The value of privacy in Web search

Sören Preibusch
Microsoft Research Cambridge
spr@microsoft.com

Abstract. Search is the most prevalent Web activity. The query logs of a user paint a picture of her interests and lifestyle and may allow inferring her identity. The integration of search engines with advertising networks and account-based services such as email or app stores heightens privacy concerns around Web search. This paper reports on the first laboratory experiment that assesses the value that consumers attach to their privacy when using a Web search engine. 189 participants were invited to enable multiple privacy-enhancing options on a query-per-query basis. Usage of these options was high (up to 79% adoption), and increased with search query sensitivity. However, only 15% to 16% of participants were willing to spend half a penny extra for keeping queries out of their search history or for preventing data sharing with third-parties. Usage decisions were found not to be systematically dependent on consumers' privacy concerns or the importance they attach to the privacy-enhancing features.

1. Privacy and the economic value of personal information in Web search

Web search is the most prevalent online activity ahead of email [**2**], and Web search engines are amongst the most popular destinations on the Web [**1**]. They act as gatekeepers to the Web and channel consumers to Web content. Continuous innovations in search engine functionality have lead to new ways in answering users' queries, such as entity answers, currency converters or calculators. Modern search engines combine information about the user and her query to infer its intent.

In a typical search engine, advertising is displayed along the organic search results. Advertising generates revenues that allow offering the search functionality to consumers free of charge; it is targeted and can be based on the query and the user's behaviour [**9**].

Through her sequence of search queries, a user exposes a wealth of information, including preferences, interests, geographic location and Web sites visited. These queries may contain sensitive or embarrassing material, such as credit card numbers or adult content. From their behavioural data, users' identity may be learned, as demonstrated when AOL released its supposedly anonymised search logs [**3**]. While the AOL case demonstrated the feasibility of manual re-identification for select individuals, it is also possible to automatically infer users' demographics at scale, even from scrubbed logs. From a user's queries, simple classifiers can predict gender at 84% accuracy or age at an absolute error of 7 years [**13**]. Beyond the queries themselves, potentially privacy-threatening details now include the clicks on the result page and metadata such as location and time of day. Some large search engines are also integrated with Webmail services, app stores, electronic commerce payment solutions, and advertising networks. This combination of services generates even broader user profiles.

Privacy concerns among the online population are high. Looking at Web search in particular, the majority of Web users are uncomfortable with their personal information being monetised to finance free search [**18**]. Less than four in ten Europeans are comfortable with search engines to use information

about their online activity to tailor advertising or content to their interests and hobbies [**18**]. At the same time, Web search engines are amongst the least trusted companies to collect and to store personal information [**18**].

Previous research has engineered technical approaches to prevent the search engine from learning a wealth of personal details about its users. Without the need to rely on the search engine's cooperation, a community of users would shuffle their queries amongst each other and then broadcast the corresponding results to all participants [**7**]. Additional cryptography could be used to hide the search queries amongst the participating users [**15**]. In practice, existing systems rely on a central authority, more trusted than the search engine itself, to aggregate search queries over different users, with the aim to hide user identifiers through proxying. Examples of these privacy-enhanced Web search engines are Startpage or Ixquick [**11**], which was awarded the first European Privacy Seal [**14**].

Although obfuscation of search queries or their dissociation from the user reduces the ability for search engines to build fine-grained profiles, it equally hampers their ability to serve the user with personalised search results. Whilst personalisation can be undesirable, it improves the quality of the search results and of query suggestions. Both applications rely on a user's personal search history, functionality available in mainstream Web search engines such as Bing or Google, but also in site-specific search engines, such as the product search on Amazon. Users can typically inspect their search history, clear it, or turn off the feature. Users may also curate their search history by selectively removing queries from it. This is an example of privacy empowerment, because the consumer can decide for herself what to remove and what to keep—instead of having to choose between full exposure and total anonymity.

As an aside, privacy empowerment in Web search can also happen as part of its gatekeeper role. Privacy Finder is an example of a search engine that annotates search results with a visual privacy-rating and also ranks privacy-friendly sites higher [**8**]. These visual indicators of superior privacy practices can be strong drivers for consumers' purchase decisions and also their willingness to spend a premium on privacy. In an experimental study, users were found to pay around $0.60 more when shopping through a Website for which a good privacy rating was displayed [**19**]. In this case, however the privacy safeguards are aimed at the user's final destination on the Web, rather than the search experience.

Contribution. This paper reports on the first laboratory experiment that measures users' contextual appreciation and use of several privacy-enhancing features in a search engine. By considering search queries of different sensitivity, this paper also provides first empirical evidence into consumers' varying willingness to pay for privacy-enhanced Web searching.

