Investigating the Interplay Between Searchers' Privacy Concerns and Their Search Behavior

Steven Zimmerman, Alistair Thorpe, Chris Fox, Udo Kruschwitz
University of Essex
Colchester, United Kingdom
{szimme,athorpb,foxcj,udo}@essex.ac.uk

ABSTRACT

Privacy concerns are becoming a dominant focus in search applications, thus there is a growing need to understand implications of efforts to address these concerns. Our research investigates a search system with privacy warning labels, an approach inspired by decision making research on food nutrition labels. This approach is designed to alert users to potential privacy threats in their search for information as one possible avenue to address privacy concerns. Our primary goal is to understand the extent to which attitudes towards privacy are linked to behaviors that protect privacy.

In the present study, participants were given a set of fact-based decision tasks from the domain of health search. Participants were rotated through variations of search engine results pages (SERPs) including a SERP with a privacy warning light system. Lastly, participants completed a survey to capture attitudes towards privacy, behaviors to protect privacy, and other demographic information.

In addition to the comparison of interactive search behaviors of a privacy warning SERP with a control SERP, we compared self-report privacy measures with interactive search behaviors. Participants reported strong concerns around privacy of health information while simultaneously placing high importance on the correctness of this information. Analysis of our interactive experiment and self-report privacy measures indicate that 1) choice of privacy-protective browsers has a significant link to privacy attitudes and privacy-protective behaviors in a SERP and 2) there are no significant links between reported concerns towards privacy and recorded behavior in an information retrieval system with warnings that enable users to protect their privacy.

KEYWORDS

Privacy; Information Retrieval; Interactive Studies

ACM Reference Format:
Steven Zimmerman, Alistair Thorpe, Chris Fox, Udo Kruschwitz. 2019. Investigating the Interplay Between Searchers' Privacy Concerns and Their Search Behavior. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331280

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331280

1 INTRODUCTION

There are many examples where personal data is collected via modern technology, of which 3rd-party tracking is just one. For information retrieval (IR) platforms, such as Google, this data combined with user queries, has potentially high value with respect to personalization [15] and advertising [17]. This data can be traced to an individual via linkage of "pseudo-anonymous" data, such as the IP address of the home router being used. While this data may be useful for demographic segmentation, including personal health, it raises many privacy concerns [1, 7, 16, 19, 20]. While seemingly innocuous, collection of our personal data (via 3rd-party tracking and other means) comes with potential threats and harms [3, 4, 12, 21].

Several search engines (e.g. Duck Duck Go) now promise not to collect personal data and web browsing tools (e.g. Tor) are also available to prevent identification. Evidence suggests only a minority of people[1] make use of 3rd-party blocking software, even though this technology demonstrates promising results [3, 21].

Combining these matters and knowledge of the current IR landscape, no search platform is currently available that allows the user to easily assess the level of personal privacy exchanged for a click on a result. It is our goal to begin exploration into such a system that provides privacy risk assessment. Understanding impacts on a user's information need is one avenue to explore with regards to system changes to address privacy concerns [22].

We identify two additional problems. First, to our knowledge, no SERP designed with a privacy warning system has demonstrated a reduction in privacy impacts in search. Second, a multitude of survey questions[2] attempt to capture privacy preferences, however these are often misleading and rarely linked to actual behaviors [8]; also, they are frequently proprietary and thus costly to use. Taking [10] as one example, strong reliability is noted with self-report measures, but no validity is provided through linkage to actual behavior. We also note limited attempts to capture self-report behaviors for enhancing privacy protection (e.g. usage of anonymous networks and browsers that prevent fingerprinting), which we identify as promising areas to explore.

To address problem 1, we designed a SERP with privacy warnings inspired by nutrition labels; nutrition labels are a proposed approach [6] to address misinformation in IR. To address problem 2, we use the survey in [10] as a starting point, however we followed suggestions in [5] and by authors of [9] to develop a set of questions with the goal of identifying individuals that do behave as they report. One main contribution of our work is a set of questions that may be used by future researchers to identify individuals that express concern about privacy and are more likely to take action

[1] In our study, only 11% subjects reported usage.
[2] see examples in [10, 17] and Table 3.1 in [8]

to protect their privacy, counter to the "privacy paradox" [13]. The other contribution being evidence that a stoplight SERP approach designed to reduce privacy impacts does in fact reduce privacy impacts for this subset of privacy-concerned individuals.

