A Comparative Study of Online Privacy Policies and Formats

Aleecia M. McDonald,[1] Robert W. Reeder,[2] Patrick Gage Kelley,[1]
Lorrie Faith Cranor[1]

[1] Carnegie Mellon, Pittsburgh, PA
[2] Microsoft, Redmond, WA

Abstract. Online privacy policies are difficult to understand. Most privacy policies require a college reading level and an ability to decode legalistic, confusing, or jargon-laden phrases. Privacy researchers and industry groups have devised several standardized privacy policy formats to address these issues and help people compare policies. We evaluated three formats in this paper: layered policies, which present a short form with standardized components in addition to a full policy; the Privacy Finder privacy report, which standardizes the text descriptions of privacy practices in a brief bulleted format; and conventional non-standardized human-readable policies. We contrasted six companies' policies, deliberately selected to span the range from unusually readable to challenging. Based on the results of our online study of 749 Internet users, we found participants were not able to reliably understand companies' privacy practices with any of the formats. Compared to natural language, participants were faster with standardized formats but at the expense of accuracy for layered policies. Privacy Finder formats supported accuracy more than natural language for harder questions. Improved readability scores did not translate to improved performance. All formats and policies were similarly disliked. We discuss our findings as well as public policy implications.

* Funded by NSF Cyber Trust grant CNS-0627513, Microsoft through the Carnegie Mellon Center for Computational Thinking, Army Research Office grant number DAAD19-02-1-0389 to Carnegie Mellon CyLab, and FCT through the CMU/Portugal Information and Communication Technologies Institute. Thanks to Robert McGuire and Keisha How for programming assistance.

1 Introduction

The United States relies on a self-regulation approach to Internet privacy. There are some Internet privacy laws, for example the Children's Online Privacy Protection Act of 1998 (COPPA), which protects children's privacy [6], and the Gramm-Leach-Bliley Act (GLB), which applies to financial data [11]. But by and large the theory of Internet privacy hinges on two assumptions:

- Consumers will choose companies with acceptable privacy policies.
- Companies will not violate their privacy policies because the Federal Trade Commission (FTC) can bring action for unfair and deceptive practices.

In both cases privacy policies play a vital role in Internet privacy. Self-reports show three quarters of Internet users take active measures to protect their privacy, ranging from installing privacy protective technology to providing false information to web sites [1]. Yet only 26% read privacy policies during a recent study and readership outside of laboratory conditions is believed to be far lower [14]. To study the effectiveness of various approaches to improving the readability of privacy policies, we investigated the performance of three different formats for privacy policies and compared policies from six different companies.

In section two we describe related work and the formats we contrasted. We describe our methods in section three. We present accuracy and time to answer results in section four, and psychological acceptability results in section five. We discuss implications from these results and conclude in section six.

2 Related Work

Several studies frame willingness to read privacy policies as an economic proposition and conclude that asymmetric information is one reason why people find it not worth their time to read privacy policies [28,1]. Other studies show that privacy policies and financial disclosures require a college reading level to understand [12,24,10,2]. A study of ambiguities in privacy policies shows they contain language that downplays privacy issues [20]. The 2006 Kleimann report on GLB financial privacy notices found that subheadings and standard formats dramatically improved readability [22]. In response to these issues, privacy researchers and industry groups devised several standardized formats for privacy policies based on the expectation that standardized formats would improve comprehension. Our study is a comparative analysis to analyze how well standardized policies work in practice.

While not in the realm of privacy policies, Kay and Terry's research on open source license agreements includes testing multiple formats. Early work found modest improvements in likelihood to read well designed agreements but no improvement in retention of the material [15]. Tsai found when study participants searched for products to purchase and saw a single icon view that evaluated the privacy practices for each site, they were willing to pay a small premium for more privacy-protective sites [27,8]. On the other hand, translating an entire privacy

policy into a grid that conveyed information by icons and colors did not improve comprehension [21]. Attempts at visualizing privacy are ongoing, including a set of icons modeled after Creative Commons [3]. This study, in contrast, examines three text-based formats as described below.

2.1 Privacy Finder

Privacy Finder (PF) is a privacy-enhanced front end to Yahoo! and Google search that was developed by AT&T and refined at the Cylab Usable Privacy and Security (CUPS) Laboratory. Privacy Finder includes a privacy report that displays standardized text generated automatically from Platform for Privacy Preferences (P3P) policies. P3P is a standardized format for privacy policies, and is formally recommended by the World Wide Web Consortium (W3C) [29]. P3P policies are encoded in XML (eXtended Markup Language), which is computer readable and thus allows software tools to help people manage their privacy preferences.

Because Privacy Finder generates text from P3P tags, the Privacy Finder report avoids emotionally charged language and ensures uniform presentation. However, Privacy Finder reports allow a free-form text description of the highest level of policy statements. This can improve readability by providing context for readers, but also means that companies with identical practices may have different Privacy Finder reports.

2.2 Layered Notices

The law firm Hunton & Williams popularized the notion of layered notices [25] which include a short one-screen overview with standardized headings which then links to the full natural language policy. Although the headings for the first layer are standardized the text within each section is free form.

By 2005, several large companies deployed layered policies including Microsoft (MSN), Procter & Gamble, IBM, and JP Morgan [17]. European Union Information Commissioner Richard Thomas called for the use of layered policies in response to research showing nearly 75% of participants said they would read privacy policies if they were better designed [19]. Article 29 of European Union Directive created the "Working Party on the Protection of Individuals with regard to the processing of Personal Data," which issued guidance on how to create layered policies [4]. Privacy commissioners in EU countries supported layered policies. In Australia, the Privacy Commissioner released a layered policy for their own office, intending it "as a model for other agencies and organisations" [26].

2.3 Natural language

Most privacy policies are in natural language format: companies explain their practices in prose. One noted disadvantage to current natural language policies is that companies can choose which information to present, which does not

necessarily solve the problem of information asymmetry between companies and consumers. Further, companies use what have been termed "weasel words" — legalistic, ambiguous, or slanted phrases — to describe their practices [20]. Natural language policies are often long and require college-level reading skills. Furthermore, there are no standards for which information is disclosed, no standard place to find particular information, and data practices are not described using consistent language.

3 Methods

We conducted an online study from August to December 2008 in which we presented a privacy policy to participants and asked them to answer questions about it. We posted advertisements on craigslist and used personal networks to recruit participants. We offered a lottery for a chance to win one of several $75 Amazon gift certificates as incentive for participating in the study.

We used a between subjects design and assigned each participant to one of 15 privacy policy representations. We used a between subjects design rather than within group design because in this context it is unrealistic to eliminate learning effects simply by reordering policies. Reading the questions could affect how participants read subsequent policies. It is also unrealistic to expect participants to spend more than 20 minutes completing an online survey. Questions remained constant over all conditions; only the policy differed.

3.1 Study Conditions

We contrasted six different companies' conventional natural language (NL) policies and their corresponding Privacy Finder privacy report format (PF) plus three layered policies. We refer to these companies as A through F. We analyzed 749 participants across 15 conditions, for an average of 50 participants per condition. The study conditions are listed in Table 1.

Table 1. Participants per Condition

Company	Designation	NL	PF	Layered
Disney	A	41	50	N/A
Microsoft	B	47	46	52
Nextag	C	46	41	N/A
IBM	D	47	47	49
Walmart	E	52	51	N/A
O'Reilly	F	62	55	63

We replaced all companies' names with "Acme" to avoid bias from brand effects. For natural language polices we used black text on white backgrounds

regardless of the original graphic design. We left other formatting that might aide comprehension (for example, bulleted lists) intact.

Note that we did not study layered policies for companies A, C, and E. Of the six companies, only B and D had layered policies. We followed the directions from the Center for Information Policy Leadership [5] to create a third layered policy for company F as part of a prior study [21] and used it here to facilitate comparisons between studies.

As deployed in practice, Privacy Finder highlights the most important information at the top of the report and provides links to expand details. We discovered in earlier testing that people rarely expanded the Privacy Finder report. We were interested in testing how well people are able to use the information in the Privacy Finder report, not how well they are able to navigate the user interface, so in our research we presented all information in a single flat file.

We selected privacy policies from six popular websites that engage in e-commerce, and thus must collect a variety of personal information as part of their business. We chose what we believe to be a comparatively easy to read and a comparatively difficult to read policy with several typical policies. We selected policies guided by several measurements of readability summarized in Table 2. For each company, we noted the length of the natural language policy. We calculated the Flesch-Kincaid Reading Ease Score, which ranges from a low of 1 to a high of 100 based on syllable count and line lengths. High Flesch-Kincaid scores are more readable than low scores. In general, experts suggest a score of at least 60—70, which is considered easily understandable by 8th and 9th graders [18]. *Reader's Digest* has a readability index in the mid 60s, *Time* is in the low 50s, and *Harvard Law Review* in the low 30s [13]. Note that while the policies we selected span a range from 32 to 46, even the most readable policy is more challenging than is normally recommended for a general audience.

We calculated the percentage of sentences written in the passive voice, which is both more difficult for readers to understand and an indicator the company may not be comfortable taking full responsibility for their privacy practices. We counted the number of cross references within each policy; the more times readers are asked to refer to other parts of the document the more difficult it is to understand. Finally, we note that the standardized Privacy Finder format also has a range of lengths due to differing numbers of statements, how much information they collect, and how much text the policy authors elected to supply.

3.2 Study Questions

Study questions comprised several groups:

- *Comprehension.* Participants answered a series of multiple choice questions to determine how well they were able to understand the policy. These questions are realistic information retrieval tasks based on typical privacy concerns, and are similar to questions used in an earlier study by Cranor et al [7]. In the study, we conducted three rounds of pilot tests with over two dozen people to ensure the questions were well-worded and understandable.

Table 2. Attributes of six companies' privacy policies

Co.	NL Words	NL Pages	Flesch	% Passive	Cross ref.s	PF Words
A	6329	13	31.8	11%	27	880
B	3725	7	35.5	22%	0	1964
C	2920	6	36.3	17%	7	2011
D	2586	8	42.8	18%	2	554
E	2550	8	44.9	11%	0	1373
F	928	3	46.3	9%	1	1843

We randomized the order of these questions to mitigate learning effects and captured both accuracy and time to respond. We also included a warm-up task which we did not score.

- *Psychological Acceptability.* Saltzer and Schroeder coined the term psychological acceptability to convey that if people do not like a system they will not use it. They wrote, "It is essential that the human interface be designed for ease of use, so that users routinely and automatically apply the protection mechanisms correctly." [23] Participants answered subjective questions on a seven-point Likert scale.
- *Demographics.* We collected basic information like gender, educational attainment, and income so we could understand how closely our study population resembles Internet users as a whole.

We also measured the time it took for participants to answer each one of the comprehension questions. When not engaged in a research study, few people even skim privacy policies let alone read them to find answers to their concerns [15]. The times we measured do not reflect normal practices, but they do allow us to compare performance between formats, which is our goal.

3.3 Research Questions

Standardized formats were designed with care to help readers make sense of online privacy policies. With all of the resources invested in standardized policies we expected they would help people understand privacy policies. We held multiple hypotheses:

- Participants will have (a) higher accuracy scores, (b) shorter times to answer, and (c) greater psychological acceptability with both of the standardized formats than with their natural language counterparts.
- Participants will have (a) higher accuracy scores, (b) shorter times to answer, and (c) greater psychological acceptability with highly readable natural language than they will on natural language policies with low readability metrics.

Understanding these issues contributes to determining the most effective ways to present policies to end users. This is particularly relevant given Gramm-

Leach-Bliley regulations on paper-based financial privacy policies; similar legislation could apply to online privacy policies in the future. The FTC's most recent report on behavioral advertising was described by the FTC Chairman Leibowitz as the last chance to make industry self-regulation work [9]. If we move away from industry self-regulated content, what should we do instead? Do any of the standardized approaches help enough to warrant considering regulation of policy formats?

3.4 Analysis

We performed a comparative analysis across all three formats (Natural Language, Privacy Finder, and Layered) and from all six companies to see if there were statistically significant differences in the mean scores for accuracy, time to completion, and psychological acceptability questions.

After we removed outliers[3] we performed ANOVA analysis for both time data and psychological acceptability, which we recorded on a seven point Likert scale and treated as continuous variables. We performed all tests of statistical significance at the $\alpha = 5\%$ confidence level. For the sake of readability, all details of statistical significance tests are in the Appendix.

4 Accuracy and Speed Results

Accuracy scores are all reported as the percentage of people who answered the question correctly.[4] As compared to natural language, we found that layered policies led to lower accuracy scores for topics not in the short layer. Privacy Finder

[3] We only included results from participants who completed all of the accuracy questions. Because this was an online study to enter a drawing for a gift certificate, a few people just "clicked through" answers without engaging with the material. We picked a fixed lower threshold of 1.5 seconds per question and removed participants entirely if they had two or more questions they answered in under 1.5 seconds (7 participants removed out of an original 756 for a total of 749.) For participants with only one time under 1.5 seconds, it is possible they accidently double-clicked once but answered other questions properly. We removed the time and accuracy data for just the affected question (3 question/time pairs out of 3000.) At the other extreme, sometimes people were diverted by other tasks while answering questions and we recorded unduly long times to answer. We discarded question times in excess of 2.5 times the mean for their condition along with their corresponding answers. This resulted in $N = 723$ for cookies, 728 for opt out, 726 for share email, and 723 for the telemarketing questions.

[4] Interpreting results is complicated by potential confusion of how participants answered when answers are inferred. For example, we asked about opt out practices for policies where there is no opt out link. The straight-forward answer we envisioned is "No." However, participants may also have replied that the policy "Does Not Say," intending to convey the same information since there is no opt out link within the policy. Arguably, in that case the correct way to score responses is to combine the correct answer with "Does Not Say." We analyzed the combined percentage for each question and found in all but one case there was no difference in

was indistinguishable from natural language until questions became harder, at which point Privacy Finder was slightly superior to natural language.

Accuracy spanned a wide range. An average of 91% of participants answered correctly when asked about cookies, 61% answered correctly about opt out links, 60% understood when their email address would be "shared" with a third party, and only 46% answered correctly regarding telemarketing. With only three possible answers, if participants guessed randomly we would expect 33% accuracy.

All other things being equal, lower times are better because they reflect participants were better able to comprehend the policy. Participants answered more quickly with both layered and Privacy Finder formats. Times to answer increased with question difficulty, with an average of 2.3 minutes to answer the question about cookies, 4.7 minutes to answer about opt out links, 5.3 minutes for email sharing, and 6.7 minutes for telemarketing.

4.1 Cookies

We asked: Does the Acme website use cookies?

Answer: Yes for all policies.

Most participants got the cookie question right (91%). This was an easy question to answer because our question is phrased with the same term the policies use. All policies, in all formats, call out cookies use explicitly. For example, one policy has a heading of "Cookies and Other Computer Information" with a paragraph that begins: "When you visit Acme.com, you will be assigned a permanent 'cookie' (a small text file) to be stored on your computer's hard drive." There is no ambiguity. Even someone who has no idea what a cookie is, or what the implications for privacy are, can skim through any of the natural language policies to find the word "cookie" and answer correctly.

Table 3. Percentage correct and minutes to answer, cookies question.

Policy	% correct	Time
A NL	87%	3.6
A PF	96%	1.5
B NL	96%	2.0
B PF	98%	1.6
B Layered	86%	2.3
C NL	93%	2.4
C PF	98%	3.5
D NL	86%	2.6
D PF	91%	1.9
D Layered	69%	2.2
E NL	96%	2.6
E PF	96%	1.8
F NL	100%	2.3
F PF	94%	2.7
F Layered	80%	2.3

We found significant differences in accuracy for company and format. The six companies have a relatively small span between the worst performance (D, 82%) and best performance (E, 96%.) See Table 3 for a summary of results.