Structure. The remainder of this paper is organised as follows. Section 2 gives a detailed account of the experiment methodology. Section 3 presents the results and provides descriptive statistics before testing the research hypotheses in Section 4. A summary and outlook on future research conclude the paper in Section 5

2. Experiment methodology

2.1. Research hypotheses

In addition to providing descriptive insights into consumers' appreciation of privacy-enhancing Web search features, the experiment was designed to test five research hypotheses:

H1 The price of privacy-enhancing features and the proportion of users enabling them are negatively associated.
H2 The more sensitive the search task, the more likely users will enable privacy-enhancing features.
H3 The more sensitive the search task, the less likely users will enable privacy-invasive features.
H4 Users who are more concerned about privacy will enable privacy-enhancing features more often.
H5 Users who consider privacy-enhancing features more important will enable them more often.

Hypothesis H1 is examining users' willingness to pay for privacy-enhancing features. Previous research has established price mark-ups that consumers are willing to pay for better privacy in the context of online shopping. In a US lab experiment, participants paid a privacy premium of approximately $0.60 when the Website they shopped with was labelled as having good privacy practices. Shoppers exhibited higher propensity to spend extra for better privacy when shopping for more sensitive products [**19**]. A series of lab and field experiments in Europe established that users prefer shopping with a privacy-friendly online retailer but few would be willing to incur higher costs. A price discount of €1 overrides customers' genuine privacy preferences and makes them buy from a shop with a more invasive data collection scheme [**5**]. When buyers can compare data collection schemes side-by-side, around a third is willing to spend €1 extra for not revealing their mobile phone number. Much fewer buyers would pay extra for not receiving advertising to their email address [**12**].

Hypotheses H2 and H3 examine the impact of search query sensitivity on users' propensity of enabling privacy-enhancing options. They rest on earlier observations that people are more inclined to protect information about them that is more sensitive [**6**].

2.2. Experiment procedures

The experiment was carried out in sessions at University College London (UCL); recruitment was done locally at the university. The experiment was framed as trialling a new Web search engine in collaboration with an industrial partner. Neither the recruitment advert nor the information sheet mentioned privacy. The design was approved by the ethics committees at UCL and Technical University of Madrid.

Two pilot sessions of the experiment were administered to 32 participants in September 2012 to test technical reliability, procedures, usability and design. Further 44 participants were recruited for the first main deployment in October 2012. Afterwards, two treatments were deployed, yielding a total of 189 valid cases, whose data will be used for the subsequent analysis. The experiment was structured in seven phases.

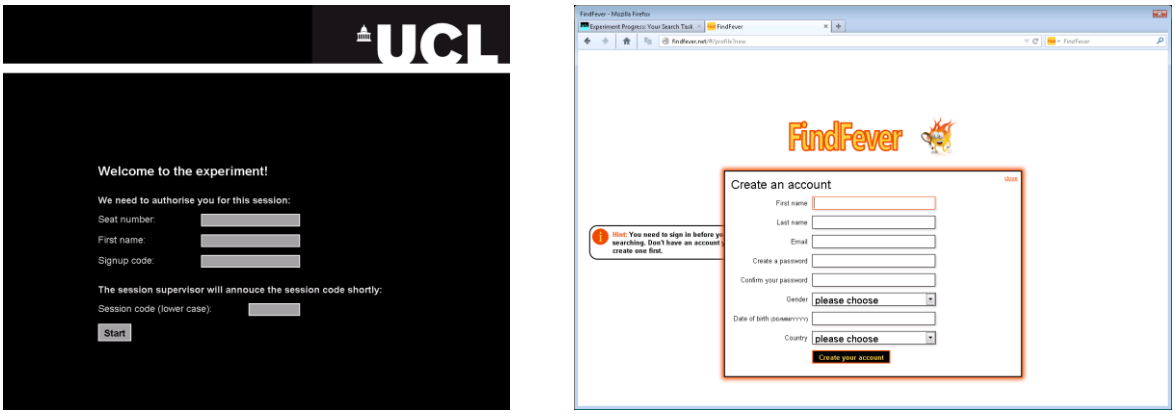

Figure 1: Examples of the computerised experiment interface. All interfaces relating to the administration of the experiment were black and branded with the university logo. These style elements were absent from pages pertaining to the FindFever search engine.

Registration. Participants signed up online through a self-service system. There were two recruitment tracks: through a pool of pre-registered volunteers and through fly-posting / notices. All participants needed to sign up through a dedicated Website that also issued a unique sign-up code. Participation was voluntary and participants did not receive course credits for taking part in the experiment.

Admission. Admission to the experiment was conditional on prior registration. Any participant could participate only once. As the entire experiment was computerised, admission checks were carried out electronically. Participants had to enter their first name and their sign-up code before they could begin with the experiment. A session code, shared amongst all attendees of a particular session was also necessary. This code was announced by the experimenter, after all participants had read the information sheet about the experiment procedures and signed the consent form.

Comprehension check. At the first phase of the experiment, participants had to answer two comprehension questions relating to the instructions, which were available to the participants throughout the experiment. Progress to the next phases was not dependent on answering the check questions correctly.