2 QUESTIONS AND HYPOTHESES

The central questions to our experiment are 1) *"Does a search system that alerts users to potential privacy threats for results, significantly reduce privacy impacts when compared with a baseline search system?"*, 2) *"To what extent are self-reported attitudes towards privacy linked to behaviors to protect privacy?"*, 3) *"For a set of self-reported privacy-protective beliefs and behaviors, what self-report measures are both reliable and valid?"*. Based on user interactions in a search interface linked to the privacy proxy of 3rd-party trackers and their responses in a post-experiment survey, we address each of these questions.

Guided by previous work demonstrating that stop light approaches produce healthier eating behaviors [11, 18], we expect the baseline to have significantly higher encounters with privacy trackers. For any self-report measures (attitudes and behaviors regarding privacy) found to be reliable, we expect that attitudes more strongly supporting privacy protection will be significantly linked to self-reported behaviors that protect privacy. For any reliable self-reported attitudes and behaviors more strongly supporting privacy, we expect a significant link to privacy-protective behaviors in the SERP offering privacy-protective warning lights, with significance being an indicator of the survey's predictive validity.

3 METHODS

We used the same methods and design as in [22], which were an adoption of methods by Pogacar et al. [14]. The updated corpus developed by [22] was also used in our study, as well as use of the result ranking method. The questionnaire developed for our study is introduced in the measures section below.

3.1 Experimental Design

A within-group design was used to conduct the experiment, in which users encountered SERP variants and multiple search tasks. SERP results were retrieved from the corpus, which included links to 3rd-party trackers. The Ghostery 3rd-party tracking tool[3] was used to determine the number of trackers for each result in the corpus. 3rd-party trackers are a relevant proxy as they often collect pseudo anonymous information such as IP addresses, which can be easily linked to who and/or where you are. The 3rd-party trackers for each website are treated independently from each other.[4]

3.1.1 Search Tasks. 10 fact based health search tasks dependent upon findings from Cochrane Medical reviews were used. The medical reviews determined if research evidence indicated a particular medical treatment was helpful or not helpful for a given medical condition, with the possibility that evidence was inconclusive. Thus, as was performed in [14] and [22], participants in this study had to determine if the treatment was *helpful, not helpful* or *inconclusive* based on the information they had available. Participants were not made aware (until post-experiment debriefing) that there were an equal number of helpful and not helpful tasks and that no inconclusive reviews were included.

3.1.2 SERP. Participants encountered 5 different interfaces, 4 SERP variants and one control question (as was done with [14] and [22]). Using the same Graeco-Latin design method as [14], 2 search tasks (1 helpful and 1 not helpful) were assigned to each interface. A baseline SERP, similar to modern search engines was used, where results were ranked based on the cumulative distribution function (CDF) used in [22]. The second SERP variation, the nutrition-label-based stoplight approach, displayed results with the same ranking method, however there was a stoplight to the left of results based upon the number of privacy trackers. The coloration of stoplights was based upon the median and upper quartile of privacy trackers for documents assigned to the search task (see Figure 1). Two other SERP variants, not in the scope of research questions covered herein, were also encountered by participants.[5] As did [14] and [22], definitions of the medical question and the 3 possible decisions (helpful, not helpful or inconclusive) were provided to participants throughout the search task. Users could click through multiple pages, with 10 results per result page and 19-22 documents available per search task, at most users had 3 pages available. A button was provided to take users to a page where they made a decision on the appropriate answer to the medical question.

3.1.3 Questionnaire. The survey (implemented with Qualtrics) was provided to participants after the main experiment, which was done to minimize the possibility of priming. Demographic information was collected first (including age, gender and educational information). The survey questions that followed were grouped in blocks as outlined in Section 3.2.1. Within each of these blocks, the questions were randomized, however we did not randomize the ordering of the blocks. Privacy attitudes were captured first, followed by reported behaviors (e.g. what browser type and frequency they used), after which we collected topical privacy attitudes (e.g. health and finance) and lastly preferences about the SERP. At the very end of the questionnaire users were provided with exit information to ensure the correct information about search tasks was provided. Section 3.3 covers methods for scoring.

