy. 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, 9]. 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 [6].

A Privacy “Nutrition Label”

We previously proposed a privacy “nutrition label” to assist consumer understanding of privacy policies [5]. This work was influenced by studies of the design and consumer acceptance of nutrition labeling programs [3, 1]. This tabu-

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.

lar privacy format was designed to enhance user understanding of privacy practices, increase the speed of information 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 [11, 12], 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” [13]. 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” [6]. 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 [6].

In December 2008, the second phase report was published by Levy and Hastak [7]. This report detailed a 1032-participant mail/interview study that tested four privacy notice formats for three fictional financial institutions. 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 two other notices were both heavier in text, with the “current notice” mimicking notices that financial institutions currently use, and the “sample clause” notice generated from GLBA-provided phrases. Levy and Hastak conclude that the KCG table notice performed best. They attribute this performance 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

¹The tabular format can be filled in automatically if a site uses a W3C protocol called the Platform for Privacy Preferences (P3P) [14, 2]. P3P specifies a standardized format for machine-readable privacy policies. Previous work evaluated a much more complicated tabular format based on P3P policies [10].

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.

POLICY FORMATS

We tested five privacy policy formats in this experiment: standardized table (std. table), short standardized table (std. short table), short standardized text (std. 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 the formats. These definitions define the row and column headers in the table conditions and the text tokens in the short natural language condition. 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. Each policy was stripped of all formatting, retaining only internal hyperlinks to reference other areas of the policy, if available in the original. All identifying branding was anonymized, including company and product names, affiliates, and contact information.

Standardized Table

The standardized table format, (Figure 1 on 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 [5], modified based on follow-up design iterations.

Standardized Short Table

The short standardized table (Figure 1 on 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 form, natural language format (Figure 2) by translating each row in the std. 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 the tradeoffs between textual and tabular formats.

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 & govt 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

Legend:
opt out by default, we will collect and use your information in this way unless you tell us not to by opting out
opt in 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 short standardized 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.

Layered Text

The final policy format we tested is the layered privacy notice [11, 12]. This format involves a summarized, one-screen privacy policy in a tabular format that links to the full natural language policy. Layered policies have already been deployed by major corporations, making them a viable, real world summary form of privacy policies. These policies were stripped of identifiable 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 Surveyor’s Point. Mechanical Turk offers workers across the world the ability to perform short tasks and get compensated. People can place jobs through Mechanical Turk, specifying the number of people they are looking for, qualifications those people must have, the amount they are willing to pay, and details about the task. Mechanical Turk payments must be calibrated for the length of the task. For our approximately fifteen-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 without needing to open up multiple browser tabs or windows. This also allowed us to track the number of users who looked at each policy and the number of times they switched between them. Additionally, we

wanted to be able to instrument the policies we were testing to understand the way users interacted with them. Not only did we collect the amount of time that users spent reading the policies, we also collected 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 a few 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. Each of our 789 participants was shown two privacy policies and asked to answer 26 questions. We designed the experiment to use a between-subjects design, with each participant assigned one of two policy pairs, in one of five different formats. The between-subjects design was chosen to remove learning effects and to be completed within about 15 minute. All participants answered the same questions; only the policy formatting and content differed.

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

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

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

Conditions

We selected the privacy policies of four well-known companies and compared them in two pairings across five policy formats (participants per condition can be seen in Table 1). We refer to anonymized versions of the Target and Disney privacy policies as Group A and to anonymized versions of the Microsoft and IBM privacy policies as Group B. Participants were assigned to either Group A or Group B, and one of the five formats. Participants in Group A, saw Target’s policy first, which we represented as the fictitious company “Acme,” and for tasks that involved comparisons they saw Disney’s policy which we referred to as the “Bell” policy. Participants in Group B saw Microsoft’s privacy policy as “Acme,” and IBM’s as “Bell.” As layered text policies are not widely used, only Policy Group B had a layered text format option. We used two pairs of policies rather than just one so that we could test a wider variety of real policies (thus the A and B groups). We discuss the differences we observed in how participants interacted with these two pairs of policies in our analysis.

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

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 including: gender, age, and current occupation.
2. *Internet & Privacy* We asked the participants four questions to better understand their internet usage and their prior knowledge of privacy.

	Std. Table	Std. Short Table	Std. Short Text	Full Policy Text	Layered Text
Group A	105	84	88	88	
Group B	90	88	90	77	79

Table 1. 789 study participants, spread across our nine conditions

	Group A		Group B	
	Acme	Bell	Acme	Bell
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.