Search tasks. After the comprehension test, participants started to work on the search tasks. Twenty search tasks were given, which had to be completed in order. Participants could skip a search task or indicate they were unable to find the answer. There was no correctness check at the point of submitting the answer. All search tasks were given as questions; a comprehensive list is given below (Table 1). From the search task page, there was a direct link to the experiment search engine.

Sign-up with the search engine. Before participants could issue their first query, they had to create individual accounts with the experiment search engine and then sign in with their credentials (Figure 1). Before the first search could be issued, users also had to configure their search options, which were displayed on-screen. There was no default for the search options, so that all options had to be turned on or off; they could not start searching before making a decision on each. Users remained logged in unless they signed out manually. Their search options remained configured for the next query, but changes to these options could be made at any time.

Exit-questionnaire. Once participants had finished all search tasks, they were redirected to the exit-questionnaire, which asked for demographics and psychometrics. The questionnaire also included questions about the experience with search engine, the search tasks and the available search options. General privacy attitudes and self-reported behaviour and computer literacy were also asked for.

Payoff. The experiment session finished upon completion of the exit-questionnaire. The farewell screen displayed the participant's individual payoff, which was made up of a constant show-up fee (£8) and a dependent component that varied with participants' configuration choices and with how many questions they had answered: on top of an initial endowment of 1000 credits, each answer was rewarded with 40 credits. Credits could be spent on searches at 1 credit per query, or on search options at varying prices (Figure 2). On average, it took participants 89 minutes to complete the experiment, from admission to payoff, and they were paid an average total reward of £12.46.

2.3. Treatments, materials and apparatus

Treatments. After pretesting and piloting, two treatments were administered that differed in the price at which privacy-enhancing options were charged. In treatment T0, all a priori privacy-enhancing search options were free—no extra credits needed to be spent to enable them. In T2, all privacy-enhancing search options required two extra credits each from the user. Two credits correspond to half

a penny or twice the base price for issuing a query without any options enabled. All other parameters remained constant across the two treatments, including the prices of all other options and their presentation on the screen.

Handouts. At the beginning of the session, participants were given an information sheet that welcomed them to an experiment into "trialling a new Web search engine". It was explained to the participants that they had to complete twenty search tasks and that for every query, they could set certain search options. The information sheet also included basic instructions such as switching off mobile phones and not talking with other participants in the session, but privacy was not mentioned on the information sheet. Participants were also given a standard consent form which was then collected, once signed, prior to starting with the session. At the same time, a laminated sheet explaining all available search options was given to the participants. This sheet explained how to turn an option on or off and listed all available options as displayed on the screen, including their prices and together with a two-line explanation. For instance, the option "Do not record in my search history" was explained as: "Your search history is a log of all your queries. If you enable this option, your query will not be kept in your search history." Again, privacy was not mentioned. Participants kept the information sheet and the explanation sheet throughout the session.

Apparatus: the FindFever search engine. The experiment provided access to FindFever, the new search engine on trial. FindFever featured a consistent branding with colours, buttons, fonts, logo and other images. The Web search functionality was provided through the public Bing Search API (http://datamarket.azure.com/dataset/bing/search), requests to which were made in real-time and according to the functionally relevant search options: geo-targeting, search term highlighting and safe search were all honoured by setting the corresponding API parameters.

For the search option of improved quality, results were re-ranked to introduce a controlled quality gradient: unless the option to improve search quality was enabled, search result quality was artificially degraded by interspersing lower-ranked results in the top results. As indicated by the exit-questionnaire, 88% of participants agreed the "extra quality" option would improve the search result quality, which is the highest proportion recorded for any search option.

On the FindFever homepage, where search options could be configured, the user was greeted using her sign-up name; this practice is similar to other search engines which display the user's name when logged in (e.g., Bing, Google). The homepage also displayed a balance counter showing the remaining credits. Again, a similar balance counter is seen in the Bing Rewards programme. The price for a query with the currently selected options was displayed clearly under each search button on the homepage or on the results page.

Apparatus: the browser. The browser was configured so that FindFever was set as the default search engine and would also be used when a search term was entered into the location bar. Other search engines, including Bing, Google, Yahoo! Search and Ask were disabled and their sites redirected to FindFever. Users could otherwise freely browse the Web while answering their search tasks; the click-stream was not recorded.

2.4. Privacy impact

The experiment featured three kinds of deliberate invasions of privacy, in the relationship between the user and the search engine FindFever. First, any user had to register with FindFever. Upon registration, the user needed to provide the following information on a mandatory basis: first and last name, email address, gender, date of birth and country. The email address needed to be validated by means of a confirmation email. Second, the search queries themselves issued to FindFever revealed behavioural

information. Given that users needed to be logged in for searching, all queries that users issued were linked to their account. On the search engine results page, the search history was displayed in a vertical pane on the left hand side, thereby increasing salience of the search history and logs kept about the user. The user had the opportunity to remove queries from their search history. Third, users could also opt-in to privacy-invasive search options for which they would be rewarded with additional credits: two extra credits could be earned by having the clicked links on search results recorded, or by having one's search added to FindFever's public Twitter feed.