Layered policies gave participants a little more trouble (78%) than other formats. Cookie information was under the heading "Personal Information" in F Layered (80%,) which may not be where people expected to look. In D Layered (69%,) the policy mentions in passing that "You may also turn off cookies in your browser,"

the threshold for statistical significance. Further, the relative ranking of formats and companies remained stable.

without explicitly saying they use cookies. People must deduce that information or go to the full policy for a direct statement that the site uses cookies. This highlights two results we will see again: first, when participants needed to think about an answer rather than just perform a search for information, accuracy dropped. Second, it appears few people ventured beyond the first page of the layered policies. Kay and Terry found similar issues with layered policies [15].

In another sign that this was an easy question for most participants, times to answer were shorter than the other questions (2.3 minutes.) We found no significance for time based on company but format was significant. Privacy Finder (2.1 minutes) and Layered (2.3 minutes) supported faster responses than Natural Language, but the Layered condition was also more likely to result in incorrect answers.

4.2 Opt Out Link

We asked: Does the company provide a link to a webform that allows you to remove yourself from Acme's email marketing list?

Answer: Yes for all policies except: B NL, D NL, D Layered, E NL, which are No.[5]

This question is a little more difficult than the question about cookies. Policies refer to this concept as "opting out." For example, company C's natural language policy phrases it as "To opt out of receiving all other Acme mailings after you have registered, click here or click the appropriate unsubscribe link contained within the email that you receive." Participants need to map the concept of removing themselves from an email marketing list to the technical jargon of opting out. However, this question is again fairly straight forward. Either there is an opt out link or there is not. See Table 4 for a summary of results.

Table 4. Percentage correct and minutes to answer for the opt out question.

Policy	% correct	Time
A NL	33%	5.7
A PF	85%	3.7
B NL	33%	9.3
B PF	91%	4.6
B Layered	18%	4.8
C NL	80%	3.2
C PF	73%	5.1
D NL	29%	6.1
D PF	71%	3.8
D Layered	19%	5.5
E NL	55%	5.4
E PF	51%	4.6
F NL	93%	3.4
F PF	79%	3.7
F Layered	92%	2.2

We found significant differences for company and format. Natural language policy accuracy rates are dissimilar, with averages ranging from 93% (F) to 33% (A). Finding the opt out link in the A NL policy was looking

[5] Answers are not the same across a given company because the companies elected to provide different information in different formats. P3P requires an opt out link, which is then included in Privacy Finder.

for a needle in a haystack: there is one link halfway through the policy in the middle of a paragraph without any headings or other cues—and the policy runs to 13 pages when printed.

It would seem Privacy Finder should have consistent results across all six policies, since an opt out link is a standard part of Privacy Finder reports. However, companies with an opt out default have additional links for each category of opt out data. As a result, policies with opt out practices fared better, ranging from 85% correct (A PF) with less privacy protective practices and many prominent opt out links, to 51% correct (E PF) which required opt out for all data collection and had only one opt out link. Interestingly, the F PF policy (79%) has identical practices as E PF (51%) yet different accuracy scores. The author of the F PF policy included an additional opt out link in the text at the very end of the policy, which is prime real estate for readers' attention. Policy authors choices affect outcomes, even within the PF standardized presentation.

Since there is no requirement to discuss opt out choices within the layered format, once again we see dissimilar results across a standardized format. B layered policy (18%) required clicking the opt out link to see what it did, phrased as "For more information about our privacy practices, go to the full Acme Online Privacy Statement. Or use our Web form," with a link from "Web form" to the opt out page. In contrast, results were quite good with F layered (92%), which contained the same opt out text as at the end of the F PF (79%) policy.

We found significant differences in time to answer for company as well as format. We would expect longer times for longer policies since this is in many ways an information search task. Instead, time appears to be based on the underlying practices: policies without opt out links took longer. Since some of the policies with opt out links mentioned them at the end, it is unlikely the difference in times is based on reading through the entire policy to determine the absence of a link. Instead, participants likely re-read to satisfy themselves that they had not missed anything. Once again participants completed the task more quickly with layered (4.0 minutes) and Privacy Finder (4.2 minutes) than Natural Language (5.4 minutes,) but the wide variance and sometimes poor performance for standardized policies reduces the strength of this result.

Table 5. Percentage correct and minutes to answer for the email sharing question.

Policy	% correct	Time
A NL	76%	3.2
A PF	53%	5.4
B NL	49%	5.9
B PF	64%	5.9
B Layered	52%	4.8
C NL	80%	4.7
C PF	72%	6.9
D NL	67%	4.6
D PF	78%	4.0
D Layered	56%	4.7
E NL	53%	6.9
E PF	44%	6.2
F NL	50%	6.0
F PF	54%	4.4
F Layered	62%	5.0

4.3 Share Email

We asked: Does this privacy policy allow Acme to share your email address with a company that might put you on their email marketing list (with or without your consent)?

Answer Yes for all policies except: companies E and F (all formats) which are No.

We tested the wording of this question in multiple pilot studies to ensure people understood it without asking something pejorative or jargon-laden like "will Acme sell your email address to spammers." This question requires participants to understand the question, read the policy carefully, and make inferences for most policies. For example, C NL reads: "We may provide your contact information and other personal data to trusted third parties to provide information on products and services that may be of interest to you." Participants need to understand that "contact information" includes email, that "trusted third parties" are companies other than Acme, and that "provide information on products and services" means marketing messages, in order to correctly answer "Yes." See Table 5 for a summary of results.

Overall accuracy was only 60%. We found significant differences for company but not format. Times to answer averaged 5.3 minutes, which indicates people had a harder time completing this task. We found no significant results for time based on company or format.

As the answers to our questions become more nuanced we would expect the more readable policies to shine, yet that is not the case. Company A, with the hardest to read policy, had a higher accuracy score (64%) than F (55%) with the most readable policy and there was no overall discernible pattern based on readability. Similarly, we would expect standardized policies to convey information better, especially the Privacy Finder format which avoids the emotion-rich wording of "trusted third parties" and "valuable offers," yet we did not find significant differences between formats. Privacy Finder summarizes "With whom this site may share your information" as "Companies that have privacy policies similar to this site's" which again requires participants to refer to a separate section to determine if the parent company may engage in email marketing.

4.4 Telemarketing

We asked: Does this privacy policy allow Acme to use your phone number for telemarketing?

Answer Yes for all policies except companies A, E and F (all formats) which are No.

Participants struggled with this question as shown in Table 6. Except in the Privacy Finder version where companies are required to provide information about their telemarketing practices, policies typically do not highlight telemarketing practices. The way to answer this question correctly was typically to read through the entire policy for all mentions of when the company collects phone numbers, then see what policies they have around that data. For example, B NL

discloses telemarketing as: "You may also have the option of proactively making choices about the receipt of promotional e-mail, telephone calls, and postal mail from particular Acme sites or services." Sometimes policies were even more vague, for example D NL, "The information you provide to Acme on certain Acme Web sites may also be used by Acme and selected third parties for marketing purposes. Before we use it, however, we will offer you the opportunity to choose whether or not to have your information used in this way." Not only is telemarketing swept under the phrase "marketing purposes," telephone numbers are not mentioned explicitly either. It was necessary to deduce practices from a very careful and nuanced reading, frequently referring to multiple sections of the policy and then putting pieces together like a jigsaw puzzle. One could even make the case that answering "The policy does not say" is correct in cases as above where "information you provide" may be used for "marketing purposes" is by no means an explicit statement about telemarketing. However, we think it is important to note that the company likely does believe they have conveyed their practices: privacy policies are vetted by lawyers and are generally expected to be able to withstand a court or FTC challenge. If necessary, companies can point to the language in their policy and show that they did not violate the text by telemarketing.

We found significant differences in accuracy scores for company and format.[6] We found no significant results for time based on company but format does have significant differences. Once again layered (5.7 minutes) and Privacy Finder (5.5 minutes) are an improvement over natural language (8.2 minutes) but with the caveat that layered does not do as well for accuracy.

Even though we called out D NL as particularly indirect, it falls solidly in the middle of the accuracy scores (42%.) When participants cannot find information in layered policies, by design they should continue to the full policy for more details. In practice this appears not to happen, with a very low accuracy of 28%.

Privacy Finder does support more accurate answers (61%) even in contrast to natural language (39%.) Privacy Finder is the only format that requires a company to disclose, yes or no,

Table 6. Percentage correct and minutes to answer for the telemarketing question.

Policy	% correct	Time
A NL	23%	8.7
A PF	43%	5.9
B NL	41%	6.7
B PF	67%	5.9
B Layered	16%	6.2
C NL	42%	9.2
C PF	68%	5.5
D NL	42%	7.6
D PF	82%	3.2
D Layered	33%	5.5
E NL	65%	10.2
E PF	56%	5.4
F NL	26%	7.1
F PF	55%	7.4
F Layered	34%	5.9

[6] Accuracy scores for telemarketing are the single exception where including "Does Not Say" as a correct answer changes whether we find significance between formats.

if they telemarket. For example, under the heading "The ways your information may be used" D PF includes "To contact you by telephone to market services or products – unless you opt-out." Again there is a lot of variation between Privacy Finder policies based on the supplemental text they provide. For example B PF, is particularly confusing by stating in free form text "While Acme does not currently support telemarketing, it is possible that in the future Acme properties may contact you by voice telephone," directly above an automatically generated statement that they may use information for telemarketing.

5 Psychological Acceptability Results

After completing the initial accuracy questions, participants answered a series of questions designed to elicit their emotional reactions. Participants responded on a scale from 1 = strongly disagree to 7 = strongly agree. Most answers hovered right around 4, which is a neutral reaction. Higher numbers are always better.

5.1 Ease of Finding Information

We asked four questions about how easy it was to find information. We expected responses to these questions to reflect how well participants were able to understand a particular policy, and thus be related to the accuracy questions and times. However, we found few significant results. Participants found layered easier to understand even though they were less accurate with the layered format.

- "I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read" (M = 4.7, s.d. = 1.5.) We found significant effects for company but not format. A, B, and F (M = 4.8 for all) scored better than C, D, and E (M=4.4 for C and D; M=4.5 for E.)
- "I feel confident in my understanding of what I read of Acme's privacy policy" (M = 4.7, s.d. = 1.6.) We found no significant differences between companies or formats.
- "This privacy policy was easier to understand than most policies" (M = 4.5, s.d. = 1.5.) We found no significant differences between companies but did find significant results for formats. Layered (M=4.8) scored better than natural language (M=4.4) or Privacy Finder (M=4.4.)
- "It was hard to find information in Acme's policy" (M = 3.8, s.d. = 1.6.) We found no significant differences between companies or formats. (Note that based on the wording for this question we had to report the inverse of responses to keep higher numbers as better.)

5.2 Trust

If a format conveys information well but results in lack of trust of the company, it is unlikely that corporations will adopt the format. Participants trusted Privacy Finder formats slightly more than other formats.

– "I feel secure about sharing my personal information with Acme after viewing their privacy practices" (M = 4.0, s.d = 1.7.) We found significant effects for both company and format.
– "I believe Acme will protect my personal information more than other companies" (M = 4.0, s.d = 1.6.) We found significant effects for both company and format.

5.3 Enjoyment

We asked two questions to gauge how much participants liked reading the privacy policy. If people are unwilling to read policies then improving them does not provide much benefit. We found no significant differences between formats.

– "Finding information in Acme's privacy policy was a pleasurable experience" (M = 3.7, s.d. = 1.7.) We found no significant differences between companies or formats. This was the lowest score of all eight psychological acceptability questions.
– "If all privacy policies looked just like this I would be more likely to read them" (M = 4.2, s.d. = 1.7.) We found significant effects for format but not company.

6 Discussion

Our hypotheses were not fully supported and in some cases were refuted. Both layered and Privacy Finder formats did improve times to answer, but not by much, and at the expense of accuracy for layered policies. Privacy Finder policies showed modest improvement in accuracy for complex questions but no improvement for easy questions. While the accuracy scores for Privacy Finder were low in some cases, the format does represent a step forward from the status quo. Readability did not determine outcomes for natural language policies. For natural language, in some cases it appears the practices of the company were greater determinants than the words they used to describe those practices. We found few statistically significant differences in psychological acceptability.

Many researchers start from the observation that privacy policies are not usable in their current format and suggest ways to fix the problem. All of the formats were tested were unsatisfactory with a low rate of comprehension on questions that required synthesis of information. Participants did not like privacy policies of any type, and the highest mean score on the psychological acceptability questions was barely above neutral.

Privacy researchers tend to talk about policies as being uniformly bad. We expected that more readable natural language policies would have higher accuracy scores, lower times, and improved psychological acceptability than less readable policies, but that was not the case. These results could suggest that readability metrics are not a good way to differentiate between policies. This seems unlikely because the Flesch index has proven robust in many contexts

and we do not immediately see any reason why privacy policies should be dramatically different from other types of textual analysis. It seems more likely that the range from 32 to 46 on the Flesch index is too similar to see major variations in outcome: even the most readable policies are too difficult for most people to understand and even the best policies are confusing.

Our results are robust across a variety of different policies, but our study does not concretely identify what makes a given policy comprehensible. However, we can offer three observations. First, results from the layered format suggest participants did not continue to the full policy when the information they sought was not available on the short notice. Unless it is possible to identify all of the topics users care about and summarize to one page, the layered notice effectively hides information and reduces transparency. Second, participants struggled to map concepts in the questions to the terms used in policies. It may prove fruitful to research how people internally represent privacy concepts: which terms do they currently use and which industry terms do they understand? As suggested in the Kleimann report for printed financial statements, online privacy policies may need an educational component so readers understand what it means for a site to engage in a given practice [22]. Third, the standardized formats we studied still offer policy authors quite a bit of leeway. Companies with identical practices conveyed different information, and these differences were reflected in participants' ability to understand the policies. The flexibility of the standardized formats may undermine their expected benefits to consumers.

Our study used a between subjects rather than within subjects structure. We expect that we would see larger differences, particularly in psychological acceptability, if we were to place policies side-by-side. Prior work[7] found that when participants have both the natural language and the Privacy Finder versions available, Privacy Finder fares well. If people are reading multiple companies' policies to compare them, Privacy Finder may be advantageous. However, for just understanding a single policy, we find differences between formats are not as pronounced. By only showing one policy, our study did not capture one of the potential advantages to standardized formats. Standardized formats should be more useful once readers understand where to find information. Learning effects may play a role over time when people can take greater advantage of standardized formats as they become more familiar with their layout.

At this time, we do not recommend regulating the format of online privacy policies. While we did not find substantial benefit from the standardized formats we tested, that is not an inditement of the concept of standardized formats. Early results testing a new format for privacy policies based around a nutrition label concept are encouraging [16]. Ideally, future formats will identify problems with existing approaches and attempt to improve upon what has come before. In the future, we encourage rigorous testing for new formats before their supporters encourage wide-spread adoption.