3. *Simple Tasks* Participants were shown 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* Participants were asked six questions (again only 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* Participants were shown 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. They were asked five comparison questions that required looking at both policies.
7. *Policy Comparison Likeability* Participants were asked three more Likert questions to collect qualitative feedback on the task of comparing two policies.

Analysis

Our analysis, detailed below, is split into three portions.

1. We scored Simple, Complex, and Comparison tasks for accuracy. We marked all questions as correct or incorrect (although we will later discuss varying degrees of incorrectness). We performed factorial logistic regressions across the policy formats.
2. We performed an ANOVA analysis on the log normalized timing information for the above tasks.

	Group A	Group B
<i>Total Participants</i>	351	413
<i>Gender</i>		
Male	45.2%	47.5%
Female	54.7%	52.5%
<i>Age</i>		
18-28 years old	44.7%	39.7%
28-40 years old	30.5%	34.6%
40-55 years old	15.1%	15.3%
55-70 years old	4.8%	3.4%
did not disclose	4.9%	7.0%
<i>Number of Privacy Policies Read in the last 6 months</i>		
Never read a privacy policy	23.3%	25.9%
None in the last six months	16.2%	17.7%
1 policy	12.3%	13.8%
2-5 policies	31.9%	28.6%
5+ policies	14.0%	12.3%

Table 3. Participant Demographics across conditions

3. We also performed an ANOVA analysis for the nine 7-point Likert questions.

We excluded participant data from analysis if they did not complete the entire question set. In addition, data from participants who completed the study in less than two standard deviations from the mean of the log-normalized³ times were excluded. (Group A: $n = 14$, Group B: $n = 11$)⁴

The data from the remaining 764 participants will be discussed for the rest of the paper. Table 3 shows the gender and age breakdown of the participants by group, 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 1 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 may have sought to seem more knowledgeable about privacy policies. Nutritional and drug labeling literature reports that

³Log-normalization is used on analysis of timing information for the remainder of the paper to force a normalized distribution, allowing us to perform ANOVA analysis. Timing information in charts will be displayed in seconds to assist understanding.

⁴Mean A: 847 seconds or 14.1 minutes, B: 806 seconds or 13.4 minutes. Cutoff point (2 standard deviations below the mean) A: 268 seconds, B: 262 seconds.

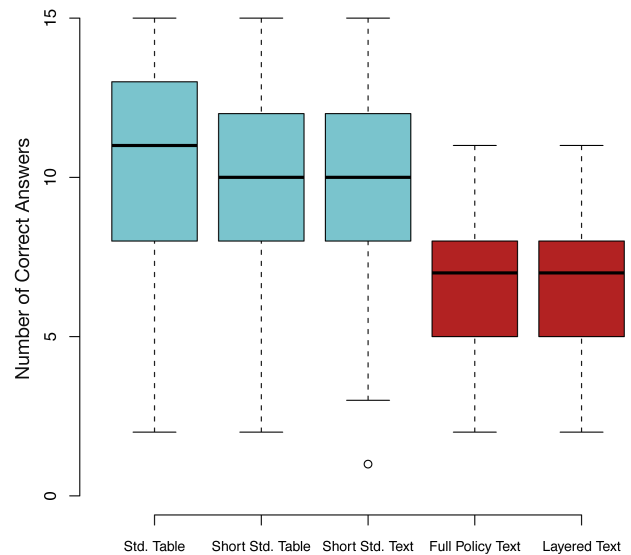


Figure 3. Accuracy results for each of the five policy formats (combined results for conditions A and B).

standardization efforts assist most those who seek out the information. If participants on Mechanical Turk do read more privacy policies than the general population then we may in fact be refining our label to help the group that will be most likely to leverage the information.

RESULTS

We describe our big-picture accuracy results, followed by a more in-depth analysis of several specific tasks, summarize our timing results, and conclude with an analysis of participants' enjoyment in 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 by conditions since the policy content varied across conditions. We present those aggregate results in Figure 3. This summary shows a very clear divide, with the three standardized formats scoring 62-69%, in light blue; while the two real-world text policies are 43-46%, in red. In both policy groups, the standardized table significantly outperformed both of the real-world text policies (standard linear regression, $p < 0.05$). In Policy Group B, the standardized formats did not perform statistically differently, while in Policy Group A the standardized table did significantly outperform the standardized short text policy ($p = 0.03$).