Although these may seem as big privacy impacts, similar procedures are customary in today's Web search engines: if logged into their Web-based Gmail account, for instance, users of the Google search engine have their queries attached to their personal account. A search history is kept by Google and Bing for instance, and may be displayed prominently on the front page of the search engine. Similarly, these search engines expose functionality for managing the search history. The practice of recording outgoing clicks is also implemented in major search engines [**17**].

2.5. Stimuli: search tasks

The stimuli in the experiment were the twenty, sequentially presented search tasks. Presented in form of questions, they were devised to be answerable with around three queries. The search tasks did not relate to the individual but were fact-based questions. For instance, there was no question such as: "Can your credit card number be found on the Internet?" Questions also avoided highly sensitive topics for ethical reasons, although there was a continuum in the sensitivity of the topics covered. There were more innocent search tasks than there were sensitive tasks. Table 1 below gives an overview of the search tasks, including their sensitivity, as recorded in the exit-questionnaire. For low sensitivity search tasks, less than 5% of participants agree that this is a sensitive question. For medium and high sensitivity search tasks, less than 20% or respectively more than 20% agree that this is a sensitive search task. Questions were not ordered by sensitivity; high sensitivity questions were well dispersed throughout the sequence of tasks (Figure 4). The questions underwent much editing during the design phase of the experiment and were further improved after pretesting the study.

search task	sensitivity
What is the population size of Little Shelford, England?	4%
What is the weight of an adult European Robin?	4%
Where can you buy lingerie in Chelsea?	**28%**
What is the street address of the tourist information in York, England?	5%
What would be a good honeymoon suite in Singapore?	8%
One can typically look for the "Way out" sign to leave the underground. Which British city or metro system uses different signs?	7%
What is the maximum penalty for attempted sexual intercourse with girl under 13?	**43%**
Which airport is closest to you right now?	4%
Which hotel is closest to you right now?	4%
What is the pH value of tap water in Croydon, England?	4%
How large is the East of England region?	6%
What is the maximum penalty charge for deliberately giving wrong information on your	**21%**

tax claim?	
Where is the mental health centre in Peterborough?	15%
What is the proportion of water in human faeces?	**29%**
What is the fee for renewing your shotgun certificate in Camden?	13%
Which medication is used to treat leprosy?	13%
What is the average income in the area where you are living?	10%
What is a popular lap dance bar in Westminster?	**43%**
How many people were living with HIV in 2011?	13%
What is Justin Bieber's message he wants to convey?	**26%**

Table 1: All twenty search tasks in the order presented during the experiment. The column 'sensitivity' gives the proportion of participants who agreed or strongly agreed on the exit-questionnaire that this is a sensitive question. High sensitivity questions are highlighted.

2.6. Privacy choices: search options

For each search query anew, participants could configure their search options. Nine search options were available, affecting functionality, usability and privacy. Most importantly, none of these options was enabled or disabled by default. Users had to click to turn an option on or off. Options also varied by price: there were free options and others required extra credits (1 or 2). Two privacy-invasive options (posting a query to Twitter, recording outgoing clicks) allowed participants to earn two extra credits. One credit is equivalent to a quarter of a penny (approximately 0.3 Euro cent or 0.4 US Dollar cents). Figure 2 gives an overview of the different search options, as seen by the participants.

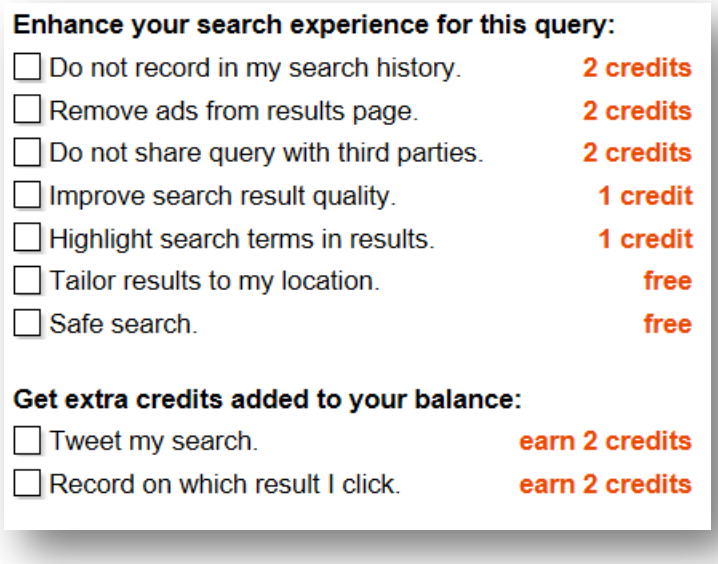

Figure 2: Search options available on FindFever and their prices, as shown in treatment where privacy-enhancing options were paying (T2). Cropped screenshot from the experiment.

3. Results and descriptive statistics

3.1. Sample description

In treatments T0 and T2, when privacy-enhancing search options were free, respectively required 2 credits, 97 and 94 valid cases were recorded.