References

1. Acquisti, A., and Grossklags, J. Privacy and rationality in individual decision

making. *Security & Privacy Magazine, IEEE 3*, 1 (January-February 2005), 26–33.
2. Anton, A., Earp, J. B., Qingfeng, H., Stufflebeam, W., Bolchini, D., and Jensen, C. Financial privacy policies and the need for standardization. *IEEE Security & Privacy 2*, 2 (Mar-Apr 2004), 36–45.
3. Bendrath, R. Icons of privacy, May 2007. `http://bendrath.blogspot.com/2007/05/icons-of-privacy.html` Accessed 22 Feb 2009.
4. Business Wire. European union issues guidance on privacy notices; new notices make it easier for consumers to understand, compare policies, January 2005. `http://www.tmcnet.com/usubmit/2005/jan/1104731.htm` Accessed 19 May 2009.
5. Center for Information Policy Leadership. Ten steps to develop a multilayered privacy policy, 2007. `www.hunton.com/files/tbl_s47Details%5CFileUpload265%5C1405%5CTen_Steps_whitepaper.pdf` Accessed 12 July 2007.
6. Children's Online Privacy Protection Act of 1998 (COPPA), Public Law No. 104–191, October 1998. `www.cdt.org/legislation/105th/privacy/coppa.html` Accessed 27 Mar 2007.
7. Cranor, L. F., Guduru, P., and Arjula, M. User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction (TOCHI)* (2006).
8. Egelman, S., Tsai, J., Cranor, L. F., and Acquisti, A. Timing is everything? the effects of timing and placement of online privacy indicators. In *CHI 2009* (Boston, MA, USA, April 2009).
9. Federal Trade Commission. Ftc staff revises online behavioral advertising principles, February 2009. `http://www.ftc.gov/opa/2009/02/behavad.shtm` Accessed 15 May 2009.
10. Graber, M. A., D'Alessandro, D. M., and Johnson-West, J. Reading level of privacy policies on internet health web sites. *Journal of Family Practice* (July 2002).
11. U.S. Gramm-Leach-Bliley Financial Modernization Act of 1999, Public Law no. 106–102, November 1999.
12. Hochhauser, M. Lost in the fine print: Readability of financial privacy notices, July 2001. `http://www.privacyrights.org/ar/GLB-Reading.htm` Accessed 27 Mar 2007.
13. Huang, H.-J. Language-focus instruction in EFL writing : Constructing relative clauses in definition paragraphs. In *2008 International Conference on English Instruction and Assessment* (2008). `http://www.ccu.edu.tw/fllcccu/2008EIA/English/C16.pdf` Accessed 22 Feb 2009.
14. Jensen, C., Potts, C., and Jensen, C. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies 63* (July 2005), 203–227.
15. Kay, M., and Terry, M. Textured agreements: Re-envisioning electronic consent. Technical report cs-2009-19, David R. Cheriton School of Computer Science, University of Waterloo, 2009.
16. Kelley, P. G., Bresee, J., Reeder, R. W., and Cranor, L. F. A "nutrition label" for privacy. In *Symposium on Usable Privacy and Security, (SOUPS)* (2009).
17. Lemos, R. MSN sites get easy-to-read privacy label. *CNET News.com* (2005). `http://news.com.com/2100-1038_3-5611894.html` Accessed 30 May 2007.
18. My Byline Media. The Flesch reading ease readability formula. `http://www.readabilityformulas.com/flesch-reading-ease-readability-formula.php` Accessed 9 Mar 2009.
19. OUT-LAW News. Drop the jargon from privacy policies, says privacy chief, September 2005. `http://www.out-law.com/page-5791` Accessed 23 Mar 2007.

20. Pollach, I. What's wrong with online privacy policies? *Communications of the ACM 30*, 5 (September 2007), 103–108.
21. Reeder, R. W., Kelley, P. G., McDonald, A. M., and Cranor, L. F. A user study of the expandable grid applied to P3P privacy policy visualization. In *WPES '08: Proceedings of the 7th ACM workshop on Privacy in the electronic society* (2008), ACM, pp. 45–54. http://portal.acm.org/citation.cfm?id=1456403.1456413# Accessed 22 Feb 2009.
22. Report by Kleimann Communication Group for the FTC. Evolution of a prototype financial privacy notice, 2006. http://www.ftc.gov/privacy/privacyinitiatives/ftcfinalreport060228.pdf Accessed 2 Mar 2007.
23. Saltzer, J. H., and Schroeder, M. D. The protection of information in computer systems. *Proceedings of the IEEE 63* (September 1975), 1278–1308.
24. Sheng, X., and Cranor, L. F. An evaluation of the effect of US financial privacy legislation through the analysis of privacy policies. *I/S - A Journal of Law and Policy for the Information Society 2*, 3 (Fall 2006), 943–980.
25. The Center for Information Policy Leadership, H. . W. L. Multi-layered notices. http://www.hunton.com/Resources/Sites/general.aspx?id=328 Accessed 23 Mar 2007.
26. The Office of the Privacy Commissioner. Release of privacy impact assessment guide and layered privacy policy, August 2006. http://www.privacy.gov.au/news/06_17.html Accessed 22 Feb 2009.
27. Tsai, J., Egelman, S., Cranor, L. F., and Acquisti, A. The effect of online privacy information on purchasing behavior: An experimental study. In *The 6th Workshop on the Economics of Information Security (WEIS)* (2008). http://weis2007.econinfosec.org/papers/57.pdf Accessed 22 Feb 2009.
28. Vila, T., Greenstadt, R., and Molnar, D. Why we can't be bothered to read privacy policies: models of privacy economics as a lemons market. *ACM International Conference Proceeding Series 5* (2003), 403–407.
29. W3C Working Group. The platform for privacy preferences 1.1 (P3P1.1) specification, November 2006. http://www.w3.org/TR/P3P11/ Accessed 28 Mar 2007.

Appendix

This appendix includes supporting statistical details. We performed all tests of statistical significance at the $\alpha = 5\%$ confidence level. We performed ANOVA analysis for both time data and psychological acceptability, which we recorded on a seven point Likert scale and treated as continuous variables. Accuracy questions were categorical data (either accurate or inaccurate) so we used Chi Squared tests. Details of that analysis follows.

6.1 Accuracy

Accuracy scores are all reported as the percentage of people who answered the question correctly. Answers are always either Yes, No, or the policy Does Not Say. We tested for statistically significant differences in mean accuracy rates by company (Table 7) and by format (Table 8).

Table 7. Statistical Significance Tests for Accuracy Questions by Company

Question	d.f.	χ^2 value	p	Significant?
Cookies	5	12.16	.033	✓
Opt Out Link	5	108.31	$< .001$	✓
Share Email	5	22.43	$< .001$	✓
Telemarketing	5	24.99	$< .001$	✓

Table 8. Statistical Significance Tests for Accuracy Questions by Format

Question	d.f.	χ^2 value	p	Significant?
Cookies	2	28.95	$< .001$	✓
Opt Out Link	2	40.80	$< .001$	✓
Share Email	2	1.90	.387	
Telemarketing	2	50.08	$< .001$	✓

6.2 Time

We recorded time in milliseconds though we reported it in minutes to assist readability. With such a fine grain unit of measure time is nearly continuous and we used ANOVA for analysis. We tested for statistically significant differences in mean times to answer by company (Table 9) and by format (Table 10).

Table 9. Statistical Significance Tests for Time to Answer by Company

Question	d.f.	F value	p	Significant?
Cookies	5	1.18	.320	
Opt Out Link	5	5.58	$< .001$	✓
Share Email	5	1.81	.109	
Telemarketing	5	1.75	.122	

6.3 Psychological Acceptability

We asked a series of questions to capture subjective impressions of the privacy policies. Responses were on a seven point Likert scale which is sufficient granularity to treat them as continuous variables. We performed ANOVA analysis to test for statistically significant differences in mean Likert scores by company (Table 11) and by format (Table 12).

Table 10. Statistical Significance Tests for Time to Answer by Format

Question	d.f.	F value	p	Significant?
Cookies	2	4.50	$< .012$	✓
Opt Out Link	2	3.59	.028	✓
Share Email	2	0.15	.864	
Telemarketing	2	8.59	$< .001$	✓

Table 11. Statistical Significance Tests for Psychological Acceptability by Company

Topic	Question	d.f.	F value	p	Significant?
Finding Info.	Explained thoroughly	5	1.9	.038	✓
Finding Info.	Confident understood	5	1.9	.099	
Finding Info.	Easier to understand	5	1.6	.148	
Finding Info.	Hard to find	5	.75	.589	
Trust	Feel secure	5	7.0	$< .001$	✓
Trust	Protect more	5	3.9	.020	✓
Enjoyment	Pleasurable	5	1.7	.135	
Enjoyment	Likely to read	5	2.4	.096	

Table 12. Statistical Significance Tests for Psychological Acceptability by Format

Topic	Question	d.f.	F value	p	Significant?
Finding Info.	Explained thoroughly	2	1.6	.203	
Finding Info.	Confident understood	2	.33	.722	
Finding Info.	Easier to understand	2	2.89	.051	
Finding Info.	Hard to find	2	.60	.549	
Trust	Feel secure	2	14.4	$< .001$	✓
Trust	Protect more	2	8.0	$< .001$	✓
Enjoyment	Pleasurable	2	.62	.539	
Enjoyment	Likely to read	2	2.4	.032	✓

A "Nutrition Label" for Privacy

Patrick Gage Kelley,[*] Joanna Bresee,[*] Lorrie Faith Cranor,[*] Robert W. Reeder[**]

[*] Carnegie Mellon University
School of Computer Science

[**] Microsoft
Trust User Experience (TUX)

ABSTRACT

We used an iterative design process to develop a privacy label that presents to consumers the ways organizations collect, use, and share personal information. Many surveys have shown that consumers are concerned about online privacy, yet current mechanisms to present website privacy policies have not been successful. This research addresses the present gap in the communication and understanding of privacy policies, by creating an information design that improves the visual presentation and comprehensibility of privacy policies. Drawing from nutrition, warning, and energy labeling, as well as from the effort towards creating a standardized banking privacy notification, we present our process for constructing and refining a label tuned to privacy. This paper describes our design methodology; findings from two focus groups; and accuracy, timing, and likeability results from a laboratory study with 24 participants. Our study results demonstrate that compared to existing natural language privacy policies, the proposed privacy label allows participants to find information more quickly and accurately, and provides a more enjoyable information seeking experience.

Categories and Subject Descriptors

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

General Terms

Design, Experimentation, Human Factors, Standardization

Keywords

privacy, P3P, policy, user interface, information design, labeling, nutrition label.

1. INTRODUCTION

Website privacy policies are intended to *assist* consumers. By notifying them of what information will be collected, how it will be used, and with whom it will be shared, consumers are, in theory, able to make informed decisions. These policies are also meant to inform consumers of the choices they have in managing their information: whether use of their information or sharing with third parties can be limited, and if it is possible to request modification or removal of their information.

However, Internet privacy is largely unregulated in the United States (except for children's privacy and some sector-specific regulations) and the privacy policies created by companies are frequently difficult for consumers to understand. Online privacy policies are confusing due to the use of specific terms that many people do not understand, descriptions of activities that people have difficulty relating to their own use of websites, a readability level that is congruent with a college education, and a non-committal attitude towards specifics [14]. These issues are complicated by companies creating policies that are tested by their lawyers, not their customers. It has further been established through numerous studies that people do not read privacy policies [21] and make mistaken assumptions based upon seeing that a site has a link to a privacy policy [26]. A recent study estimated that if consumers were somehow convinced to read the policies of all the companies they interact with, it would cost an estimated 365 billion dollars per year in lost productivity [20].

In addition, research has shown that consumers do not actually *believe* they have choices when it comes to their privacy. Based solely on expectations, they believe there are no options for limiting or controlling companies' use of their personal information [16]. This is a finding that we again validated in our work.

In short, today's online privacy policies are failing consumers because finding information in them is difficult, consumers do not understand that there are differences between privacy policies, and policies take too long to read. We set out to design a clear, uniform, single-page summary of a company's privacy policy that would help to remedy each of these three concerns.

This paper first presents related work describing standardization efforts in other domains in which companies present information to consumers to aid in their decision making, as well as early standardization efforts for privacy policies. Our approach comes from a broad survey of work that provides consumers with information: nutrition labeling, drug facts, energy information, and most recently work commissioned by the Federal Trade Commission to create a standard financial privacy notice. We discuss our iterative design approach, including focus group testing, as we developed and refined our information design over several months. Finally, we describe our 24-participant laboratory study and discuss the results of our initial evaluation.

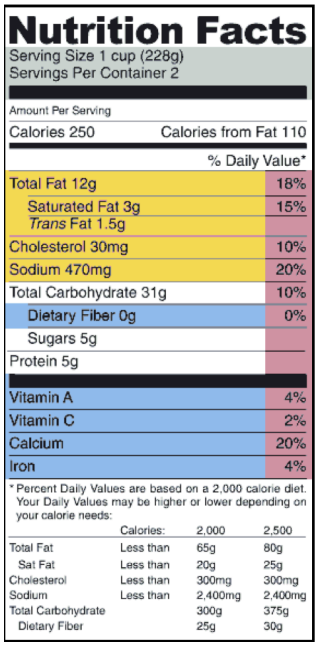
2. RELATED WORK

To better inform our design process we surveyed the literature surrounding other consumer labeling efforts: the "Nutrition Facts" panel, energy and drug labeling, and recent work on creating a standardized financial privacy notice. Additionally, we summarize our previous work on a standardized privacy policy format.

2.1 The "Nutrition Facts" Panel

In the United States, the nutrition label seen in Figure 1, has become iconic after being mandated by the Nutrition Labeling and Education Act of 1990 (NLEA) [28]. In the last nineteen years, its increasing ubiquity has led to a number of studies examining the

Symposium On Usable Privacy and Security (SOUPS) 2009, July 15-17, 2009, Mountain View, CA, USA.



Nutrition Facts

Serving Size 1 cup (228g)
Servings Per Container 2

Amount Per Serving

Calories 250	Calories from Fat 110
	% Daily Value*
Total Fat 12g	18%
Saturated Fat 3g	15%
Trans Fat 1.5g	
Cholesterol 30mg	10%
Sodium 470mg	20%
Total Carbohydrate 31g	10%
Dietary Fiber 0g	0%
Sugars 5g	
Protein 5g	
Vitamin A	4%
Vitamin C	2%
Calcium	20%
Iron	4%

*Percent Daily Values are based on a 2,000 calorie diet. Your Daily Values may be higher or lower depending on your calorie needs:

	Calories:	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Figure 1. The Food and Drug Administration's Nutrition Facts panel as regulated by the NLEA. Source: www fda.gov

costs of adoption and the ability to inform and change consumer purchasing decisions.

The sparse literature around the design of the nutrition label [3] focuses on the decisions made to simplify the information as much as possible for consumers. These decisions were made in part to address low literacy rates and the needs of older Americans. These guidelines include defining a zone of authority, providing quantitative information about nutrients, defining minimum font sizes, and equalizing labels across products by providing defined serving sizes and calculating percentages based on standardized daily amounts.

Surveys indicate that consumers would prefer that nutrition labels include more information. However, studies have shown that including more information would not actually be beneficial [10]. Studies conducted to examine the impact of the NLEA have found that it is the populations of people who are educated and already motivated to investigate nutritional information who benefit the most from nutrition labels [2][10]. Another study found that nutrition information had the greatest impact when there was a limited number of items from which to make a selection [24]. This result implies that the nutrition label made it easier to compare between a small set of items, allowing consumers to benefit, through informed decision making. Studies have demonstrated that nutrition labels have an impact on consumer decision making, with some user-reported effect sizes up to 48% after the initiation of NLEA [10]. For most studies, however, the effect of the nutrition label is small and most studies focus on specific nutrients such as fat intake or specific products such as salad dressings. We are not aware of controlled studies that measure the impact of nutrition labels on consumer behavior over an extended period of time.

Other studies have found that the effects of providing calorie information (not a complete nutrition facts label) in restaurant menus are often very small and the effects may vary depending on the population studied. In a study of meal choices at a sandwich shop, Downs et al. found that if participants were given menus that included calorie information, they ordered meals with about 50 fewer total calories than participants who did not receive calorie information. However, the authors stated that this was "an effect smaller than this study was powered to test." Nonetheless, they pointed out that if the finding proved reliable, it could be significant if it caused people to reduce their caloric intake by a similar amount for multiple meals each day. In a related study of food purchases at three New York City restaurants before and after a law went into effect mandating the posting of calorie information on menu boards, the authors found no effects of the legislation at two of the three restaurants. At the third restaurant they found a small effect. They noted that the effect was larger for dieters than for non-dieters, suggesting that the availability of label information may again be most useful to people who are already interested in the information provided by the label [9].

2.2 Other Privacy Notices

Layered Privacy Policies, a policy display format popularized by the law firm Hunton & Williams [25], involve a short form or summarized version of a privacy policy created using a step by step process. This summary has standardized headings for the policy information, but the information itself is provided by the company, in free-form natural language text.