Accuracy Results

The complete accuracy results are presented in Table 4. For analysis on a per question basis, we performed factorial logistic regressions with the standardized table as the base for comparison across formats. We have shown significant differences ($p < 0.05$) in formats with boldface accuracy scores. We did not compare performance quantitatively between companies due to differences in practices and answers. However,

we do discuss a number of observations about differences between policies across conditions and their impact on performance. We highlight several of the questions, below.

Information that is out of scope

Question 4 asked: “Based on the policy will Acme register their secure certificate with VeriSign or some other company?” As shown in Table 4, 79-88% of participants in the three standardized conditions were able to correctly answer, “The policy does not say.” However accuracy dropped to 31-52% for the full policy text and layered text conditions. Neither of the policies mentioned Verisign or any other certificate registrar, nor did either policy contain the word “certificate.” While this was clear in the three standardized formats, participants with the full policy text format had to scan for the absence of information over several pages of text.

Information that is collected

Question 5 asked: “Based on the policy may Acme store cookies on your computer?” The answer for both policies was “Yes.” While this question was straightforward for most conditions, the standardized short text format did not perform as well in group A, with only 73% of participants answering correctly (compared 80-96% across all other conditions). “Cookie information” is in the middle of a more substantial block of text in Policy Group A’s std. short table which is significantly longer, than Policy Group B’s. However, at only 175 words participants may not use the search or find functionality of their browser, thus missing the word cookie. This is speculative; however, and a study with the paragraphs rearranged may lead us to better understand if any blind spots exist in this format.

Sensitive information collection

Question 6 asked: “Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?” For Policy A, the answer was Yes, and for Policy B, No. The full text policy, again, performed badly, especially for Policy B. Here, 29% of participants correctly answered that Acme did not collect their medical information; however, 41% answered the policy does not say, which we marked as incorrect because the absence of this information means that they cannot collect health-related information. One of the benefits of a standardized form is an empty row in a table, or a required text notice that lists information that is not collected. For Policy A, only 49% of the participants correctly answered that they do collect medical information. The policy itself references “counseling from pharmacists,” an “online prescription refill service,” and “prescription medications.” The standardized short table format performed poorly (59%) when medical information was absent; however, the standardized short text format performed best (81%) when medical information was absent, though both had identical notices describing this absence. This is probably due to the standardized short text format reserving the largest font size for things a company does not do, including in this case, collecting medical information. While this is in the same font-size for the short table, the table itself overpowers the notice.

Understanding if any information is shared publicly

Moving onto complex tasks, Question 7 asked: “Does the policy allow Acme to share some of your information on public bulletin boards?” For Policy Group A, the answer was “Only if I allow them to,” which translates to an opt in, while for Policy Group B, the answer was simply “No.” In the 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%. In both policy groups, incorrect answers across the standardized formats were evenly distributed, with no clear incorrect answer trends. Participants given the full policy text format have strikingly low results for this question, regardless of the policy they were assigned. For Policy A, the largest contingent, 32% of participants, (incorrectly) reported the policy did not specify whether information would be shared on public bulletin boards.

Sharing of a specific piece of personal information

Question 8 asked: “Does the policy allow Acme to share your home phone number with other companies?” For Policy Group A, the answer was “Yes, unless I tell them not to,” while for Policy B, the answer was “Yes.” To answer this question correctly in all conditions, the participants needed to realize that a home phone number was considered “contact information” in the standardized conditions or that it fell under a broad umbrella of “personal information” in both full text policies. Looking at the std. short text format for Policy Group B, 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 believe that this again comes from misreading the text in the std. short text. An option existed in that paragraph, but only for telemarketing, not sharing. The remaining five standardized format conditions had accuracy ranging from 63-69% with incorrect answers split evenly across answers.

Comparison Results

The final five task-based questions (13-17) called for participants to answer questions that were based on two different policies. The first policy was the Acme policy, which they had looked at for the prior 12 questions. The second policy was representing a company we called “The Bell Group.” The first and final questions, 13 and 17, were the same, asking participants to select which of the two companies they would prefer to make an online purchase from, while questions 14-16 asked for specific information comparisons. For each of these questions, the participants started with only the question on the screen, and were presented with a “policy switcher” that allowed them to view either policy. The Acme policy, which participants had already answered twelve questions about, was infrequently viewed. Less than one-third of our participants reviewed the Acme policy, while nearly all participants reviewed the Bell policy for each of the comparison questions.