Demographics. Based on the exit-questionnaire, the overall sample had the following characteristics. The median age of the participants was 22 years (range 18 to 62). Of all participants, 62% were women and 42% had completed a degree course at university.

Privacy and security. Of all participants, 74% reported being very or fairly concerned that their personal information is being protected by the organisations that hold it. Only 3% are not concerned at all. 44% do not believe the transfer of their information through the Internet is secure. The instrument by Smith et al. was used to measure privacy concerns [**16**], which yields scores from 1 (low concerns) to 7 (high concerns). The upper third of the scale was occupied by 71%.

Regarding computer literacy, between 78% and 86% kept a back-up of their data or had changed the homepage or the default search engine in their browser. Almost a third had creative Web experienced (designed a Website, registered a domain), and a quarter had configured a firewall.

Regarding computer fraud exposure, more than 90% had received spam and 52% had caught a computer virus. Other experiences of cybercrime were much less frequent, but in combination 38% had been victims of at least one of phishing, credit-card fraud, data breach or misuse or identity theft.

3.2. Search tasks

The experiment featured twenty search tasks of varying sensitivity. The sensitivity scores for each task are given in Table 1 as the proportion of participants who agreed in the exit-questionnaire that a task was sensitive. On average, 37% of participants agreed that the search task was typical for their search behaviour. Among the high sensitivity search tasks, the most typical ones related to buying lingerie and giving wrong information on the tax claim; the least typical one related to Justin Bieber.

The median number of queries issued per participant is 53 with a wide range between 16 and 344. The upper bound is an example of four participants who issued 200 queries and more. These queries had the credit-earning options enabled, thereby increasing the participants' payoff to the capped maximum. It seems plausible that such behaviour is not an artefact of the experiment but could also occur on other query-rewarding search engines. Consequently, the Bing Rewards Programme enforces a cap on the credits that can be earned per day.

3.3. Search options

In the experiment, participants could enable or disable nine search options. Their original on-screen presentation is reproduced in Figure 2. Figure 3 gives the popularity of different options by treatment. The first three options were designed to be privacy-enhancing, only their price varied across treatments; the last two were designed to be privacy-invasive. A significant difference in adoption across treatments can only be observed for the former ($p < 0.0001$, Fisher's exact test). There is no statistical difference in adoption for any other search option and consequently, aggregate scores are reported.

All search options exhibit a good adoption, even the paying options such as highlighted search terms in the results or improved search quality (47% and 61% respectively). At a quarter of a penny per search, usage of these two options alone corresponds to £17 spend; the money spent by all participants combined on the privacy-enhancing options corresponds to £72. Free options were also received well,

in particular when promising an improvement in result relevance (geo-targeting: 90%). The privacy-invasive options do not see the highest uptake even though their usage was rewarded. A quarter or participants renounced 2 credits by opting out of having their searches tweeted; still 12% did not want to have their clicked results recorded.

Of particular interest is users' perceived benefit from the various search options. The exit-questionnaire asked participants whether they agreed that "[t]his feature increases my privacy". The options "no third-parties" and "no history" top the list with 90% and 80% of users respectively finding this a privacy-enhancing option. Conversely, the option to tweet one's searches is only considered privacy-enhancing by one in eight—87% disagree this option would enhance their privacy. In consequence, these three options will be considered as *the* privacy-enhancing resp. -invasive options. Accordingly, 15% to 16% of participants paid half a penny per search extra to enhance their privacy in Web search by keeping queries out of their search history or by disabling data sharing with third-parties respectively. More than three-and-half times as many users opted for extra quality than for "no history", albeit the price difference was only twofold.

There is a noticeable decline in the adoption of privacy-enhancing options when they require payment (treatment T2) (Figure 4). Irrespective of search task sensitivity, fewer participants spend credits on the "no history" and the "no third parties" options, as the experiment progresses. Speculatively, this decline may be attributed to a last-round effect, as participants reduced their spending with the payoffs approaching.

The experiment required participants to configure their search options prior to their first query as no on/off defaults were provided. Settings were subsequently carried forward from query to query, but could be changed at any time. Most participants never or rarely changed their search settings: 29% kept their own initial configuration for all subsequent queries. Of all participants, 27% changed their settings only once. Only 3% used ten or more different combinations of search options. On average, participants changed their settings twice over all twenty search tasks. This should be interpreted as a high number. Users of existing search engines change their search options much less frequently.

The option to remove advertisements from the search results page was not perceived as particularly privacy-enhancing. Targeted and behavioural advertising is regularly flagged up by mainstream media as an invasion of privacy, but only 63% of participants agreed that the option to remove ads would enhance their privacy. This number is lower than for safe-search (67%) and significantly lower than the proportion of users rating "no history" as privacy-enhancing ($p < 0.003$, Fisher's exact test). Still, 91% of participants considered "no ads" an important option and it featured highest adoption rates across all free enhancements.