The US Federal Trade Commission (FTC) is currently leading an effort to develop a standardized financial privacy notice. The Kleimann Group used an iterative design process to develop a prototype notice for the FTC, focusing on user comprehension, allowing users to "identify differences in sharing practices," and compliance with the regulations surrounding financial privacy notices specified in the Gramm-Leach-Bliely Act. Over a 12-month period the Kleimann Group iterated on several design prototypes, conducting focus groups and diagnostic usability testing [16]. Our iterative design approach followed a similar process of testing labels for comprehension and then overall design through focus groups.

The Kleimann Group final prototype consists of four parts: the title, the frame, the disclosure table, and the opt-out form. The disclosure table, which actually displays the company's privacy practices, makes up the majority of our label. The rest of the Kleimann Group prototype was educational information to build a foundation of terms and understanding for the user [16].

More recently, the Levy-Hastak report was released, detailing the results of a 1032-participant mail/interview study [17]. The authors conclude that the table format performed the best "on a diverse set of … measures." Additionally, this success is attributed to the table providing a more holistic context for the particular sharing of each financial institution.

2.3 Other Labeling Programs

We also explored energy labeling programs from the European Union [12] and Australia [11], the US Consumer Products Safety Commission's toy and game warnings [8], and the US FDA Drug Facts label [29], to gain a broader understanding of practices used in designing and defining labeling requirements.

In general, the standards documents [7][12][28] are occupied with defining precise guidelines to describe compliance with the various labeling requirements. This includes point sizes of rules and text, allowable typefaces, allowable colors, and minimum sizes. In some instances, such as choking warnings on children's games, standards also include placement requirements.

Recently, a number of labels have been introduced to provide ratings to consumers on a fixed scale, focusing on a single metric or small number of metrics. The Australian Water Efficiency Labeling System (WELS) [32] and the British Food Standards Agency's Signposting (or Traffic Light) [13] use very small indicators with accompanying ratings. The WELS program uses an indicator with a possible score out of six blue stars. The Signposting initiative rates the quantities of fat, saturates, sugar, and salt in foods using a red, amber, green traffic light coloring system. Early research [2][18] has shown that Signposting enhances consumers' ability to evaluate products more accurately and surveys show that ninety percent of consumers find this type of label useful.

2.4 The Platform for Privacy Preferences

Due to the difficulties surrounding the use of text privacy policies, the World Wide Web Consortium created the Platform for Privacy Preferences (P3P) [30]. P3P is a standard machine-readable format for encoding the online privacy policy of a company or organization. Once this P3P policy has been provided, consumers must use a user agent to interpret it into something understandable. Unfortunately, widely available P3P user agents have limited functionality. These include the P3P policy processing elements of common web browsers and a few privacy specific browser add-ons [6].

To provide consumers with an active tool where they can investigate and explore the privacy policy of a website, earlier work from the CyLab Usable Privacy and Security Lab (CUPS) produced the P3P Expandable Grid. This user agent was based on one of the central Expandable Grid objectives of displaying a holistic policy view [22]. The interface was created to use the entire P3P specification, broken down by categories. An example of the grid is shown in Figure 2.

The P3P Expandable Grid has two main parts: the header and the information display. In the header, there is a title, a legend that explains the 10 possible symbols (8 pictured) that may appear in the body of the grid, as well as expandable column headers that explain how that company uses data, and who they will share it with. Finally, in the top-right corner of the header is a button that toggles between showing and hiding information that isn't collected (i.e., hide rows that would be blank).

In the body, information is displayed in blocks that correspond to P3P Statements. Each block starts with a title and a short textual description (if available) and is followed by a hierarchy of expandable rows, which list what information this company collects. The symbols in each row show how that specific piece of information could be used or shared according to the policy. In this way we were able to show the entire depth of the P3P specification in a two-dimensional grid.

Based on an online survey of over 800 people in the summer of 2007, we found further evidence that people generally do not understand the information presented in privacy policies and also do not enjoy reading them. When comparing three formats: a standard natural language policy; PrivacyFinder, which is a simplified human-readable version based on a P3P policy and consisting mostly of bulleted lists; and the above version of the

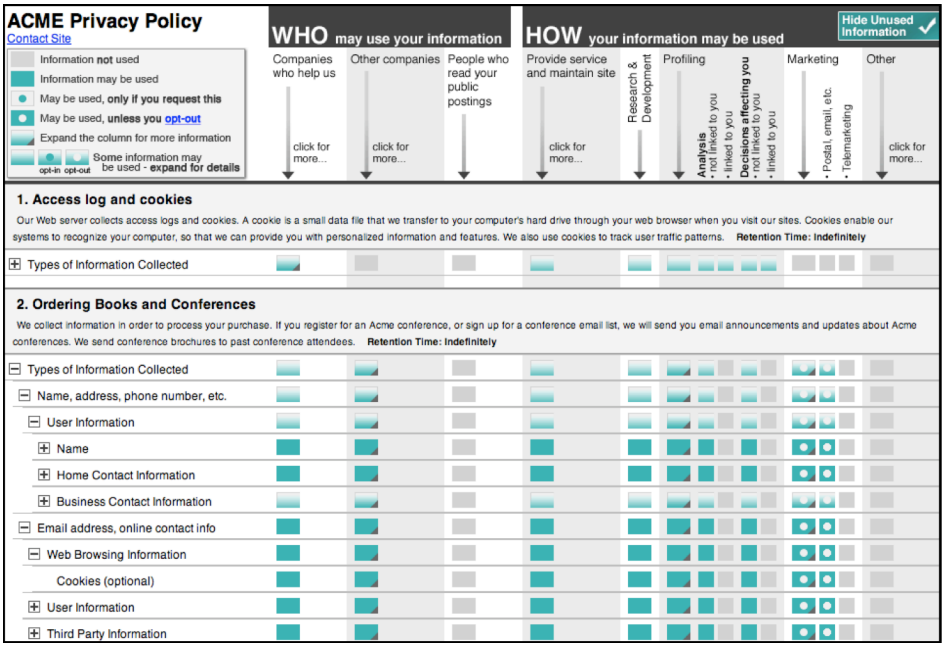


Figure 2. Our *P3P Expandable Grid,* an early attempt at a standardized information design for privacy policies. Due to its implementation of the entire P3P specification its complexity prevented large performance gains.

P3P Expandable Grid, we found that none of the three formats were found to be pleasurable to read or easy to comprehend. Notably, we found the P3P Expandable Grid to be slightly worse than the other formats, both in enjoyment and comprehension [23].

3. DESIGN METHODOLOGY

This section elaborates on our iterative design process, presenting several prototype labels with benefits and criticisms, and highlighting where knowledge from other label designs was applied. Throughout this process we leveraged informal user feedback as well as focus groups, which are discussed in detail in Section 4.

3.1 Problems with the P3P Expandable Grid

Based on the analysis of the previously mentioned P3P Expandable Grid study results and a subsequent lab evaluation, we identified five major problems with the Expandable Grid [15]:

- Many of the P3P labels are not clear to users. For example, "Profiling" and "Miscellaneous Data" are not terms that users encounter in the context of their use of websites.
- The legend has a large number of symbols including multiple symbols for expansion (depending on directionality), which the user may not understand.
- Multiple statements that may be related to the same types of information in a P3P policy are displayed separately, possibly requiring the user to check multiple rows to answer a single question.
- The Hide Used Information button in the top right only condenses unused rows, not columns.
- Rows with a plus symbol may be expanded; however, many users (40.7%) never expanded any data types. By not expanding data types, users never saw some important parts of the policy [23].

With these initial five problems in mind we abstracted several general principles from the nutrition labeling literature [3][4][27][28].

- Putting a box around the label identifies the boundaries of the information, and, importantly, defines the areas that are "regulated" or should be trusted. This is a common issue when the label is placed in close proximity to other information, but may not be as significant an issue online.
- Using bold rules to separate sets of information gives the reader an easy roadmap through the label and clearly designates sections that can be grouped by similarity.
- Providing a clear and boldfaced title, e.g., Privacy Facts, communicates the content and purpose of the label specifically and assists in recognition.

While much of the labeling literature also focuses on quantifiable properties, such as amounts of fats or fiber or percentages of active ingredients or calories from a standardized expected daily value, privacy policies typically do not include quantifiable measures, and the P3P specification includes no quantifiable fields. The Kleimann Group dealt with this lack of quantifiable information by moving to binary Yes/No statements, which they found to be readily understood by focus group participants.

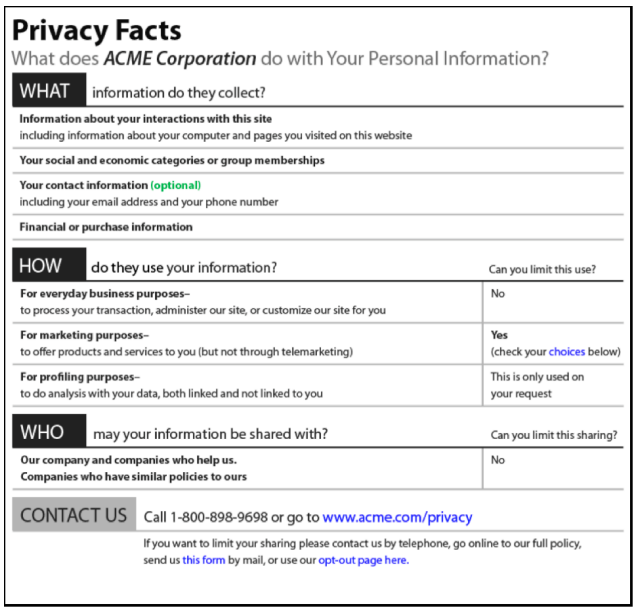


Figure 3. Our *Simplified Label*, an early attempt at a privacy label.

3.2 The Simplified Label

Our next design, following the P3P Expandable Grid, was the *Simplified Label.* In creating the Simplified Label, we used Yes/No statements and applied the three general principles discussed above. The Simplified Label is shown in Figure 3. (Note: as with each of the screenshots shown below, this is one of many variants of a similar vein. We show only one of each that we believe is representative of the entire series.)

While we made visual changes including adding a title and sub-head, adding bold lines, and simplifying the table view, the most significant change is a reduction in complexity. Two changes contributed most to simplifying the label: eliminating P3P statement groupings and eliminating the use of P3P data hierarchies. These changes are detailed below.

3.2.1 P3P Statements

P3P specifies data groupings called STATEMENT elements [31]:

> The STATEMENT element is a container that groups together a PURPOSE element, a RECIPIENT element, a RETENTION element, a DATA-GROUP element, and optionally a CONSEQUENCE element and one or more extensions. All of the data referenced by the DATA-GROUP is handled according to the disclosures made in the other elements contained by the statement.

This means that all of the collected information in a statement can be used for certain purposes, and can be shared in the same way. A useful model is to think of P3P as consisting of multiple triplets of information, {data, purpose, recipient}. We do not include retention because our analysis of over 5000 unique P3P policies collected by the Privacy Finder search engine [6] shows that the majority of P3P policies state that data is retained indefinitely. In cases where a website has a different data retention policy we include a note at the bottom of the label.

Due to P3P information naturally falling into these triplets, a display such as the list in Figure 3 suffers some information loss. For example, it is possible contact information is used for

marketing exclusively and purchase information is used for profiling purposes exclusively. Or it is possible that both contact and purchase information could be used for either purpose. By removing the triplets and only displaying a list, we lose that distinction. This tends to make privacy policies appear more permissive than they actually are.

A P3P policy may also have multiple statements. In the P3P Expandable Grid, statements were displayed in a numbered list. In the Simplified Label we have merged multiple statements into a single list. For example, consider a policy where the first statement of a policy was about cookies and the second dealt with web activity. In the P3P Expandable Grid we would list the categories twice. The first time only cookies would be highlighted; the second, web activity. With the Simplified Label we show the information from all of the statements in a single list.

3.2.2 P3P Data Hierarchies

P3P allows for two interchangeable and different hierarchies of data (collectable information). The more commonly used is categories: a list of 17 types of information that companies can collect. When a category is specified a company reserves the right to collect any information that falls under that category (i.e. "Physical Contact Information" includes name and telephone number). The other data hierarchy, the base data schema, includes every data element that can be specified using P3P, hierarchically arranged (e.g., NAME is a child of USER and includes GIVEN[name], MIDDLE[name], and FAMILY[name]). Further complicating the situation, every element belongs to one or more category (NAME is a member of both demographic data and physical contact information because one's GIVEN name is part of their contact information while one's FAMILY name provides demographic information).

In the original P3P Expandable Grid, each category was displayed in its entirety in each statement, with each element of the base data schema hierarchically arranged as children. This led to nearly 800 elements per category (if fully expanded). To simplify, we decided to display only data categories. While this affords us a list of possible information that can fit on a page, it suffers when companies state they will only collect specific items. For example Contact Information would be displayed similarly if a company collected a consumer's name, their postal address, their telephone number, or all of the above information. One way of preserving some of this detail would be to display the specific data elements a company collects when a user clicks on the name of a category.

3.2.3 Design Notes

To further reduce complexity, information that is not collected or purposes that are not mentioned in a particular policy are not shown. The Show/Hide information button has also been removed; thus, there is no way to see uncollected information.

Finally, we have defined a maximum width of 760px for this label and all following designs in this paper. One important consideration was that the privacy label design be printable to a single page and viewable in the standard width of today's internet browsers.

3.3 The Simplified Grid

While the above label is extremely simple and closely follows a pattern established by the nutrition facts panel and the financial privacy notice, we felt that it sacrificed too much detail.

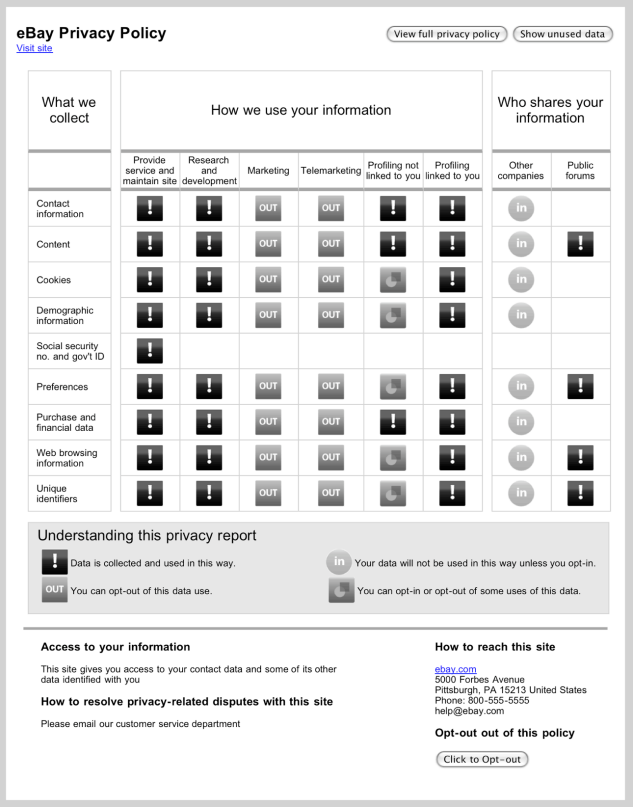


Figure 4. Our *Simplified Grid* in which the grid concept is reintroduced to the label.

The goal of our next design was to bring back more of the detailed information that privacy policies can provide without overwhelming users. To do this we decided to try to find a happy medium between our Simplified Label and the best aspects of the original P3P Expandable Grid. We adopted a two-dimensional grid layout, as shown in Figure 4. We call the resulting design the *Simplified Grid*.

3.3.1 Simplifying the P3P Expandable Grid

While the P3P Expandable Grid was not successful, this failure was not a result of the tabular display. Also, as discussed above, due to the nature of P3P Statements each reduction in dimensionality causes a loss of information and we wanted to minimize information loss to most benefit consumers. With the reintroduction of the two dimensional layout several changes were made. As mentioned in 3.2.2 we only used Data Categories to show what information companies collect, but we also simplified recipients and purposes .