3.1.4 Participants. After receiving ethics approval from the university review board, we recruited 91 participants via volunteer lists maintained by the psychology department (of which 4000+ people are possible). Each participant received £10 or a course credit for their time. The experiment and survey combined took 50 minutes on average. Each rotation of the experiment required 10 participants, thus 9 rotations were completed with one additional participant included in our analysis. Baseline results were not recorded for 6 users, and these were removed for paired t-tests. Of our sample(N = 91), 89 were students, of which 77 of those were undergraduates. The average age of participants was 23.2 years and 63 identified as

[3] https://www.ghostery.com/ (Accessed August 2018)

[4] For example, Google DoubleClick is a tracker on many different websites, and is not independent if you visit two different websites using this tracker. In the experiment, each tracking encounter is treated as separate from encounters on different websites.

[5] One variant re-ranked results based on total trackers (result 1 would have least number of trackers and last result for task would have highest number of trackers). The other variant filtered out any results with the number of trackers above the median. In all SERPs, users had the option to shut off the privacy-protective method, but no participant went into the settings to make such a change.

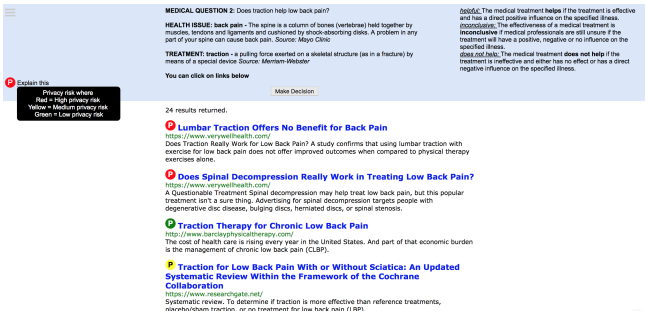

Figure 1: SERP with privacy stoplights shown. An explanation of lights was available to users. Color of stoplight was based on the number of 3rd-party trackers for the associated website in results, where red was given to the top quartile, yellow assigned to results with total trackers between median and top quartile and green given to everything else. A tooltip was provided to explain the different colors, however participants were not given any other information (until debriefing) that we were investigating their search behaviors with respect to privacy. The baseline SERP had no stoplights, but otherwise had all the same elements. In the top center of the SERP the task description and medical definitions were available. Definitions of the decisions were in the top right. Pagination was available with 10 results per page.

Female. English was reported as the native language for 27 students. STEM departments provided 53 participants, with 17 participants having a background in computer science. 42 participants reported using a privacy-protective browser at least once a month, while only 10 reported usage of a privacy-enhancing search engine.

3.2 Measures

For the interactive experiment, the SERP interface type (baseline vs. stoplight) is used as the independent variable. For the dependent variable (privacy tracking), we take the average trackers encountered for each search task and normalize by the maximum number of trackers that could be encountered for the search task. If a user did not click on a link for the task, then 0 trackers were recorded. We also use the self-report behaviors as the dependent variable for analysis of self-reported privacy attitudes.

3.2.1 Survey. We generated a survey to capture general attitudes towards privacy, attitudes towards health privacy and behaviors considered to be privacy-enhancing (we consider general behaviors, and usage of browser and search engine). The two attitude groups were on 7-point Likert scales, the general behavior group was a set of yes/no options while the search and browser behaviors were reported frequencies (daily, weekly, monthly or never). Though utility measures (such as [2]) were an avenue considered, due to concerns around participant fatigue we excluded such measures from our study.

The general attitude group consisted of 7 questions in total, and included questions such as "How likely do you think it is that personal information submitted / shared on the internet will be: **AQ1** - Shared or sold to others? **AQ2** - Used by others to harm you? ". The health attitudes group had 3 questions, such as "**HQ1** - How concerned are you about the sensitivity of the following personal information that you share? "[6]; not included here is a list of 10 other non-health topics adapted from the appendix in [10]. As final measures, for each attitude group we used the normalized mean of responses (i.e. sum responses divided by total possible points).

For the behaviors[7], we also use the normalized mean of responses. For general behaviors we asked: "**EQ1** - Which of the following do you do?" [8], with 1 point each. Browser enhancing behavior was captured by "**EQ2** - How frequently do you use the following web browsers? `[Tor; Mozilla Firefox; Brave]`". Search-enhancing behavior: "**EQ3** - How frequently do you use the following search engines? `[Duck Duck Go; Qwant]`".

We tested the reliability of the two attitude groups and found they demonstrated excellent internal consistency, with Cronbach's α = .76 for the 7 general attitudes questions and Cronbach's α = .89 for the health questions. We considered combining the 3 privacy behavior groups as one measure, but given the unsatisfactory reliability of the scale (Cronbach's α = .53), we kept the 3 measures (General behaviours, usage of browsers, and usage of search engines) separate.