Regarding the stated importance of an option, high search quality and the absence of advertisements top the list with 94% and 91% respectively (Figure 3). Usability options (highlighted search terms, 89%) also score high. It is interesting that not all important functionality is currently available in major search engines: whereas they do deliver good quality search, they are ad-sponsored. The ability to tweet a search is considered least important by far (23%). Almost two thirds consider it important that a search engine records on which links they click. Although this practice does not deliver an immediate benefit to the user and had no impact on results in the experiment, it may carry the promise of improved search quality in the medium run. In the exit-questionnaire, 58% of participants agreed it would improve search result quality to have their clicks recorded—a high albeit below average proportion.

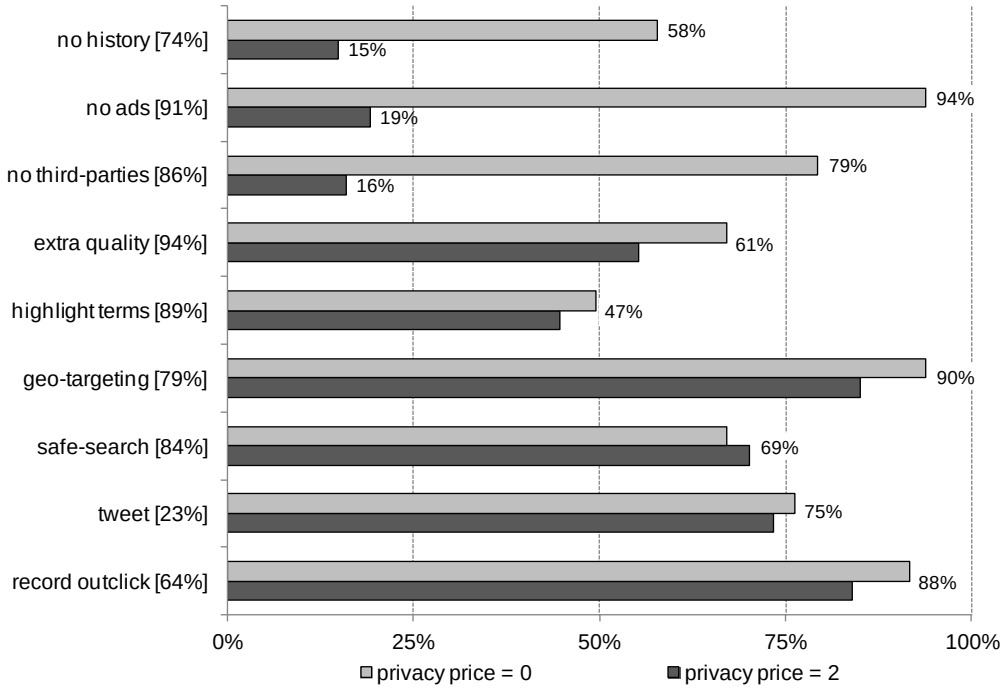

Figure 3: Proportion of users who enabled a given search option at least once, by treatment. The number in brackets in the label gives the percentage of participants who consider this search option important.