Purposes, of which there are 12 specified[1] in the P3P specification, were grouped similarly to the categories in the P3P Expandable Grid. However the sub-categories were removed. Thus, Administration, Current Transaction, and Tailoring are all

[1] The P3P specification specifies 12 purpose elements: Current, Admin, Develop, Tailoring, Pseudo-analysis, Pseudo-decision, Individual-analysis, Individual-decision, Contact, Historical, Telemarketing, and Other-Purpose [31].

A bold title is used to set the context for the information.

Short labels are used for column and row headers, with longer definitions on our Useful Terms page.

Information that is not collected has a saturated label and a row full of the lightest symbol.

Four symbols on a scale from light to dark are used to indicate the severity of certain privacy practices.

Row and column locations are consistent so that two policies side-by-side can be easily visually compared.

A legend provides information about what each symbol means.

Figure 5. Our *proposed Privacy Nutrition Label.* This label is the one we tested in the second focus group and the laboratory study.

grouped under the title "Provide service and maintain site." We split the four P3P profiling-related purposes into two categories, based on whether that profiling is linked to the users' identity or performed anonymously. However, during our user testing, this distinction proved unclear to users.

Of the 6 recipients specified by P3P[2], Ours and Delivery are both never shown, as it is implied that the given company will always maintain the information. "Other Companies" merges the three remaining types of recipients, distinguished by their own privacy practices. We decided the importance of this column was to show whether any sharing with other companies was taking place. Public forums remained unchanged.

3.3.2 Symbols & Mixed Control

While you cannot opt-in or out to the trans-fat in your salad dressing, you might be able to have control over certain aspects of your information sharing on the internet. The Yes/No dichotomy advocated by participants in the Kleimann Group's studies works when there are only one, or maybe two, columns of information. Here we would have needed 8 columns and 10 rows of Yes/No information, which would have been visually difficult to parse.

[2] The P3P specification specifies 6 recipient elements: Ours, Same, Other-recipient, Delivery, Public, Unrelated [31].

Instead we again looked back to the P3P Expandable Grid and used symbols. However, while the P3P Expandable Grid had an array of 10 symbols, the Simplified Grid uses only four:

- **Exclamation Point**: Data is collected and used in this way.
- **OUT** (in a square): You can opt-out of this data use.
- **IN** (in a circle): Your data will not be used in this way unless you opt-in.
- **Square and circle**: You can opt-in or opt-out of some uses of this data.

Each of these four symbols was defined in a legend labeled "Understanding this privacy report" directly below the policy. The legend is another device borrowed from the P3P Expandable Grid; however, it has been moved below the policy.

Again, due to the way P3P uses data statements, it is possible that in some instances consumers might be able to opt-out of allowing their demographic information to be used for profiling, but in others it is required, or opt-in. The "square and circle" or "mixed choices" symbol attempts to convey this possibility; however, in our user testing it was found to be incomprehensible.

3.3.3 Visual Intensity

The Simplified Grid is the first iteration of our label to use visual intensity to provide a high-level indication of the quality of a given policy. Each of the four symbols has been colored such that darker symbols represent what could be more privacy-invasive practices. The use of intensity allows users to make quick visual comparisons that would not have been possible with text alone.

3.3.4 Testing

The most significant issue that arose in our testing was confusion over blank areas of the label. We thought that blank areas would clearly indicate information a company does not collect; after all, natural language policies typically leave out any mention of types of information the company does not collect. However, in testing, many participants were unclear on the meaning of the blank cells. Some inferred the accurate meaning that such information uses would not occur, but others thought it allowed the company free reign to do anything in those situations or that they simply had not yet decided their practices.

3.4 Final Proposed Privacy Nutrition Label

Our *Privacy Nutrition Label,* shown in Figure 5, is a direct descendent of the Simplified Grid. With the Privacy Nutrition Label, we sought to refine the strengths of the Simplified Grid by reducing clutter, introducing color, and simplifying symbols.

3.4.1 Types of Information Displayed

We made changes in the way we present data categories as rows in the table to better facilitate comparisons between policies and to reduce confusion about what data is being collected.

All of the P3P Data Categories are now represented in rows regardless of whether they are collected or not. For example, the label shown in Figure 5 indicates health & financial information are not collected (and thus not used or shared), but they have not been removed. Any policy displayed in this format will have exactly 10 rows, and the ordering will always be consistent. This allows two policies to be easily, visually compared side-by-side.

Participants in a focus group we conducted after making this change did not understand which information companies were not collecting. We indicated the information that was not collected with rows completely filled with minus symbols, but participants believed that companies collected every piece of information listed on the grid. One participant asked, "Why would they collect all that information if they're not going to do anything with it?" In the final prototype we grayed out the labels for data that companies did not collect, and we changed the minus symbol's description from "we will not use your information in this way" to "we will not collect or we will not use your information in this way." We also changed the row-heading label from "What we Collect" to "types of information." This change was made to highlight the fact that we now show even un-collected information and to reduce confusion about what was and was not being collected.

3.4.2 Symbol Changes & Color

In the Simplified Grid design, we marked types of information that companies collected and left other cells in the policy blank. However, half of the participants were afraid of the blank spaces; for instance, one said, "Nothing is mentioned. It is completely open-ended. These guys [the company] can modify these values." Therefore, in the final version we introduced a symbol to indicate that information was not collected or used.

Focus group participants found the mixed choices symbol confusing so we removed it. Instead we now display the symbol for the most invasive practice. For example, if in some circumstance one can opt-in and in another one can opt-out, we display the opt-out symbol.

We constrained our initial designs to grayscale to facilitate easy printing without loss of information and to reserve color for highlighting differences between a policy and a user's personal preferences (something we plan to implement later). However, feedback indicates that color seems to improve user enjoyment in reading the label, although we have not yet quantified this improvement. We selected the colors used in our label with care to accommodate viewers with color-blindness, allow for grayscale reproduction, and maintain the darker-is-worse high-level visual feedback discussed in Section 3.3.3.

3.5 Useful Terms

Even with the "understanding this privacy policy" legend in place there was still confusion over many of the terms used in the label. This was also a common issue during the development of the Kleimann Group's Financial Privacy Notice, and in response they developed what they call the "Secondary Frame." This portion of the prototype notice included both frequently asked questions and a series of extended definitions, which are: "[not] information as essential for consumers to have, but consumers often commented that they liked having it included." [16 p.27]

Our version of the "Secondary Frame" is a single page hand-out of useful terms. Our useful terms information was informed by the Human Readable definitions included in the P3P 1.1 Working Group Note [31] and consists of seventeen definitions, one for each of the row and column headers. Some are straightforward, others more detailed. For example, the definition of telemarketing states: "Contacting you by telephone to market services or products," while the profiling definition is:

> Collecting information about you in order to:
>
> - Do research and analysis
> - Make decisions that directly affect you, such as to display ads based on your activity on the site.

Information that the site collects about you may be linked to an anonymous ID code, or may be linked to your identity.

In future versions, clicking on or hovering over the headers could pop-up these definitions.

4. FOCUS GROUPS

We held two, hour-long focus group sessions to review the design and discuss participants' impressions and questions. We recruited focus group participants from the Carnegie Mellon University (CMU) Center for Behavioral Decision Research (CBDR) participant recruitment website. We paid participants $10 to participate in a 60 minute focus group.

The first focus group was composed of three female and seven male CMU students. The participants reacted positively to the Simplified Grid. For example, one participant stated, "This is more convenient than scrolling through reams and reams of paragraphs. I mean who reads them?" and another participant said, "I like the chart. [It's] better than long sentences." However, we found that some participants still had problems understanding privacy concepts. For example, one participant asked, "What is the difference between opt-in or opt-out?" and many others agreed that they did not understand this distinction. Additionally, many participants had trouble distinguishing different privacy concepts. Most participants were familiar with profiling, but did not understand the difference between "Profiling linked to you" and "Profiling not linked to you." Similarly, participants did not understand the different meanings of "cookies" and "unique identifiers." It was this vein of feedback that led to the inclusion of the useful terms definitions described in Section 3.5.

By asking participants to compare two policies, we found that participants could easily isolate and describe differences. Participants noticed that Policy A had more opt-in symbols and Policy B had more opt-out symbols. However, participants were not able to make accurate judgments about the policies. When we asked the participants to chose the company with whom they would prefer to do business, five of the ten participants chose Policy B: the company that collected and used more of their personal information.

Using the feedback from the first focus group, we initiated another series of rapid iteration and prototyping, which resulted in the final label prototype. Our second focus group compared the final Privacy Nutrition Label to the Simplified Label.

The second focus group was composed of four female and three male undergraduate students from CMU and the University of Pittsburgh. When reviewing the Privacy Nutrition Label vs. the Simplified Label we found that participants better understood the grid and were able to make more accurate side-by-side comparisons. Participants understood the significance of the red symbols, saying, "Red is for 'stop' or 'danger.'" We passed out two privacy policies, Policy A and Policy B, and asked the participants to raise their hands if they believed that Policy A is the better policy. Every participant raised his or her hand, correctly identifying Policy A as the more favorable policy. Participants demonstrated a detailed understanding of the differences between the policies with comments such as "It's very clear which site is best" and "You should pick a site with more opt-ins than opt-outs." Some participants even noted subtle differences between the two policies saying, "Policy A isn't perfect either, because they share your preferences, and this may include things like your religious or political preferences."

After reviewing the grid design, we passed out the simple text policy. Participants reacted negatively to the text policy because they felt that it did not provide enough information, saying, "This is an empty policy, it says nothing. I wouldn't trust it." Participants wanted to see how each piece of information was being used. For example, one participant stated, "With the grid it's easier to see things. What information is being shared? We don't know that anymore."

5. USER STUDY METHODOLOGY

Based on the feedback from our second focus group we performed a 24-participant laboratory user study comparing a standard natural language (NL) privacy policy with privacy policies presented in our Privacy Nutrition Label.

We used a within-subjects design where participants were randomly assigned to first use either the label or the natural language format. Each participant completed 24 questions relating to the policy format they were shown first and then the same 24 questions again with the other format. These tasks are detailed below. We recorded accuracy as well as time for each participant.

5.1 Participants

We recruited the 24 participants through the CBDR website. Our only requirement was that English be the participant's native language. We offered participants $10 to participate in a 45 minute study in our laboratory.

Our participants included 16 students and 8 non-students. Of the 16 students, 5 studied humanities, 5 economics or business, 2 science, and 4 information science. 16 of our participants were male, 8 were female.

5.2 Privacy Policy Selection

Our study used two NL privacy policies and two label formatted policies. We started with the current actual P3P policy of a popular online e-commerce website. We modified this policy in three ways to produce two different label policies for the mythical companies Acme and Button. The first change was to the data collected. Acme has preference information collected but not demographic information, whereas Button Co., collects demographic, not preference. This change is not incredibly significant but does distinguish the data collection. The second change was to the data uses. Acme does not do any profiling while Button Co. does. The third change was to information sharing practices. While Acme only shares information when consumers opt-in, Button Co. shares information unless consumers opt-out. These significant differences were introduced so that there would be a clear "correct" response for participant tasks that require them to determine which company better protects their privacy (see 5.3.3).

The two NL policies for the mythical companies ABC Group and Bell General represent the exact same policies as described above. The ABC Group policy is the natural language policy of the same company whose P3P policy was used to populate the grid, again with the three modifications above made to make it match Acme's. We could not however simply make the three modifications to the policy and also present it as the other natural language option because two different companies, no matter how

Table 1. Extended Text & Readability Comparison for NL

Policy Metric	ABC	Bell
Word Count	2287	2299
Sentence Count	136	130
Flesch Reading Ease	42.06	41.69
Flesch-Kincaid Grade	11.57	11.84

similar their practices, would not share the same text. The introduction, structure, and actual language used needed to be different. Thus, to create the Bell General policy we used the text of a different, yet comparable e-commerce website, and changed the practices so as to match that of Button Co.

In editing the natural language policies we removed any references to programs that would distinguish the companies (such as specially branded programs), removed lists of links from the beginning of the policies, removed references to Safe Harbor, and additionally modified the second policy so that both were approximately the same length. For a more complete comparison see Table 1.

We chose not to use layered policies. This decision was made because layered policy adoption is not consistent or widespread, most common layered policies would not be suitable for answering the questions we asked, and finally recent research has suggested layered policies are no better at helping consumers understand privacy than full natural language policies [19].

5.3 Task Structure

The task structure for each condition was exactly the same, with 24 tasks comprising a section. These sections can be split into four parts, each of which is detailed here:

5.3.1 Information Finding

The first 8 questions were all Yes/No questions asked of a single policy (ABC Group for NL, Acme for the label). Of these 8 questions, 6 were single-element questions, involving only one element of the P3P statement triplet. For example: "Does the policy allow the Acme website to use cookies?" to which the answer was Yes, or "Does the policy allow the Acme website to share your information on public bulletin boards?" to which the answer was also Yes.

The remaining two questions all required two parts of the triplet to answer the question, for example "By default, does the policy allow the Acme website to collect your email address and use it for marketing?"

5.3.2 Perceived Privacy Policy Understanding

Following the 8 information finding questions, participants were given 6 questions on a 5-point Likert scale, from Strongly Disagree (1) to Strongly Agree (5). Each of these is described below.

The first question: **L1**: "I feel secure about sharing my personal information with Acme after viewing their privacy practices" attempts to capture participants' reaction to the actual content of the privacy policy they read. **L2**: "I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read" questions whether participants believe their practices are well displayed.

The next three questions deal with the experience of interacting with the privacy policy in the format we presented. **L3**: "Finding information in Acme's privacy policy was a pleasurable experience" has participants rate their enjoyment of finding information. **L4**: "I feel confident in my understanding of what I read of Acme's privacy policy" investigates participants' perceived accuracy in the earlier questions. **L5**: "It was hard to find information in Acme's policy" has participants rate the difficulty they had in finding information.

The final question, **L6**: "If all privacy policies looked just like this I would be more likely to read them" attempts to capture whether our proposed label would encourage more people to read privacy policies.

5.3.3 Policy Comparison Questions

The third section requires participants to compare two policies of the same format (ABC Group v. Bell General for NL or Acme v. Button Co. for the label). One of the policies in each comparison is the same policy from the initial 8 information-finding questions.

The first four questions in this section are True/False statements such as "By default, **Button Co.** can share information about your purchases with other companies, but **Acme** cannot."

The final two questions in this section are opinion questions, asking: "Which company will better protect your information online?" and "You're looking to buy a gift online. At which company would you prefer to shop?"

5.3.4 Policy Comparison Enjoyment & Ease

The final four questions are again on the 5-Likert scale presented earlier. They are in two pairs, the first pair asking if, "Looking at policies to find information was an enjoyable experience" and "Looking at policies to find information was easy to do." The second pair focuses specifically on the comparison task, "Comparing two policies was an enjoyable experience" and "Comparing two policies was easy to do."

6. RESULTS

The results from our laboratory study are presented below. First

Table 2. McNemar's p-values & Benjamini-Hochberg Correction p-values for information finding questions 1-8 (5.3.1), and policy comparison questions 15-18 (5.3.3).

	Label	NL	McNemar's	Benjamini-Hochberg Correction
1	96%	100%	NS	NS
2	**88%**	**29%**	**0.00024**	**0.0014**
3	100%	96%	NS	NS
4	92%	100%	NS	NS
5	54%	25%	0.12	0.21
6	**79%**	**21%**	**0.00012**	**0.0014**
7	75%	54%	0.3	0.45
8	88%	58%	0.09	0.18
15	96%	63%	0.06	0.14
16	92%	79%	NS	NS
17	**83%**	**38%**	**0.007**	**0.021**
18	**71%**	**25%**	**0.0009**	**0.0036**

we will address the issue of information finding through our quantifiable accuracy results. Next we describe the timing data on those questions, showing information finding is not only more accurate but also faster with label polices than with NL policies. To conclude this section we will present the "likeability" of the privacy label.

6.1 Accuracy Results

At a high level, people were able to answer more questions correctly with the label. We compared the correct number of total questions, per participant, for the label vs. the natural language policy, M = 10.13 and M = 6.83 respectively, $t(23) = 7.41$, $p < 0.001$.

