The Privacy Policy Paradox

Rena Coen
rena@ischool.berkeley.edu

Jennifer King
Jenking@ischool.berkeley.edu

Richmond Wong
richmond@ischool.berkeley.edu

UC Berkeley School of Information
102 South Hall
Berkeley, CA 94720-4600

ABSTRACT

There have been multiple studies exploring the content and efficacy of privacy policies. However, to date no one has examined them from the angle we are proposing for this study: to determine whether the presence of a privacy policy link on a website has any significant influence on one's willingness to disclose personal information. Our study intends to examine whether the link itself acts as a trust heuristic without testing a respondent's comprehension or opinion about the privacy policy itself. In this paper, we discuss a study currently in progress to examine this question.

1. INTRODUCTION

Everyday, millions of people in the U.S. (and globally) make the choice to disclose personal information online. These choices can have implications for their personal privacy, depending upon the type and sensitivity of information disclosed and to whom they disclose it. A casual decision to disclose sensitive health data to a mobile app, for instance, could have negative consequences down the road if the app developer chooses to disclose identifiable information about his users to third parties, such as health insurance companies.

Today, these choices are guided by a legal framework called "notice and consent." Notice and consent assumes that Internet users are rational actors who are capable of perusing legal documents to evaluate the privacy practices of individual websites and mobile applications prior to use or download. Subsequent use/download stands in as *de facto* consent of the site's/app's policies and terms of use. In California, state law (the California Online Privacy Protection Act of 2003, or Cal-OPPA) requires that any website or mobile application with users located in-state which collects personal information must "conspicuously post a privacy policy on the site and to comply with its policy."[1] Given California's size and political sway, this law essentially forces the vast majority of websites operating in the U.S. (and potentially abroad) to comply with Cal-OPPA, especially given the absence of Federal legislation in this area (the Federal Trade Commission, for example, can file suit if a company violates its stated privacy policy, but cannot force a company to post one).

As a result, links to privacy policies abound on most commercial and popular U.S. websites. The unofficial standard location for these links is in a website's footer. Multiple studies have confirmed what most of us know instinctively—these policies are rarely read [2, 3]; to read the privacy policy of every site you visit could take up an unreasonable portion of your life [4]; they are often written in language far beyond what most people can process. In short, they are legal documents written for other lawyers, not for end users. Further, a 2009 survey by co-author King and others found that the majority of the respondents believed that privacy policies were affirmative statements of rights, which is false.[5] Privacy policies simply state what personal information the website or app operator collects and their intended usage of it (e.g., sharing with third parties). Due to this, a privacy policy can include terms that are in fact privacy infringing.

2. OUR STUDY

There have been multiple studies exploring the content and efficacy of privacy policies. However, to date no one has examined them from the angle we are proposing for this study: to determine whether the presence of a privacy policy link on a website has any significant influence on one's willingness to disclose personal information. In other words, does a privacy policy link itself act as a trust heuristic?

We intend to examine this question without testing a respondent's comprehension or opinion about the privacy policy itself. However, we presume that in most cases, users rarely seek out the link, let alone click to read the policy it links to, unless provoked by a disclosure environment that raises their concern about how their information will be used. Thus, our study design seeks to ask respondents questions designed to be maximally intrusive while violating a reasonable sense of context.

2.1 Proposed Study Details

We are conducting a controlled experiment wherein respondents are asked to disclose highly personal information through a staged online service sign-up form.[1] We are in the process of piloting the full study and initial findings will be presented at the workshop. The proposed study design is as follows: respondents will be randomly assigned to one of two conditions: having a privacy

Symposium on Usable Privacy and Security (SOUPS) 2016, June 22-24, 2016, Denver, Colorado.

[1] This study employs deception and has been approved by UC Berkeley's Committee to Protect Human Subjects.

policy link displayed or no privacy policy link. Our primary goal is to gauge the amount of personal information participants are willing to reveal between the control and experimental conditions.