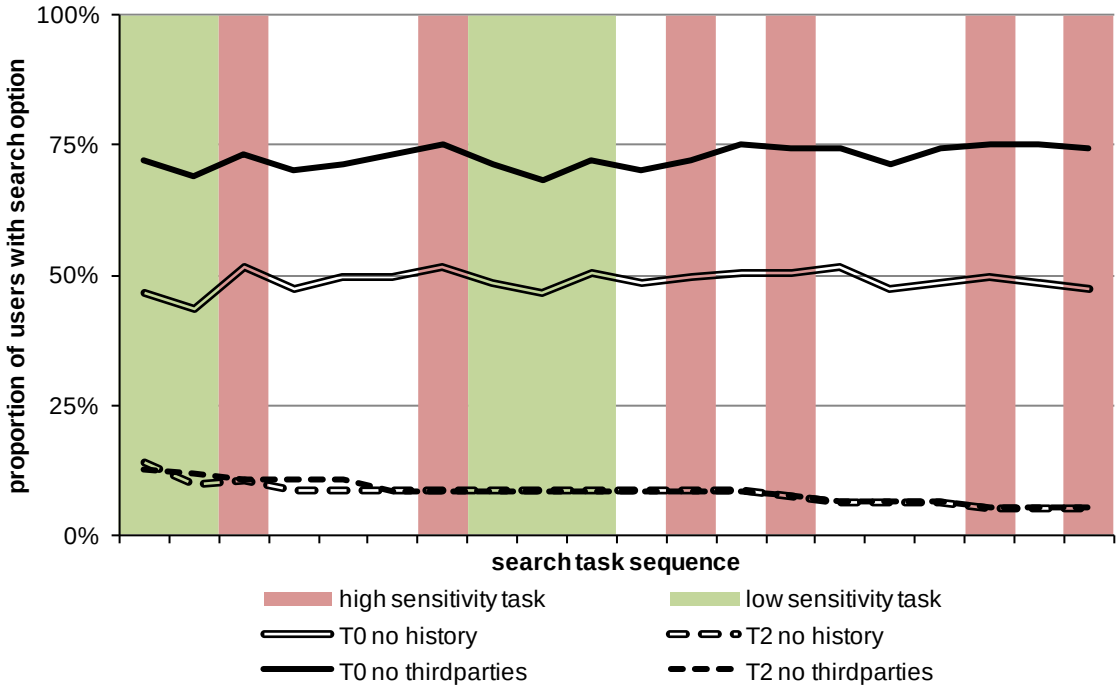

Figure 4: Proportion of users who enabled a privacy-enhancing option, by treatment and by search task (sequentially from left to right from the 1st to the 20th search task). Search tasks with low/high sensitivity are highlighted; the exact wording is given in Table 1.

4. Analysis

This section returns to the research hypotheses (Section 2.1).

4.1. H1: The price of privacy-enhancing features and the proportion of users enabling them are negatively associated. (supported)

In comparing the adoption of the perceived privacy-enhancing search options (no history, no third-parties) between the treatments T0 and T2, one notices a sharp drop when the search option is charged at 2 credits rather than provided for free ($p < 0.0001$, Fisher's exact test). Hypothesis H1 is thus supported.

The proportion of participants who enabled the no history option almost quartered when the price increased from zero to two credits. For the no third-parties option, the drop is even more pronounced (63 percentage points). Interestingly, the introduction of a price for privacy takes the variance out of the different adoption ratios for privacy-enhancing options. Whilst T0 exhibits a spread between 58% and 79% for no history and no third-parties resp. ($p = 0.002$, Fisher's exact test across the two options), this flattens to 15% and 16% ($p = 1$). It seems the fee has reduced to demand for privacy to those truly interested in it. One could estimate that one in seven consumers are privacy-conscious Web search users.

4.2. H2: The more sensitive the search task, the more likely users will enable privacy-enhancing features. (supported)

Search tasks are divided by whether or not they pertain to the high sensitivity group. In T0, queries corresponding to high sensitivity tasks are issued with privacy-enhancing options at a significantly higher rate ($p = 0.003$ for no history; $p = 0.01$ for no third-parties; both G-test of independence). Hypothesis H2 is thus supported.

When privacy-enhancing options are charged at two credits per query (T2), higher query sensitivity no longer results in users taking more privacy-protective action.

4.3. H3: The more sensitive the search task, the less likely users will enable privacy-invasive features. (supported)

In T0, queries corresponding to high sensitivity tasks were tweeted significantly less often ($p = 0.03$, G-test). Hypothesis H3 is thus supported.

4.4. H4: Users who are more concerned about privacy will enable privacy-enhancing features more often. (not supported)

The privacy concern as measured in the exit-questionnaire is not systematically associated with users' propensity to enable privacy-enhancing search options. Neither in T0, nor in T2, is there a significant relationship between high privacy concerns and enabling the no history or no third-parties options. Also, there is no impact of privacy concern on users' tweeting behaviour.

Interestingly, a post-hoc analysis reveals that disabling advertisements is significantly associated with high privacy concerns ($p = 0.02$ in T0, $p = 0.04$ in T2; Fisher's exact test). These findings not only corroborate earlier work on the discrepancy between users' self-professed privacy preferences and their actual behaviour [**4**] [**5**], but also provide valuable insights into which aspects of privacy concerns the Smith et al. instrument is measuring.

4.5. H5: Users who consider privacy-enhancing features more important will enable them more often. (partially supported)

Participants who stated that the no history option is important to them are not significantly more likely to enable this feature during their search. For the no third-parties option, there is a significant effect only for T0 ($p = 0.01$, Fisher's exact test).

5. Summary and conclusions

The status quo regarding privacy in Web search looks as follows: consumers enjoy Web search free of charge. The service is financed by exploiting their personal data, but users are uncomfortable with this practice. However, mainstream Web search engines do not currently offer privacy-enhanced subscriptions whereby users would pay with small money rather than their personal details.

This paper reports on the first user study into consumers' valuation of privacy in Web search. A laboratory experiment with 189 participants has delivered a number of key insights.

- First, privacy-enhancing search features are important to users and universally appreciated. Amongst the participants, 86% indicate they would like to see a feature that prevents data sharing with third-parties and 74% find it important to remove queries from their search history.
- Second, when these features are available, 58% of Web search users profit from the ability to selectively keep certain queries out of their search history. Almost four in five users turn on the feature that disables data sharing with third-parties.
- Third, there are 15% to 16% of consumers who turn on privacy-enhancing features in a search engine even if they are priced at pay half a penny per query.
- Fourth, much higher proportions of users were willing to incur an invasion of their privacy for a monetary reward of the same amount: 88% agree to have recorded on which search results they click and 75% opt in to having their search queries published on Twitter. These findings confirm earlier results on the divergence between a willingness to pay for privacy and a willingness to accept compensation for a privacy invasion [**10**].
- Fifth, users are significantly more likely to turn on privacy-enhancing search options when issuing sensitive queries. Still, neither users' privacy concerns nor their stated importance of a privacy-enhancing option are systematically associated with their privacy choices in Web search.