We explored each of the questions individually by testing the proportions of correctness for each question by condition, using McNemar's test. These results combine participants who saw the label first and with participants who saw the label second as accuracy differences were not significant between these two conditions. These comparisons show that the label is significantly more accurate in 2 of the 8 information-finding questions and 2 of the 4 policy-comparison questions. The accuracy rates for each question are shown in Table 2, with statistically significant comparisons shown in bold.

We performed a Benjamini-Hochberg correction to account for multiple testing across comparisons. Each of the paired proportions are shown in Table 2 along with the McNemar's p-values and the corrected p-values.

6.2 Timing Data

For each of the information-finding and policy-comparison questions we collected time-to-task completion data. As shown in Table 3, the label was significantly faster than the natural language policies for both the group of information-finding questions and the group of policy-comparison questions ($p<0.001$).

To test the mean task completion time for accurate answers we removed all timing results where the resulting answer was inaccurate and calculated means per question, per condition. Using a 2-sided t-test the label is significantly faster in 2 of the 8 information-finding questions and significantly faster in 3 of the 4 policy-comparison questions. In only one question was the average time faster for participants using the natural language policy, and this difference was not significant. The full results for this test can be found in Table 4.

6.3 Satisfaction Results

The satisfaction results were captured based on participants' responses on a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). We computed the mean response for the label and for natural language, both combined, and also separated by which format was viewed first. For each of these questions higher is better, including Question L5 "information was hard to find," which was reversed to be consistent with the remaining questions.

We performed t-tests for each of these questions, to compare the label to the natural language policies. All but 2 of these 10 questions resulted in significant results. The label was rated significantly more pleasurable, easier to find information in, and easier and more enjoyable to use when comparing two policies.

Table 3. Time-to-task comparisons between the label and natural language policies. Shorter times are better. Information Finding is questions 1-8 (5.3.1), Policy Comparison, questions 15-18 (5.3.3)

Times in seconds.	**Label**	**NL**
Information Finding	174.5	349.6
Policy Comparison	120.0	292.4
Average Total Time	339.9	692.0

Table 4. Time differences and p-values for average time per question comparing only correct answers. All times reported in seconds.

	Label	NL	Difference	**p-value**
1	37.58	61.27	23.69	0.07
2	**21.67**	**85.7**	**64.03**	**0.04**
3	**14.35**	**50.07**	**35.72**	**<.001**
4	18.89	23.09	4.2	0.4
5	34.51	29.95	-4.56	0.46
6	20.19	50.24	30.05	0.06
7	16.32	22.82	6.5	0.88
8	26.93	36.79	9.86	0.73
15	**46.58**	**132.69**	**86.11**	**0.0006**
16	**34.36**	**68.32**	**33.96**	**0.05**
17	21.91	35.48	13.57	0.28
18	**12.24**	**47.36**	**35.12**	**0.03**

The results from each of these questions are shown with means and p-values in Figure 8.

Additionally we performed 2-sample t-tests between conditions to exploring priming effects, where opinions have changed based on the policy format a participant viewed first. When looking at how participants answered the Likert scale questions about the label by condition, 3 questions had significant results. Participants felt significantly more secure when viewing the grid if they saw the NL policy first, (label first=2.92, NL policy first=3.92, $p=0.03$) reported they were significantly more likely to read policies more in the label format if they saw the NL policy first (label first=4, NL policy first=4.5, $p=0.04$), and found comparisons on the label significantly easier when viewing the NL policy first (label first=3.92, NL policy first=4.58, $p=0.004$). These results show significant priming to appreciate the grid more when the NL policy was viewed first.

6.4 Observations

The initial results we have presented above are very strong, however there is still much room for improvement. We observed that some participants still found elements of the label confusing. We began an additional round of iterative design and testing to address some of the issues we observed during the lab study.

Several participants were confused by the symbols we used to indicate opt-in and opt-out. For instance, one participant did not understand what "out" meant, saying, "I've been messing things up because I thought 'out' meant 'out of the question.'" To

improve users' comprehension, we will alter the symbol design to include the full phrases "opt-out" and "opt-in."

In addition, several participants in the lab study were completely unfamiliar with the terms opt-out and opt-in, and they assumed that the terms meant exactly the same thing. We will continue to refine our glossary definitions to help educate users about these concepts. The original definitions did not explain the terms opt-in and opt-out, with the legend reading "we will collect and use your information in this way unless you opt-out." The new definitions help explain the concepts, stating: "we will collect and user your information in this way unless you tell us not to by opting out." We plan to further test our design changes in focus groups, and believe that the design iterations will continue to improve the speed and comprehensibility of the Privacy Nutrition Label.

7. DISCUSSION

We began this paper with three factors in mind: the ability to find information, the understanding that there are differences between privacy policies and control over one's information, and the simple time-based costs of reading privacy policies. We strove to design a single page summary of a company's privacy policy that would help to remedy each of these three concerns and at the same time be enjoyable.

We believe that the results presented above clearly show that each of these areas was addressed. Accuracy results were better or similar for information finding and policy comparison. Task completion times were significantly lower when using the label than when using a natural language policy. And across the board, participants believed information was easier to find and had a more pleasurable time finding it using the label.

The final label design allows for information to be found in the same place every time. It removes wiggle room and complicated terminology by using four standard symbols that can be compared easily. It allows for quick high-level visual feedback by looking at the overall intensity of the page, can be printed, can fit in a browser window, and has a glossary of useful terms attached. People who have used it to find privacy information rated it as pleasurable. They not only rated it better than the natural language, but actually rated it enjoyable to use.

When using the label people far more consistently selected the company that had the stronger privacy policy. Participants also realized the benefits of the label for comparison: "This may actually be the biggest advantage of this system because you can put down two polices that are formatted the same and see the exact differences between them. It's really easy." Even more directly one participant said "I guess I'll look to see which policy has more blue," exactly capturing one of our intended design goals.

A number of open questions remain about how people will use the label in practice. Will people make more use of the label than they currently do of privacy policies? How will their use change as they become more familiar with the labels through continued use over time?

Our next step will be to iterate on a number of additional minor changes and then run a large online study, similar to Reeder et al.'s original test of the P3P Expandable Grid [23]. This will further confirm over a much larger and more diverse group of people that the label is in fact, more accurate, faster, and more pleasurable. Additionally as this study will be conducted online, people will be viewing privacy policies just as they normally would, at their computer, which is very different than performing these tasks in our laboratory on paper.

Finally, we plan to integrate a version of the privacy label into Privacy Finder, a privacy search engine maintained by the CUPS Laboratory. This will allow people to use the label outside of the context of a research study and will allow us to monitor frequency of use while collecting feedback on the label design. It is likely this public online deployment that will bring us closer to answering how much a standardized label design assists people over time as they become accustomed to using it.

8. ACKNOWLEDGMENTS

The authors would like to acknowledge Sungjoon Steve Won for his early designs, including the simplified grid; Janice Tsai for her statistical expertise; Daniel Rhim, Robert McGuire, and Cristian Bravo-Lillo for their technical assistance and assistance in conducting user studies; Norman Sadeh and Aleecia McDonald for their guidance and advice; and everyone who provided input throughout the design process.

This work was supported in part by U.S. Army Research Office contract DAAD19-02-1-0389 ("Perpetually Available and Secure Information Systems") to Carnegie Mellon University's CyLab, by NSF Cyber Trust grant CNS-0627513, by Microsoft through the Carnegie Mellon Center for Computational Thinking, FCT through the CMU/Portugal Information and Communication Technologies Institute, and the IBM OCR project on Privacy and Security Policy Management.

9. REFERENCES

[1] Balasubramanian, S. and Cole, C. "Consumers' Search and Use of Nutrition Information: The Challenge and Promise of the Nutrition Labeling and Education Act." *Journal of Marketing.* 2002. Vol. 66, 112-127.

[2] Beard, T.C., Nowson, C.A., Riley, M.D. "Traffic-light food labels." *Med J Aust.* 2007;186:19.

[3] Belser, B. Designing the Food Label: Nutrition Facts. *AIGA Journal.* 1994.

[4] Buckley, P. and Shepherd, R. Ergonomic factors: The clarity of food labels. *British Food Journal.* 1993. 95

[5] Byrd-Bredbenner, C., Alfieri, L., Wong, A., and Cottee, P. The Inherent Educatiional Qualities of Nutrition Labels. Family and Consumer Sciences Research Journal, Vol 29, No 3, March 2001 265-280.

[6] Cranor, L., Egelman, S., Sheng, S., McDonald, A., and Chowdhury, A. P3P Deployment on Websites. *Electronic Commerce Research and Applications*, Volume 7, Issue 3, Autumn 2008, Pages 274-293.

[7] Consumer Product Safety Commission. "Labeling Requirements for Toy and Game Advertisements." 2008. http://cpsc.gov/library/foia/foia08/brief/toygameads.pdf

[8] DeJoy, D.M., Cameron, K.A., and Della, L.J. Post-exposure evaluation of warning effectiveness: A review of field studies and population-based research. *The Handbook of Warnings.* 2006. *(35-48)*.

[9] Downs J.S., Loewenstein G., and Wisdom J. Strategies for Promoting Healthier Food Choices. *American Economic Review*. 2009, vol. 99, issue 2, pages 159-64

[10] Drichoutis AC, Lazaridis P, Nayga RM.2006. Consumers' use of nutritional labels: a review of research studies and issues. Acad Marketing Sci Rev, no. 9.

[11] The Energy Label. 2007. www.energyrating.gov.au

[12] European Union Commission Directive 98/11/EC "Energy Labeling." 1998. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1998:071:0001:0008:EN:PDF

[13] Food Standards Agency. "Signpost Labeling Research." 2005 http://www.food.gov.uk/foodlabelling/signposting/siognpostlabelresearch/

[14] Jensen, C. and Potts, C. Privacy policies as decision-making tools: an evaluation of online privacy notices. SIGCHI. 2004.

[15] Kelley, P., A. McDonald, R. Reeder, and L. Cranor. P3P Expandable Grids. Poster at Privacy MindSwap Carnegie Mellon University. 2007. http://cups.cs.cmu.edu/soups/2008/posters/kelley.pdf

[16] Kleimann Communication Group, Inc. Evolution of a Prototype Financial Privacy Notice. February 2006. Available: http://www ftc.gov/privacy/privacyinitiatives/ftcfinalreport060228.pdf

[17] Levy, A. and Hastak, M. Consumer Comprehension of Financial Privacy Notices. December 2008. Available: http://www ftc.gov/privacy/privacyinitiatives/Levy-Hastak-Report.pdf

[18] Maubach, N., Hoek J. "The Effect of Alternative Nutrition Information Formats on Consumers' Evaluations of a Children's Breakfast Cereal" Proceedings of the EParternships, Proof and Practice – International Nonprofit and Social Marketing Conference 2008.

[19] McDonald, A., Reeder, R.W., Kelley, P.G., and Cranor, L.F. A Comparison of Online Privacy Policies and Formats. Privacy Enhancing Technologies 2009.

[20] McDonald, A, and Cranor, L. The Cost of Reading Privacy Policies. Telecommunications Policy Research Conference, 2008.

[21] Privacy Leadership Initiative. Privacy Notices Research Final Results, November 2001, Available at: http://www.understandingprivacy.org/ content/library/datasum.pdf.

[22] Reeder, R.W. *Expandable Grids: A user interface visualization technique and a policy semantics to support fast, accurate security and privacy policy authoring.* PhD thesis, Carnegie Mellon. 2008. http://www robreeder.com/pubs/ReederThesis.pdf

[23] Reeder, R., Cranor, L., Kelley, P., and McDonald, A. A User Study of the Expandable Grid Applied to P3P Privacy Policy Visualization. *Workshop on Privacy in the Electronic Society.* 2008

[24] Seymore, J.D., Lazarus Yaroch, A., Serdula M., Blanck, H.M., and Khan, L.K. "Impact of nutrition environmental interventions on point-of-purchase behavior in adults a review." *Preventative Medicine* 2004. 29: S108-S136.

[25] The Center for Information Policy Leadership, H. . W. L. Multi-layered notices.

[26] Turow, J. Feldman, L., and Meltzer, K. Open to Exploitation: American Shoppers Online and Offline. The Annenberg Public Policy Center. 2005. http://www.annenbergpublicpolicycenter.org/NewsDetails.aspx?myId=31

[27] U.S. Food and Drug Administration. A Food Labeling Guide. Center for Food Safety & Applied Nutrition. 1999. http://vm.cfsan fda.gov/%7Edms/flg-toc html.

[28] U.S. Food and Drug Administration. "Guide to Nutrition Labeling and Education Act Requirements" 1994. http://www fda.gov/ora/inspect_ref/igs/nleatxt html

[29] U.S. Food and Drug Administration. "New OTC Drug Facts Label" *FDA Consumer Magazine.* 2002. http://www fda.gov/FDAC/features/2002/402_otc html

[30] W3C. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. http://www.w3.org/TR/P3P/

[31] W3C. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. http://www.w3.org/TR/P3P11/

[32] WELS Regulator (Australian Government). "WELS and Watermark." 2005. http://www.waterrating.gov.au/compliance html

An Empirical Study of How People Perceive Online Behavioral Advertising

Aleecia M. McDonald and Lorrie Faith Cranor

November 10, 2009

CMU-CyLab-09-015

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213

An Empirical Study of How People Perceive Online Behavioral Advertising

Aleecia M. McDonald[1] and Lorrie Faith Cranor[2]
Carnegie Mellon University
November 10, 2009

Abstract

We performed a series of in-depth qualitative interviews with 14 subjects who answered advertisements to participate in a university study about Internet advertising. Subjects were not informed this study had to do with behavioral advertising privacy, but raised privacy concerns on their own unprompted. We asked, "what are the best and worst things about Internet advertising?" and "what do you think about Internet advertising?" Participants held a wide range of views ranging from enthusiasm about ads that inform them of new products and discounts they would not otherwise know about, to resignation that ads are "a fact of life," to resentment of ads that they find "insulting." Many participants raised privacy issues in the first few minutes of discussion without any prompting about privacy. We discovered that many participants have a poor understanding of how Internet advertising works, do not understand the use of first-party cookies, let alone third-party cookies, did not realize that behavioral advertising already takes place, believe that their actions online are completely anonymous unless they are logged into a website, and believe that there are legal protections that prohibit companies from sharing information they collect online. We found that participants have substantial confusion about the results of the actions they take within their browsers, do not understand the technology they work with now, and clear cookies as much out of a notion of hygiene as for privacy. When we asked participants to read the NAI opt-out cookie description, only one understood the text. One participant expressed concern the NAI opt-out program was actually a scam to gather additional personal information. No participants had heard of opt-out cookies or flash cookies. We also found divergent views on what constitutes advertising. Industry self-regulation guidelines assume consumers can distinguish third-party widgets from first-party content, and further assume that consumers understand data flows to third-party advertisers. Instead, we find some people are not even aware of when they are being advertised to, let alone aware of what data is collected or how it is used.

1. Introduction

Behavioral advertising, also known as targeted advertising, is the practice of collecting data about an individual's online activities for use in selecting which advertisement to display. Third party cookies are one of several of the mechanisms to enable behavioral advertising: a central advertising network with ads across thousands of websites can set and read cookies, noting every

---

[1] Aleecia M. McDonald is a PhD candidate in the Engineering & Public Policy Department of Carnegie Mellon University. http://www.aleecia.com

[2] Lorrie Faith Cranor is an associate professor in the School of Computer Science and in the Engineering & Public Policy Department of Carnegie Mellon University. http://lorrie.cranor.org/

time a given user visits any of the sites in the network.[3] By correlating which sites an individual visits, advertisers can build profiles of likely characteristics and interests, and display advertisements to people most likely to purchase a given product or service. Targeted ads command a premium but also offer the potential for more cost-effective advertisements. While each advertisement costs slightly more, the specific ads go to fewer people than they would in a non-targeted campaign, and the hope is that a higher percentage of ad views will result in sales.