Checking for any options regarding a data category

The first comparison task, Question 14 asked: “Does either company give you options with regards to cookies?” For Policy Group A, the Acme policy does provide an opt out

Percentage Accuracy By Question, By Condition

#	Question	Answer	Std. Table	Std. Short Table	Std. Short Text	Full Policy Text	Layered Text
<i>Group A</i>	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	
<i>Group B</i>		Yes	87.21	85.06	89.53	92.11	84.62
<i>Group A</i>	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	
<i>Group B</i>		Yes	76.74	77.01	86.05	89.47	64.10
<i>Group A</i>	3 Does the policy allow Acme to collect information about your current location?	No	48.04	23.75	43.37	18.60	
<i>Group B</i>		No	46.51	24.14	53.49	3.95	15.38
<i>Group A</i>	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	
<i>Group B</i>		The policy does not say	79.07	82.76	87.21	43.42	30.77
<i>Group A</i>	5 Based on the policy may Acme store cookies on your computer?	Yes	89.22	92.50	73.49	91.86	
<i>Group B</i>		Yes	89.53	80.46	87.21	96.05	88.46
<i>Group A</i>	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	
<i>Group B</i>		No	73.25	58.62	81.40	28.95	33.33
<i>Group A</i>	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	
<i>Group B</i>		No	61.63	57.47	65.12	25.00	38.46
<i>Group A</i>	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	
<i>Group B</i>		Yes	68.60	60.92	20.43	14.47	14.10
<i>Group A</i>	9 Does the policy allow Acme to use your buying history to design custom functionality targeted at you?	Yes	53.92	58.75	53.01	62.79	
<i>Group B</i>		Yes	50.00	58.62	69.77	64.47	65.38
<i>Group A</i>	10 Does the policy allow Acme to share your cookie information with other companies?	No	69.61	67.50	50.60	16.28	
<i>Group B</i>		No	79.07	71.26	74.42	26.32	44.87
<i>Group A</i>	11 Will Acme contact you with advertisements?	Yes, unless I tell them not to	54.90	61.25	55.42	38.37	
<i>Group B</i>		Yes, unless I tell them not to	44.19	49.43	51.16	14.47	39.74
<i>Group A</i>	12 Does Acme give you control regarding their sharing of your personal data?	Yes	70.59	68.75	73.49	66.28	
<i>Group B</i>		No	56.98	44.83	37.21	31.58	24.36
<i>Group A</i>	14 Does either company give you options with regards to cookies?	Only with Acme	58.82	52.50	45.78	33.72	
<i>Group B</i>		Only with Bell	63.95	48.28	19.76	15.79	42.31
<i>Group A</i>	15 Does either company collect sensitive information (such as banking or medical records)?	Acme	64.71	47.50	53.01	20.93	
<i>Group B</i>		Neither company	73.26	63.22	80.23	52.63	53.85
<i>Group A</i>	16 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	59.80	61.25	34.94	19.77	
<i>Group B</i>		False, both can	56.98	74.71	77.91	46.05	24.36

Table 4. Percentage of participants who answered each question correctly, by policy format and company group. 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.

regarding cookies while Bell does not. For Policy Group B, the Acme policy does not provide any options regarding cookies while Bell does. Focusing on Policy Group B, we note that the standardized short text received only 20% accuracy (putting it on par with the full policy text at 16% accuracy), with 49% of respondents incorrectly answering that neither company gave options with regards to cookies. For this format, a participant must have understood that the first paragraph applied to cookies, and then noticed the ability to opt out of either use or sharing practices in the last two lines. For the full policy text, 55% of the participants believed that both companies gave options regarding cookies. This means that they incorrectly answered that the Acme policy had options regarding cookies. 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 a function of most web browsers. The next line even states “If you choose to decline cookies, you may not be able to sign in...,” making it obvious that Acme sites will use cookie information. Policy Group A had a smaller range across conditions, 33% for the full policy text, the standardized formats ranging from 46-59%.

Checking for multiple data categories

Question 15 asked: “Does either company collect sensitive information (such as banking or medical records)?” For Policy Group A, the Acme policy does collect health information, whereas Bell does not. For Policy Group B, neither company collects sensitive information. Policy Group A performed, on average, worse than Group B, which we expected as the concept that neither company collects sensitive information is easier to understand and also more expected. Specifically, the full policy text comparison had only 21% accuracy for Group A, but also again performed worst across all conditions. Note that the standardized short text policy did well when the two companies’ practices were aligned but accuracy dropped nearly 30% when the policies had different practices, a larger drop than that found in the other standardized policies. This question relates back to question #6 which asked participants if the Acme policy collected medical or health information. Participants who correctly answered that question should have been more likely to answer this question correctly. For Policy Group A, 84% of the people who answered #15 correctly had already correctly determined that the Acme company collected medical information in #6. However, 44% of those who answered #6 correctly did not come to the correct conclusion for #15. For Policy Group B, 72% of the people who answered #15 correctly had already correctly determined that the Acme company did not collect medical information in #6. However, 18% of those who answered #6 correctly did not come to the correct conclusion for #15.