The experiment at hand provides empirical evidence on the contextual value that consumers place on their privacy when searching online. It complements earlier studies on the willingness to pay for privacy during Web shopping. Further research opportunities unfold in contrast between a high frequency activity that happens in passing such as search, and more singular events that involve higher investments of time and money, such as online purchases.

Being the first experimental study into users' behaviour when managing their privacy in Web search, it has uncovered limitations in the experiment design that future work should address. Firstly, additional investigation should be aimed at the pricing structure for different search options: changing from a pay-per-query model to subscription plan (e.g., over the entire lifespan of the experiment), making participants pay for search options with out-of-pocket money (e.g., the search engine becomes a slot machine), and testing positive prices for privacy-invasive options (e.g., users have to spend credits to tweet their search). Further improvements in the experiment design should address last-round effects, whereby participants' willingness to spend on privacy declines as the experiment progresses.

Another, deliberate limitation is the prescription of search tasks, although users are free to formulate their own queries. By keeping the search tasks constant across all participants, the laboratory provides a controlled environment and allows establishing findings with high internal validity. It needs to be complemented by field experimentation that taps into the diversity of consumers and their search interests and behaviours in the wild.

Acknowledgements. Kat Krol (UCL) provided essential local support with the experiment. The experiment was supported by Technical University of Madrid – Centre for Applied ICT Research.

References

[1] Alexa Internet, Inc. (2012) Alexa Top 500 Global Sites. [Online]. http://www.alexa.com/topsites

[2] ARD/ZDF-Onlinestudie. (2012) ARD - ZDF Onlinestudie: Genutzte Anwendungen. [Online]. http://www.ard-zdf-onlinestudie.de/index.php?id=onlinenutzunganwend0

[3] Michael Barbaro and Tom Zeller, "A Face Is Exposed for AOL Searcher No. 4417749," *New York Times*, Aug. 2006. [Online]. http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=0

[4] Bettina Berendt, Oliver Günther, and Sarah Spiekermann, "Privacy in e-commerce: stated preferences vs. actual behavior," *Communications of the ACM*, vol. 48, no. 4, April 2005.

[5] Alastair R. Beresford, Dorothea Kübler, and Sören Preibusch, "Unwillingness to pay for privacy: A field experiment," *Economics Letters*, vol. 117, no. 1, pp. 25--27, 2012.

[6] Alex Braunstein, Laura Granka, and Jessica Staddon, "Indirect content privacy surveys: measuring privacy without asking about it," in *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS)*, 2011.

[7] Jordi Castellà-Roca, Alexandre Viejo, and Jordi Herrera-Joancomartí, "Preserving user's privacy in web search engines," *Computer Communications*, vol. 32, no. 13-14, pp. 1541--1551, 2009.

[8] CMU Usable Privacy and Security Laboratory. PrivacyFinder - Frequently Asked Questions. [Online]. http://www.privacyfinder.org/faq

[9] Google Inc. (2012) Ten things we know to be true – Company – Google. [Online]. http://www.google.com/intl/en/about/company/philosophy/

[10] Jens Grossklags and Alessandro Acquisti, "When 25 Cents is too much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information," in *WEIS*, 2007.

[11] Ixquick.com. Ixquick Protects Your Privacy! [Online]. https://www.ixquick.com/uk/protect-privacy.html

[12] Nicola Jentzsch, Sören Preibusch, and Andreas Harasser, "Study on monetising privacy. An economic model for pricing personal information," European Network and information Security Agency (ENISA), 2012.

[13] Rosie Jones, Ravi Kumar, Bo Pang, and Andrew Tomkins, ""I know what you did last summer" -- query logs and user privacy," in *CIKM*, 2007, pp. 909--913.

[14] Christian Krause. (2008, July) First European Privacy Seal Awarded. [Online]. https://www.european-privacy-seal.eu/press-room/press-releases/20080714-europrise-press-release-en.html

[15] Yehuda Lindell and Erez Waisbard, "Private Web Search with Malicious Adversaries," in *Privacy Enhancing Technologies*, vol. 6205, 2010, pp. 220--235, Lecture Notes in Computer Science.

[16] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke, "Information Privacy: Measuring Individuals' Concerns about Organizational Practices," *MIS Quarterly*, vol. 20, no. 2, pp. 167-196, 1996.

[17] Danny Sullivan. (2009, Dec.) Google Now Personalizes Everyone's Search Results. [Online]. http://searchengineland.com/google-now-personalizes-everyones-search-results-31195

[18] TNS Opinion & Social, "Attitudes on Data Protection and Electronic Identity in the European Union," Special Eurobarometer 359, 2011.

[19] Janice Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti, "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," in *WEIS*, 2007.