Behavioral advertising has received a lot of attention in the past few years. Questions about consumer's online privacy, how easily seemingly anonymous information can be re-identified,[4] and the legality of some behavioral advertising business practices[5] are at issue. The advertising industry[6] and their allies[7] favor the continuation of an "industry self-regulation" approach. The Federal Trade Commission has held numerous workshops and released guidelines for self-regulation,[8] and there are several legislative proposals at the Federal[9] and State[10] level. In 2008, TRUSTe commissioned a report on behavioral advertising, finding 57% of respondents are "not comfortable" with browsing history-based behavioral advertising, "even when that information

---

[3] Kristol, D., "HTTP Cookies: Standards, privacy, and politics," *ACM Transactions on Internet Technology* (TOIT) Volume 1 , Issue 2 (November 2001.) Pages 151 – 198. Available from: http://doi.acm.org/10.1145/502152.502153

[4] Ohm, P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, 57 UCLA L. Rev. ___ (forthcoming 2010). Available from: http://ssrn.com/abstract=1450006

[5] In particular: did NebuAd or their business partners violate wiretap and other laws? To date, the only settled case law for NebuAd pertains to jurisdiction. See Davis, W., *Online Media Daily*, "Judge Dismisses Case Against ISPs That Worked With Closed NebuAd," (October 12, 2009). Available from: http://www.mediapost.com/publications/?fa=Articles.showArticle&art_aid=115259

[6] AAAA, ANA, BBB, DMA, and IAB. "Self-Regulatory Program for Online Behavioral Advertising," (2009). Available from: http://www.iab.net/media/file/ven-principles-07-01-09.pdf

[7] Szoka, B. M. and Thierer, A. D., Targeted Online Advertising: What's the Harm And Where Are We Heading? (February 13, 2009). Progress & Freedom Foundation Progress on Point Paper, Vol. 16, No. 2, February 2009. Available at SSRN: http://ssrn.com/abstract=1348246

[8] Federal Trade Commission Staff Report, "Self-Regulatory Principles For Online Behavioral Advertising: Tracking, Targeting, and Technology" (February 2009). Available from: http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf

[9] Boortz, A. R. "New Federal Privacy Bill in the Works: Behavioral Advertising "Beneficial," But Must Be Done "Appropriately"" *AdLaw By Request* (August 12, 2009). Available from http://www.adlawbyrequest.com/2009/08/articles/legislation/new-federal-privacy-bill-in-the-works-behavioral-advertising-beneficial-but-must-be-done-appropriately/

[10] Arias, M. L. "Internet Law – Behavioral Advertising in the United States," *Internet Business Law Services* (June 30, 2009). Available from: http://www.ibls.com/internet_law_news_portal_view.aspx?s=latestnews&id=2237

cannot be tied to their names or any other personal information."[11] Several academic scholars have also investigated this area. Anton et al. studied privacy concerns in 2002 and again in 2008, and found that "individuals have become more concerned about personalization with regard to customized browsing experiences, monitored purchasing patterns, and targeted marketing and research" in 2008.[12] Gomez et al. estimated that Google Analytics tracks at least 329,330 unique domains, and found confusion in privacy policies containing "conflicting statements that third-party sharing is not allowed but third-party tracking and affiliate sharing are."[13] Most recently, Turow et al. conducted a representative sample of Americans and found 66% do not want behavioral advertising, with three quarters or more rejecting common behavioral advertising practices.[14] While the Turow work is valuable because it quantifies the percentage of Americans holding particular views, the standardized phone interview format meant they were unable to discover why people hold those views. In this paper, we overcome that limitation to investigate people's mental models of online advertising.

2. Methods

We performed a series of in-depth qualitative interviews with 14 subjects who answered advertisements to participate in a university study about Internet advertising. Subjects were not informed this study had to do with behavioral advertising privacy, but raised privacy concerns on their own, unprompted. We followed a modified mental models protocol of semi-structured interviews, using standard preliminary questions for all participants while also following up individually to gather participants' understanding of and reaction to behavioral advertising in particular.

Our study ran from September 28th through October 1, 2009 in Pittsburgh, PA. We recruited participants with a notice on the Center for Behavioral Decision Research website, which is run by Carnegie Mellon to notify the Pittsburgh community of research opportunities. Participants were compensated $10 for an hour of their time. Of our 14 subjects, 8 were male and 6 female. Half were age 21–29 and half were age 30–59. Participants had diverse professional backgrounds including health, architecture, photography, marketing, and information technology.

---

[11] TRUSTe, "2008 Study: Consumer Attitudes About Behavioral Targeting," (March 28, 2008). Available from: http://danskprivacynet.files.wordpress.com/2009/02/truste2008_tns_bt_study_summary1.pdf

[12] Antón, A. I., Earp, J. B., and Young, J. D. "How Internet Users' Privacy Concerns Have Evolved Since 2002," North Carolina State University Computer Science Technical Report # TR-2009-16 Submitted to *IEEE Security & Privacy* (July 29, 2009) Available from: http://theprivacyplace.org/blog/wp-content/uploads/2009/07/tr_2009_16.pdf

[13] Gomez, J., Pinnick, T., and Soltani, A. "KnowPrivacy," UC Berkeley School of Information Report 2009-037, (October 10, 2009). Available from http://www.escholarship.org/uc/item/9ss1m46b

[14] Turow, J., King, J., Hoofnagle, C., Bleakley, A., Hennessy, M. "Americans Reject Tailored Advertising and Three Activities that Enable It," (September 29, 2009). Available at SSRN: http://ssrn.com/abstract=1478214

Because our sample size is small, we do not yet know how well our results generalize to the full United States population. What we can offer are insights to *why* people hold the views they hold, and their motivations behind the actions they take online. We were able to follow up on participants' comments and engage them in dialog to elicit their views, rather than just ask fixed questions. We are still analyzing our rich qualitative data set. We describe preliminary unpublished findings below and will update the CUPS website with final publications as they become available (http://cups.cs.cmu.edu). We also expect to conduct a follow-up survey to determine the prevalence of the views held by our interview participants in the larger population.

3. Consumer Expectations

Much of the current self-regulation approach to online privacy is grounded in the Fair Information Principle of notice. Notice, by its nature, requires communication. As Morgan et al. wrote, "An effective communication must focus on the things that people need to know but do not already. This seemingly simple norm is violated remarkably often in risk communication."[15] To follow this guidance we must find out what people already know about online privacy, what they do not, and what information they require to make decisions based on privacy policies. We need to investigate people's pre-existing mental models to see what beliefs they hold about online privacy risks, remedies, and mitigation. *Mental models* are the beliefs people hold about how a system works, interacts, or behaves. Incorrect views may form a view of the world that leads to poor evaluation of options and ultimately to bad decisions. For example, if people hold the mental model that any company with a privacy policy is bound by law not to release data to third parties, and if that is the only threat that worries them, why would people bother to read the policy? The existence of a link to a privacy policy would seem sufficient in and of itself.[16] Our research contributes to understanding consumer expectations.

3.1. Limited Knowledge of Types of Internet Advertising

We began all interviews by asking the open-ended question "What is Internet advertising?" The answer given most immediately was "pop ups," with all but four participants mentioning pop ups. This is an intriguing response since modern browsers block pop ups by default, and indeed, participants discussed their interactions with pop up blocking. However, participants call many things "pop ups," including interstitial and hover ads. For one participant the association is so strong that she talks about all ads "popping up" on her screen, even while clearly giving examples of banner ads. For her, all ads are pop ups. Banner ads are tied with pop ups for the most prevalent response when we asked participants, "What is Internet advertising?" Banner ads were not usually mentioned first (as pop ups were) and were rarely mentioned by name. However, participants were quite capable of describing banner ads even without the vocabulary to name them. Over a third of respondents mentioned spam as a form of Internet advertising. We found it surprising that a few participants mentioned Google AdSense by name. While Google's brand is well known, we had not expected AdSense to reach beyond the advertising industry. Instead, several participants had either used AdSense to try to monetize their own blogs or knew friends who had used AdSense.

---

[15] Morgan, M. G., Fischhoff, B., Bostrom, A., and Atman, C. J. *Risk Communication: A Mental Models Approach*. (Cambridge: Cambridge University Press, 2002).

[16] Research shows that people do, in fact, believe the words "privacy policy" mean they are protected by law. See Hoofnagle, C. and King, J. "What Californians Understand About Privacy Online," (September 3, 2008) <http://ssrn.com/abstract=1262130> Accessed 11 September 2008.

Some participants gave characteristics of ads, rather than examples of ads. Less than half mentioned video and audio ads, usually while expressing displeasure at ads they find distracting. Participants also mentioned difficulty closing ads, and in particular complained that pop ups do not necessarily have a close button in the same place (here, again, we see confusion between true pop up ads and similar forms of advertising like interstitials.) The following concepts were mentioned by one participant each: viruses, hijacked links within articles, a constant stream of pop ups, and behavioral advertising (not mentioned my name, but described by the participant as a way to "exploit a person's history"). The other thirteen respondents did not mention or allude to behavioral advertising at all when asked to define Internet advertising. Overall, the picture that emerges includes only a general familiarity with advertising, and some user frustration with specific advertising methods and modalities.

3.1.1. Mixed Identification of Internet Advertising

Contextual search advertisements are well understood. All participants said Google is their search engine of choice. When asked if Google has ads, all participants answered correctly. Participants knew there are ads down the right hand side, that "sponsored" links frequently appear at the top of results pages, and that these links are also advertisements. They were all able to recall these details of Google's advertisements with no prompting beyond asking if there are ads and where they are located.

We asked how advertising on Google works. All participants understood that advertisers pay Google to run ads. Participants were less clear on the mechanics of payment. Some expected Google charges for all ads displayed, and some thought Google only charges for ads when people click on them. No one thought anything that was impossible or has not occurred at one time. All told, this is a surprisingly sophisticated understanding of Google's contextual advertising during search tasks.

In contrast, when we gave participants a printout of a webpage from the *New York Times* and asked them to identify the advertisements, answers varied widely. On the low end, participants looked at the graphics only, and discounted anything that came from the *Times* itself (e.g. home delivery and subscriptions.) At the other extreme, one participant counted every single item on the page as an advertisement, including hyperlinks in the article to other *Times* articles — and even the article itself. She reasoned the article text was likely a press release and therefore also an advertisement. Some of the differences in answers stemmed from participants skipping over parts of the page, discounting anything other than an image as a possible advertisement. Even while asking specifically about ads, a few people suffered from "ad blindness" and simply did not notice smaller ads that were in unexpected places (e.g. flush against the masthead instead of the right-hand column.) But much of the difference was definitional. While they did not phrase it this way, some participants saw advertisement as strictly a third party endeavor. Anything from the *Times* itself was therefore not an ad. Some participants also discounted all text as a potential source of advertisement.

Clearly participants do understand that text can be advertising, or they would not have been able to answer correctly about Google search ads. Why do some people then discount text as a source of advertisement on the *Times*? We have two hypotheses. First, it could be that Google is uncommonly good at communicating with their users. Ads are always in the same place, the "sponsored" label and yellow background are understood, and the right side is the place people expect to find ads. Second, it could be that people's pre-existing mental models of print media come into play with the *Times*. People have learned with experience that ads in printed newspapers and magazines are usually graphics. To look for text ads on the *Times* people must first unlearn what they already knew, where Google was a blank slate with no direct offline

analog. Or it may be a combination of factors that people react to in different ways, which might account for why participants reacted uniformly to Google but with great variance to *Times* advertisements.

3.1.2. Inability to Distinguish Widgets

Regardless of the cause, what the *Times* results mean is that even absent any confusion over technology, participants had different mental models of advertising. We find participants have a wide range of expectations on the simple question of what is or is not an advertisement. Industry guidelines assume people can distinguish third party widgets from first party content and assume that people understand that data flows differently to third party advertisers. Therefore they treat third party widget providers as first party data collectors, subject to fewer guidelines[17]:

> *In addition, in certain situations where it is clear that the consumer is interacting with a portion of a Web site that is not an advertisement and is being operated by a different entity than the owner of the Web site, the different entity would not be a Third Party for purposes of the Principles, because the consumer would reasonably understand the nature of the direct interaction with that entity. The situation where this occurs most frequently today is where an entity through a "widget" or "video player" enables content on a Web site and it is clear that such content is not an advertisement and that portion of the Web site is provided by the other entity and not the First Party Web site. The other entity (e.g., the "widget" or "video player") is directly interacting with the consumer and, from the consumer's perspective, acting as a First Party. Thus, it is unnecessary to apply to these activities the Principles governing data collection and use by Third Parties with which the consumer is not directly interacting.*

Instead, we find some people are not even aware of when they are being advertised to, never mind being aware of what data is collected or how it is used. It appears that self-regulatory guidelines may assume an unrealistic level of media literacy on the part of Internet users.

3.2. Misperceptions of First Party Cookies

We asked several questions regarding cookies. All participants had heard of cookies before. However, there was widespread confusion about what cookies are or how they are used. When asked, "What is a cookie?" nearly a third of participants replied immediately that they were not sure. Slightly more than a third of participants gave an answer that was at least partially correct without also saying something factually incorrect. Only one person articulated that a cookie can contain a unique identifier. We asked follow up questions of "are there ways cookies can help you?" and "are there ways cookies do not help you?"

More than half of participants confused cookies with browser history, including one participant who believed the backward and forward arrows in a web browser depend on cookies. Participants did not understand that browser history is stored independently of cookies. One participant told us cookies contain a "history of websites" visited and that if he deletes cookies, then "hyperlinks in different colors goes away, that's what it does. It clears the navigation history." He related how

---

[17] AAAA, ANA, BBB, DMA, and IAB. "Self-Regulatory Program for Online Behavioral Advertising," (2009) page 24. Available from: http://www.iab.net/media/file/ven-principles-07-01-09.pdf

when he was a child living at home with his mother, he lost his computer privileges because she could see where he had been based on the color of web links, which he blamed on cookies. Cookies mean "someone else can follow your previous path, and can see what you've read before, but that means they can get into your [computer]." More exploration revealed that in his view, cookies were only an issue on computers where he shared a single account with multiple people as he had in his mother's home. At work, where he signed into his computer account with his own password, he believed cookies could not provide details of his browsing history because he was the only one with access to the account. Notice the confusion around password-protected accounts and privacy protections: several participants had confusion in similar areas. From follow up questions we learned that participants clear cookies and browser history at the same time, so they do not distinguish the effects. Browser user interfaces may contribute greatly to confusion as we describe later.

While participants generally did not understand what cookies store locally rather than what cookies provide a key to in a remote database, perhaps it is more important that they understand the effects of cookies rather than their mechanism. Over a third of participants said that cookies can be related to saving passwords, though during follow-up questions they did not know if their web browsers were storing the passwords or if cookies were involved. Similarly, three participants answered that cookies allow them to remain logged in to websites without retyping a password. Three participants believed cookies store their preferences for websites, including details like preferred colors and placement of site elements.

Most people believed something that was not correct about cookies. A small number of people mistakenly believed that cookies store far more than they do, such as believing cookies record all actions they take online. A couple of people thought cookies store personally identifiable or sensitive data like social security numbers, credit card numbers, and IP addresses. Three participants believe cookies are a form of malware (virus, spyware, or spam.) Several people described warnings for self-signed certificates and mistakenly believed that those warnings pertained to accepting or rejecting cookies. A few believe cookies make websites display more quickly. Again, if we consider this in the context of what cookies enable, some of these answers are more reasonable, but these answers show people do not understand what risks and benefits cookies pose.

Only three of our fourteen participants said that cookies are related to personalized advertisement. Notice three very different perspectives ranging from outright rejection to seeing benefits but finding harms outweigh them to support that is conditioned on the mistaken view that current practices are illegal.