Respondents will be assigned to one of three service contexts: a health/fitness context, a community review site context, and an educational/self-help context. The context will only be described in the introduction to the task and by the fictional name of each service. The visual presentation of each context version will be identical and will employ a neutral color palate and design. No other privacy or trust indicators will be present in the design beyond the privacy policy link (if assigned to that condition).

In addition, within each context, respondents will be assigned to a condition with no prompt (control), a (negative) privacy-heightening prompt, or a (positive) privacy-assuring prompt. The intent is to gauge whether attempting to heighten a respondent's privacy awareness has any effect on their willingness to disclose. The control condition provides simple directions:

> "We'd like your feedback on the sign-up process for a new [CONTEXT] app. You will be testing the portion of the sign-up process that immediately follows user account creation. Typically, you would sign up for this service using an email address and a password you create. (You will not be asked for your email address or to create an actual account during this process.)"

The privacy primed directions include an explicit suggestion appended to the end of the directions, designed to either trigger concern about one's information privacy (negatively priming for privacy), or to provide reassurance about one's information privacy (positively priming for privacy). Examples of a negative priming statement might include "by signing up for this service, you consent to the disclosure of your personal information" or "we may share your information with our partners." Examples of a positive priming statement might include "we will not disclose your personal information without your consent" or "your information will not be shared."

During the "sign-up process," respondents will be asked to answer a number of highly personal questions pertaining to their mental and physical health, sexual status and sexual practices, as well as demographics, educational attainment, domestic habits, employment status, and income. Within each theme are questions designed to be maximally personally intrusive. The next pilot test we will perform is an assessment of the 14 questions we have identified as intrusive. We will launch a short survey on Mechanical Turk in which we ask participants to rank each question on a scale of 0-10 based on how personally intrusive they find the question (independent of any specific context). Our goal is to yield a core set of intrusive questions, validated using correlation analysis.

After completing the "sign-up process," respondents are taken to a debrief and consent screen informing them of the intent of the study and asking for their informed participation. If they refuse, their answers will not be included in our final dataset. If they accept, participants are then taken to a post-test survey soliciting responses about the intrusiveness of the sign-up process, their level of comfort with the types of questions asked and their appropriateness (given the context assigned). They are also asked several questions about their experience with and perception of privacy policies. We will compare answers between those in the link vs. no link conditions; in instances where participants were shown the link, we will track whether or not they clicked on it to view the privacy policy.

The control and experimental groups will be compared on the basis of rates of disclosure as well as on responses to the post-test survey. We expect to test this study design on Mechanical Turk with approximately 1,500 subjects, in which we will have a minimum of 50 participants in each of our 18 experimental conditions. Our goals are to determine whether disclosure rates are constant across contexts or if there are variations; if priming in both positive and negative directions has any effect on disclosure, and ultimately to determine whether we should reject any of the contexts or priming statements in our final study.

2.2 Conceptual Pilot Study

In December of 2014, we ran a conceptual pilot for this study (N=236) on Mechanical Turk as a component of a UC Berkeley graduate course. The format and content of the pilot were similar to what is included in the current proposed study design (absent a privacy priming condition). We used a single context: the sign-up process for a stress mitigation application, the Calm Coach. Participants were asked to provide personal information about their: demographics; work and finances; physical, medical, and psychological health; lifestyle; and sexual activities. We randomly assigned respondents to one of two conditions: one in which a privacy policy link was visible at the bottom of the screen, and a control in which no link was present. We tracked how many participants actually clicked on the link; no participants did.

It appears that our efforts to make the "Calm Coach" seem like a legitimate context actually worked against us—in fact, based on our results, we appeared *too* legitimate. Most respondents answered all of our questions, with only about 2-3% choosing "prefer not to say" across the entire survey (per question, not by respondent). When asked, "did we ask any questions that seemed too personal or too sensitive," 85% of the respondents told us no. In our free text responses, many people said that the questions seemed legitimate for the type of service we were offering. In terms of comfort level, the questions we asked about sexual behaviors were considered to be the most sensitive, followed by mental health and income. Even so, the mean ratings for these questions skewed slightly from neutral to the more comfortable end of the spectrum. Our findings indicated that we should proceed with a context that could be less easily tied to the invasive questions that we ask, and a less reassuring or more neutral user interface.