3.3 Analysis

We used linear regression to test our hypotheses with respect to the self-report constructs on the dependent SERP experiment privacy metrics variables (in Section 3.2), for which we moderated the analysis by the type of interface (baseline or stoplight) associated with the search task. Additionally, we performed linear regression to model possible effects of self-report attitudes onto self-report behaviors. Two-sided independent and paired t-tests were used to compare differences for the interactive design, with significance values set to $p < .05$.

4 RESULTS AND DISCUSSION

With respect to our first question, similar to findings with a smaller sample (n = 40) in [22], we find no evidence in our entire sample suggesting that a privacy stoplight approach will reduce privacy impacts for the population as a whole when compared with the baseline.

Our findings in Table 1 are helpful for answering our remaining two questions around reported attitudes and behaviors. The privacy attitude construct, while reliable appears to have limited predictive validity, we find no evidence suggesting it is a useful measure to predict how individuals will behave when presented with a stoplight privacy warning system in a SERP. We are not entirely surprised by this finding based on existing literature. Conversely, self-reported privacy behaviors are more promising at predicting how users

[6]For brevity, questions **AQ3-7 and HQ2-3** are not included. These questions can be found in a copy of the full survey given at end of experiment at https://github.com/stevenzim/sigir-2019

[7]Note, surveys capturing enhancing behavior are limited, therefore our survey items are based on suggestions from https://www.privacytools.io (Accessed August 2018)

[8]`"Please select all that apply: Use = [Anonymous communications networks, end-to-end encryption, software to block 3rd-party trackers, a VPN, software to prevent browser fingerprinting, software to delete cookies automatically, software to ensure HTTPS communication], Disable = [javascript, cookies]"`

will behave in our privacy-enhancing SERP, where we find small effects when considering reported usage of browsers that enhance privacy. In our analysis, we noted a distinct split between subjects (those who reported usage of privacy-enhancing browsers, and those who do not). Using this split, we performed a Welch's two-sided independent t-test comparing % of max trackers encountered in the stoplight SERP for participants that use privacy-enhancing browsers ($n = 42, M = 0.22, SD = 0.19$) with those that do not ($n = 49, M = 0.28, SD = 0.24$) and find these differences to be significant $t(179.79) = 2.11$, $p = .036, d = .31$. The same split is used for within-group analysis to compare interactions against the baseline SERP. For the group not using privacy-enhancing browsers, no differences were found for comparison of the stoplight SERP ($M = 0.30, SD = 0.15$) with the baseline SERP ($M = .31, SD = 0.22$), $t(47) = -0.46$, $p = .646$, $d = -0.07$. For the group reporting usage of privacy-enhancing browsers there is a trend towards reduction in privacy encounters with the stoplight SERP ($M = 0.23, SD = 0.17$) compared to the baseline SERP ($M = 0.31, SD = 0.18$), $t(36) = 1.91$, $p = .064$, $d = 0.31$.

Table 1: Results of 4 predictive models with significant effects are included in the table. In bold are dependent variables of linear regression models, with the lines below separating each model. The first two results are for self-reported attitudes (Attd) towards general and health privacy, each on the reported Behavior (Bhv). Bhv for the browser is regressed on the % of maximum privacy tracker encounters in the SERP for the search task. $M1$ denotes model on data from stoplight SERP only (with no moderation). Model $M2$ is moderated by the SERP type (baseline or stoplight), all other models had one input variable. ** and * suggest significant results with $p < .01$ and $p < .05$ respectively. Not included in table, we find $p = 0.11$ to 0.55 for models: Attd -> Bhv (general and search), Attd -> % of maximum trackers and Bhv (general and search) -> % of maximum trackers.

Bhv-Browser	b	F	df	SE	p
Attd-General	0.30	10.96	89	0.09	**.001
Attd-Health	0.14	8.98	89	0.05	**.004
% of max trackers					
Bhv-Browser^{M1}	−0.29	5.57	180	0.03	*.020
Bhv-Browser^{M2}					
Bhv-Browser	−0.23	−2.64	360	0.09	**.009
SERP	0.03	1.34	360	0.02	.182
Bhv-Browser × SERP	0.12	0.70	360	0.17	.490

The remaining two self-report behaviors show some promise, however we note that a limited number of participants (< 10 reported use of privacy-protective "search" engines and < 20 for usage of "general" internet privacy protection methods) reported behaviors contributing to these constructs, additional studies are needed to better formulate and to confirm validity of these constructs. Results for both "general" and "search" are discussed in Table 1 caption.