- One participant said she has no choices about cookies, because if you "say no then you don't get to go to the site. That's not much of an option." She could not think of any way cookies help her. For ways cookies do not help her, she said sites use cookies to personalize, and that "could mean more personalized advertising. It makes me feel like they expect me to be gullible."
- A second said cookies are things "that programs use to gather information about sites [visited], functionality, and demographics for an ad." Cookies "factor in" when advertisers "decide if it's worth the charge they pay to advertise to that person at that time." He said that "if asked for information [people] would say no," and believes he has "no choices" about cookies. He said that cookies are good when "a set pattern of behaviors, sites, topics, or hobbies" can give "information on products and services that are more interesting," but "some [cookies] are used negatively to exploit a person's history," and "cookies open pools of information one might prefer to stay private." He

was concerned that IP addresses identify "a processor or individual computer," and that the type of information collected for advertising could also be used to guess people's passwords. He was skeptical about how well behavioral advertising works, saying "maybe you're buying a gift" but would continue to see ads about that purchase in the future. "Patterns may be a coincidence," and advertisers "may put you in more of a box than you are in." Drawing an analogy to shopping offline, he said "you may be shopping in a public place but there is a privacy issue" with companies "knowing where you spend money and time." Even with a computer collecting and storing the data, there still must be a "person manipulating and interpreting that," and that invites "bias" because "some manipulate facts to serve a goal."

- A third participant said advertisers use cookies to "find out as much as [advertisers] can without asking for names," to gain an "idea of what sort of person" you are. He mentioned ISPs trying to "find ways to catalog this wealth of information," to pair ads to an audience. He described this practice as a "smart thing" and "reasonable." He then volunteered that he believes ISPs are constrained by law not to share information. When we asked what the law entails, he answered he was not sure and perhaps constraints were not from law but that there would be a "public uproar" and a "bad image" for any company sharing even anonymous customer data. He made the analogy to phone service where recording conversations can be illegal, and said there are "certain cultural norms and expectations" to privacy.

3.3. Unclear on Clearing Cookies

Nine of our 14 participants self-reported that they clear cookies. Only one of those nine said they clear cookies on their computer for privacy. Another three clear cookies on shared machines out of privacy concerns. People told us they clear cookies for the following reasons:

- To delete history
- To avoid malware (viruses, spyware)
- To reduce clutter
- To save space
- Out of habit
- For "hygiene"

Participants have a vague notion that too many cookies are bad, do not know why, and are not sure why they should delete cookies.

3.4. Ignorance of Cookie Variants

We asked participants if they had every heard of several technologies and if so, to define them. For anything they had not heard of we asked them to guess what the phrase might mean. We asked about:

- Session cookies
- Third party cookies
- Flash cookies

A few people had heard of session cookies or third party cookies. Those few who had heard of them were able to give mostly accurate answers. No one had heard of flash cookies before, with participants guessing things like they are cookies that "appear in a flash and are gone."

4. Risks, Concerns, and Benefits

We asked participants for their views of "what are the best and worst things about Internet advertising?" and, "what do you think about Internet advertising?" Overall, participants held a wide range of views. Two participants responded enthusiastically not just to the idea of advertising-supported content, but to the ads themselves, which inform them of new products and discounts they would not otherwise know about. Two people were against online advertising, finding the content "insulting" and an attempt to reach "the vulnerable." The remaining ten participants were neutral to resigned. Ads are simply "a fact of life," multiple participants said.

4.1. Perception 1: Internet Advertising is Necessary

Participants named several benefits from Internet advertising such as:

- "Necessary" for the Internet to function and to enable free content
- "Good if you can control [them]"
- "Great" or "beneficial" because ads are a source of information
- "Can totally ignore" ads, unlike television or billboards
- "Short and sweet" ads
- Ads tend to be "related to [the] page"
- Ads are "more of what I want" and "not random"

With the notable exception of being able to ignore ads, this list is very similar to the benefits touted by advertisers themselves. Participants generally feel advertisements are annoying, but also see advertisements as an essential element of online life. They understand advertisements as the payment for otherwise "free" online content. A minority of participants volunteered a preference for relevant ads. However, this does not mean they understand or like data collection for behavioral advertising. When participants ask for more relevant advertisements, they almost always express a preference for contextual, not behavioral, targeting. "Relevance" means ads related to the website they are visiting, rather than related to them individually.

4.2. Perception 2: Internet Advertising is Annoying

The single most frequent response volunteered was that Internet advertising is "annoying," a word used by nearly half of all participants. Participants mentioned harms from Internet advertising such as:

- "Annoying"
- "Insulting"
- "Distracting"
- "Crude graphics"
- "Clogs up Internet access" / "Slower"
- Unrelated / Off topic / Awkward mismatches
- "Opens pools of information one might prefer to stay private"
- "Not regulated"

Participants complained about being distracted by ads while trying to work or perform other primary tasks, which made pop ups and streams of ads particularly unpopular. Participants mentioned flashy colors, over-reliance on primary colors, movement, and sound as distracting elements.

4.3. Perception 3: Internet Advertising is Concerning

One straight male participant complained that he kept "getting male companion [advertisements]." He explained that this "mismatch is awkward sometimes" because it "makes you feel targeted as someone you're not." A second participant explicitly raised behavioral advertising and "threats to privacy." A third participant discussed the "two way communication" of the web, and volunteered that a "privacy issue comes up" due to "creepier" advertisements "based on personal messages and keywords." A fourth participant called for a complete "reboot" of the Internet. A fifth participant worried about "obscene" and "inappropriate" ads, particularly as she is considering starting a family. She worried about how to keep children safe online. A sixth participant raised lack of regulation. She mentioned "horror stories" of friends who signed up to get free iPods but had to submit their friends' names first, and then never even received the promised iPods. She was most disturbed about an ad for a prescription drug to grow longer eyelashes, which was advertised just like mascara but without discussion of potential medical side as other media require. She said with TV it "seems more obvious what you can trust" but for Internet advertising a "well-designed website can be a scam." She concluded that regulation for online advertisements is necessary and "all of it needs some kind of change."

Four things were striking about these opening conversations. First, discussion of "relevant" ads ran the gamut from support to deep concerns about privacy. As the interviews continued the diversity of opinions became even more marked and we learned how little people understand of current practices. Second, participants were largely pragmatic about advertising. Even when they had scathing remarks about bad experiences, on the whole they understand and accept the model that advertising supports content. Their frustrations are generally not due to the existence of advertising, but rather to the specific practices. Third, participants expressed real anger and frustration about advertising tactics they see, even when they do not understand the data being amassed about their online activities that they do not see. Finally, all of the issues raised above were volunteered, not prompted, after very open-ended questions at the start of the interviews. Participants' concerns about advertising practices, content, lack of regulation, behavioral targeting, and privacy surfaced in the first few minutes of discussion. These issues are central to how participants perceive online advertising.

5. Mechanisms of Consumer Privacy Protection

From what we have observed to date, it appears behavioral advertising violates consumer expectations and is understood as a source of privacy harm. While we do not attempt a full analysis of possible policy responses here, we note several things. First and foremost, consumers cannot protect themselves from risks they do not understand. One younger participant said in frustration that she did not learn about how to protect her online privacy in school, she was just taught typing. We believe there is a serious need not just for improved notice of practices, but for the education requisite to understand disclosures. Most non-regulatory approaches require consumers to understand tradeoffs and know enough to take whatever actions will enable their privacy preferences. At the current moment that seems unrealistic, but the outlook could improve in the future. Below, we offer some preliminary findings about the industry self-regulation mechanism of NAI opt out cookies, and some observations about web browsers' roles in cookie management.

5.1. Consumers Do Not Understand Opt Out Cookies

None of our fourteen participants had heard of cookie-based methods for opting out of tracking cookies, including TACO[18] and NAI opt out cookies.[19] At the end of the protocol, we showed four participants a text description of NAI opt out cookies from the NAI opt out website.[20]

All four participants understood they would continue to see at least some online advertisements. However, there is substantial confusion about what the NAI opt out does. The text does not explain that companies may choose to continue all data collection and profiling, and that in some cases the only thing that changes is the type of ads displayed.[21] One participant understood this but the other three did not.

- The first participant believed the NAI opt out "sets your computer or ethernet so information doesn't get sent." She still expected to see ads, but now the ads would be "random." She said it might "sound old fashioned" but in a choice between "convenience and privacy, I'm going to pick privacy." She was afraid that by clicking to opt out "all these people get your information" and therefore "this could be a phishing expedition."
- A second participant began his comments by saying "Where do I click? I want this!" He believed the NAI opt out to be an "opt out tool so users opt out of being tracked." He thought "the ads are still there, they just get no data."
- A third participant believed the purpose of the NAI opt out text was "reducing the amount of online advertising you receive." He understood data collection was also involved, but not how, just "some sort of control over what companies use that information." He would choose to opt out of some companies, "ones I thought the information they would seek would be too personal to share with a group."
- Our final participant did understand the NAI text better. At first he said by way of example that it means if you use GMail the opt out cookie means "stop reading my email and tailoring ads." However, he later clarified "What you search is Google property, it's theirs. They're going to profile you but not show you that they are."

---

[18] Targeted Advertising Cookie Opt-Out (TACO) is a plugin for the Firefox browser that stores persistent opt out cookies, available from: https://addons.mozilla.org/en-US/firefox/addon/11073

[19] The Network Advertising Initiative (NAI) offers non-persistent opt out cookies for all browsers, available from: http://www.networkadvertising.org/managing/opt_out.asp

[20] Our study used printed materials so we did not test the NAI video, which may communicate more clearly. The degree to which the video's clarity is important hinges on how visitors engage the site. The NAI may be able to provide information about what percentage of website visitors watch the video to completion, but four calls to NAI asking to speak about their opt out cookies went unreturned.

[21] Anderson, Stacy. "House Subcommittees Hold Joint Hearing On Behavioral Advertising," *Security, Privacy and the Law* (July 2009.) Available from: http://www.securityprivacyandthelaw.com/2009/07/articles/recent-legislation-1/house-subcommittees-hold-joint-hearing-on-behavioral-advertising/ Original testimony available from: http://www.youtube.com/watch?v=-Wklp2qdbmw

5.2. Web Browsers May Promote Consumer Confusion

While we did not study web browser interactions specifically, participants explained ways they use their web browsers to interact with cookies. In the section "Misperceptions of First Party Cookies," we documented consumer confusion between cookies and browser history. One component of this confusion is temporal: participants reported they delete cookies and clear history at the same time, which leads them to misattribute properties of browser history to cookies. The reason participants clear cookies and history together likely stems from the way they are swirled together in the user interfaces of web browsers. For example, as shown in Figure 1, Firefox presents choices about cookies, history, and bookmarks on the same tab. There is no visual hint that these three topics are distinct. To the contrary, cookies are in the middle of options for history, which serves to convey history and cookies are related. Moreover, Firefox does not expose any cookie options unless users know to change a setting from "Remember history" to "Use custom settings for history." Anyone looking through preference tabs for cookies will not find them in the default configuration.

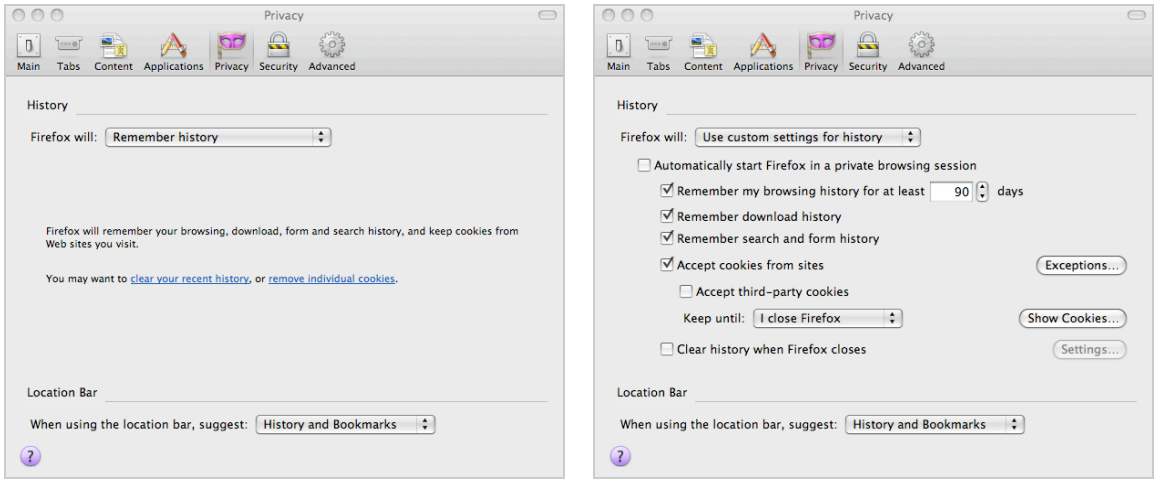


Figure 1: Firefox's Macintosh user interface mixes cookies, history, and bookmarks

Mixing cookies, history, and bookmarks is not the only area where web browsers interfaces contribute to lack of understanding of Internet privacy issues. As another example, web browsers give no notice of or access to Flash cookies. Even technologically sophisticated users are unfamiliar with Flash cookies and how they can "respawn" deleted cookies.[22] As another example, Internet Explorer implements P3P support, but information about P3P is buried in the user interface, so much so that a study of online trust markers found none of the participants were familiar with the P3P icon.[23] The Internet Explorer P3P implementation works well in that it does not require user intervention. Based on default settings, users do not accept any third party cookie that does not have an associated P3P policy with an opt out. In this way browsers can provide an enforcement mechanism that may be stronger and faster to take effect than any regulations.

---

[22] Soltani, A., Canty, S., Mayo, Q., Thomas, F., and Hoofnagle, C. "Flash Cookies and Privacy," (August 10, 2009). Available at SSRN: http://ssrn.com/abstract=1446862

[23] Jenson, C., Potts, C., and Jenson, C. "Privacy practices of Internet users: Self-reports versus observed behavior," *International Journal of Human-Computer Studies*, Volume 63, Issues 1-2, (July 2005) Pages 203-227.

However, as the early history of cookies themselves and the current example of Flash cookies and P3P amply demonstrate, just because browsers *can* provide user control does not mean they *will*. Cookies were introduced fifteen years ago, yet we observed most people do not understand even first party cookies. Browsers can be an important part of user empowerment but as the lack of cookie knowledge illustrates, informed privacy decision making is not something the free market is solving.

6. Observations

Netscape introduced cookies fifteen years ago, yet today approximately two thirds of our respondents were unable to explain what cookies do without volunteering incorrect information. Half of participants confuse cookies with browser history, and that confusion may be promoted by web browsers' user interfaces. Participants had no understanding of flash cookies or that flash cookies can respawn deleted cookies across domains.

None of our participants were familiar with NAI opt out cookies. Participants who incorrectly believed NAI opt outs mean they are no longer subject to profiling were very enthusiastic supporters. Based on NAI's text, participants had a difficult time understanding what the NAI opt out cookies do.

Consumers have a very clear understanding of when and where Google search displays advertisements. However, consumers do not understand which parts of the *New York Times* website are advertisements. They lack the knowledge to distinguish widgets from first party content. Consequently, it is overly optimistic to believe consumers know their data flows to widget providers as a first party.

One of the questions posed by the advertising industry is "where's the harm" in behavioral advertising, with a suggestion that a formal benefit cost analysis should occur before regulation. This question seems to ignore privacy loss as a distinct harm. In contrast, our participants spoke frequently about their privacy concerns. One technically savvy participant even described withdrawing from online life as a result of privacy concerns. We refer readers to our prior research where we estimated the high value of people's time if they were to actually read privacy policies. We found "it appears the balance between the costs borne by Internet users versus the benefits of targeted ads for industry is out of kilter," and "suggest that any such cost-benefit analysis should include the value of time for reading privacy policies."[24]

[24] McDonald, A. M. and Cranor, L. F. "The Cost of Reading Privacy Policies," *I/S: A Journal of Law and Policy for the Information Society* (2008). Forthcoming from http://www.is-journal.org/ and available from http://cups.cs.cmu.edu