As part of that pilot, we also asked "What (if anything) made you comfortable sharing your personal information with this website? 51% of the respondents selected "I was comfortable because this survey was offered on Mechanical Turk," while others commented in open text comments that they trusted this survey despite its personal nature because it originated at "berkeley.qualtrics.com," noting that the Berkeley name in the URL gave them reassurance.

Based on this feedback, we felt that a clear identification with Berkeley prior to taking the survey confounded our data collection process, and thus our final study will originate from Qualtrics.com instead of the Berkeley domain. Because we are attempting to control for factors that influence one's trust in a website when someone makes a decision to disclose, we think the only way to mitigate confounding the testing of our hypotheses is to reveal our affiliation and disclose our intent after the conclusion of the survey. The issues of credibility provided by hosting the survey on Mechanical Turk are a factor we feel we can grapple with both through the participant qualification process (allowing less experienced workers to participate) and also through some of the academic studies that have been conducted examining the attributes and representativeness of Mechanical Turk workers.

2.3 Hypotheses & Dependent/Independent Variables

The dependent variables we are examining are: disclosure rate (as measured by the number of questions for which respondents provide an answer other than "decline to state"), and their self-reported level of comfort with the questions asked. The independent variables have identified are: the absence/presence of the privacy policy link, service context, priming condition, and self-reported knowledge and behaviors with privacy policies. Our hypotheses are as follows:

H1: Across all conditions, the presence of the privacy policy link will have a positive effect on one's disclosure rate.

H2: In a privacy-heightened (negative) priming condition, we hypothesize that the presence of a privacy policy link will have a positive effect on disclosure rate as compared to the control condition (no privacy prime).

H3a: In a privacy-assuring (positive) priming condition, we hypothesize that the disclosure rate will be higher than compared to the control (no privacy prime).

H3b: In a privacy-assuring (positive) priming condition, we hypothesize that the disclosure rate will be higher than compared to the negative priming condition.

H4: The context a participant is assigned to may have a modest interaction effect (positive or negative) on a participant's disclosure rate.

3. IMPLICATIONS FOR PRIVACY INDICATORS

Privacy policies provide the scaffolding for the online consent framework, yet as discussed earlier, they are typically unread and do not provide any guarantee of information privacy. We are curious to explore to what extent—if any—they have evolved to become a *privacy heuristic*: merely a visual cue many of us expect to see on commercial websites, but not one that most of us ever bother to investigate.

We would not be surprised to find that our findings end up cutting both ways: that the absence of a link generally goes unnoticed, but that even under our most privacy intrusive conditions that participants still do not seek out the details of a privacy policy. Absent flaws in our study design, we think this would provide tangible justification for more pressure on regulators to move beyond the notice and consent framework as an acceptable means for informing consumers.

However, it is also possible that people use the existence of a privacy policy link as a *heuristic,* and are more likely to disclose personal information when a privacy policy link exists, even if they never click on it. This may be even more problematic, as it may indicate that people believe a privacy policy provides certain protections regardless of its actual content. This finding would not only question the efficacy of privacy policies and how they are displayed, but potentially question whether they actually cause more harm than good.

4. ACKNOWLEDGMENTS

We thank Heather Patterson for her feedback and Steve Weber for his initial support. This project is generously supported by a gift from the Intel Corporation.

5. REFERENCES

[1] California Online Privacy Protection Act of 2003. Full text available at: https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=BPC&division=8.&title=&part=&chapter=22.&article=

[2] Jensen, C., Potts, C., & Jensen, C. (2005). Privacy practices of Internet users: self-reports versus observed behavior. *International Journal of Human-Computer Studies*, *63*(1), 212.

[3] Milne, G. R. and Culnan, M. J. (2004), Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal* of *Interactive Marketing*, 18(3), 15–29.

[4] McDonald, A. M., & Cranor, L. F. (2008).. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society 2008 Privacy Year in Review Issue*, *4*, 543.

[5] Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., & Hennessy, M. (2009). Americans reject tailored advertising and three activities that enable it. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214.