Based on these findings, we conclude that the self-reported browser type is a useful pre-screening metric to identify participants more likely to be concerned about privacy, which is a key contribution of our work. We also believe that our stoplight approach is a promising direction to head for individuals that want to reduce personal privacy impacts. To increase confidence in the findings presented here, future research should consider refinement of our privacy constructs as well as the potential for online data collection to overcome the limitations of within-subjects design/resource constraints.

ACKNOWLEDGMENTS

This work was supported by the Economic and Social Research Council grant number ES/M010236/1 and The Human Rights Big Data and Technology (HRBDT) Project.

REFERENCES

[1] B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel. 2013. Inferring the demographics of search users: Social data meets search queries. In *Proc. WWW*. 131–140.
[2] J.P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira. 2013. Your Browsing Behavior for a Big Mac: Economics of Personal Information Online. In *Proc. WWW*. 189–200.
[3] S. Englehardt and A. Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proc. SIGSAC*. 1388–1401.
[4] S. Englehardt, D. Reisman, C. Eubank, P. Zimmerman, J. Mayer, A. Narayanan, and E.W. Felten. 2015. Cookies that give you away: The surveillance implications of web tracking. In *Proc. WWW*. 289–299.
[5] F.J. Fowler Jr. 2013. *Survey research methods*. Sage publications.
[6] N. Fuhr, A. Giachanou, G. Grefenstette, I. Gurevych, A. Hanselowski, K. Jarvelin, R. Jones, Y. Liu, J. Mothe, W. Nejdl, I. Peters, and B. Stein. 2017. An Information Nutritional Label for Online Documents. In *ACM SIGIR Forum*, Vol. 51. 46–66.
[7] E. Horvitz and D. Mulligan. 2015. Data, privacy, and the greater good. *Science* 349, 6245 (2015), 253–255.
[8] G. Iachello and J. Hong. 2007. End-user privacy in human–computer interaction. *Foundations and Trends® in Human–Computer Interaction* 1, 1 (2007), 1–137.
[9] S. P Jenkins, L. Cappellari, P. Lynn, A. Jäckle, and E. Sala. 2006. Patterns of consent: evidence from a general household survey. *Royal Statistical Society: Series A (Statistics in Society)* 169, 4 (2006), 701–722.
[10] P.G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L.F. Cranor. 2013. What matters to users?: factors that affect users' willingness to share information with online advertisers. In *Proc. SOUPS*.
[11] N. Maubach, J. Hoek, and D. Mather. 2014. Interpretive front-of-pack nutrition labels. Comparing competing recommendations. *Appetite* 82 (2014), 67–77.
[12] J. R. Mayer and J. C. Mitchell. 2012. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*. IEEE, 413–427.
[13] P.A. Norberg, D.R. Horne, and D.A. Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Consumer Affairs* 41, 1 (2007), 100–126.
[14] F. A. Pogacar, A. Ghenai, M. D. Smucker, and C. L. A. Clarke. 2017. The Positive and Negative Influence of Search Results on People's Decisions about the Efficacy of Medical Treatments. In *Proc. SIGIR*. 209–216.
[15] J. Teevan, S. T. Dumais, and E. Horvitz. 2005. Personalizing search via automated analysis of interests and activities. In *Proc. SIGIR*. 449–456.
[16] Z. Tufekci. 2015. Facebook said its algorithms do help form echo chambers, and the tech press missed it. *New Perspectives Quarterly* 32, 3 (2015), 9–12.
[17] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proc. SOUPS*. 4.
[18] E. van Herpen and H.C.M. van Trijp. 2011. Front-of-pack nutrition labels. Their effect on attention and choices when consumers have varying goals and time constraints. *Appetite* 57, 1 (2011), 148–160.
[19] I. Weber and C. Castillo. 2010. The demographics of web search. In *Proc. SIGIR*. 523–530.
[20] I. Weber, V.R.K. Garimella, and E. Borra. 2012. Mining Web query logs to analyze political issues. In *Proc. WebSci*. 330–334.
[21] Z. Yu, S. Macbeth, K. Modi, and J.M. Pujol. 2016. Tracking the trackers. In *Proc. WWW*. 121–132.
[22] S. Zimmerman, A. Thorpe, C. Fox, and U. Kruschwitz. 2019. Privacy Nudging in Search: Investigating Potential Impacts. In *Proc. CHIIR*. 283–287.