Questions 13 & 17: Making a purchase decision

We asked the participants “Which company (assuming prices and products are similar) would you rather make a purchase

from?” once before the above information-finding questions, and again afterwards, prefaced by: “After reviewing the policies in more detail, which company...” The results for both questions are nearly identical. Bell was preferred by Policy Group B by between 51-62% of the participants across all conditions. Policy Group A results were not consistent. We see that the std. short text format favors the Bell policy at 78%, while the full policy text format is the only condition where the participants favored the Acme policy, at 37%.

Timing Results

We examined completion times for the simple, complex, and comparison tasks, as presented in Table 5. Statistical significance was tested 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 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 about comparing policies. The results are summarized in Table 6. 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.

For the single-policy tasks, participants across the board reported that they were 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 much larger shift towards the standardized formats and away from the full policy text. The questions we asked in this section were: if 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 (t-test, 1-6, $p = 0.215$ and 7-9 $p = 0.478$).

The comments provided by participants at the end of the study also provide insights into their enjoyment. Partici-

Average Timing Information (in seconds)

	#	Std. Table		Std. Short Table		Std. Short Text		Full Policy Text		Layered Text	
		avg.	σ	avg.	σ	avg.	σ	avg.	σ	avg.	σ
Group A	1-6	233	150	216	106	248	175	391	260		
Group B	1-6	239	254	204	99	218	170	338	230	317	403
Group A	7-12	183	211	144	72	176	145	203	171		
Group B	7-12	168	172	128	72	151	92	310	483	186	208
Group A	13-17	174	141	168	87	178	128	244	270		
Group B	13-17	140	99	129	102	160	114	228	163	187	156
Group A	All	952	634	886	364	1012	552	1281	835		
Group B	All	865	480	821	439	867	462	1252	774	1089	763

Table 5. 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.

pants in the full policy text condition 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 and standardization in mind performed significantly better than the full and layered text policies that currently exist online today. These two policy formats, across the variety of measures we tested, performed consistently worse. The large amount of text in full text policies and the necessity to drill down through a layered policy to the entire 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 [5] 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 for information. Indeed, one participant suggested that 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.”

Question Number		Std. Table	Std. Short Table	Std. Short Text	Full Policy Text	Layered Text
1-6	Group A	4.52	4.46	4.52	3.95	
7-9*	Group A	5.12	5.03	4.66	3.88	
1-6*	Group B	4.44	4.42	4.13	3.63	4.12
7-9*	Group B	4.84	4.69	4.44	3.91	4.65

Table 6. Mean enjoyability scores on 7-point Likert scale for single-policy questions (1-6), and comparison questions (7-9). 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 ($p < 0.05$, ANOVA).

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 in one of the two policy groups. 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 short standardized text policy did perform well with information that was not collected, used, or shared, even in comparison to the short standardized table with which it shares an identical text notice for this information. We believe that this can be attributed to the larger type size than the short text policy itself, while underneath the colorful and larger short standardized table, the notice is not as visible.

One area where the full text policies did perform as well as the other formats was on user enjoyment after the single-policy tasks in one of the two policy groups. 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 [5]. 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 access the full policy; 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, and they perform similarly on the remaining 80% of tasks. 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 simple 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 [7]. 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, Lucian Cesca, Robert McGuire, Daniel Rhim, Norman Sadeh, and Janice Tsai.

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. 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.
6. Kleimann Communication Group Inc. Evolution of a prototype financial privacy notice., February 2006.
7. A. Levy and M. Hastak. Consumer comprehension of financial privacy notices: A report on the results of the quantitative testing, 2008.
8. A. McDonald and L. Cranor. The cost of reading privacy policies. In *Proceedings of the Technology Policy Research Conference*, September 26–28 2008.
9. 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.
10. 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.
11. The Center for Information Policy Leadership. Multi-Layered Notices Explained, 2004.
12. The Center for Information Policy Leadership. Ten steps to develop a multilayered privacy notice, 2005.
13. United States Code. 6803. Disclosure of institution privacy policy, 2008.
14. World Wide Web Consortium. The platform for privacy preferences 1.1 (P3P1.1) specification, 2006.