Your Browsing Behavior for a Big Mac: Economics of Personal Information Online

Juan Pablo Carrascal*
Universitat Pompeu Fabra
jp.carrascal@upf.edu

Christopher Riederer*
Columbia University
mani@cs.columbia.edu

Vijay Erramilli
Telefónica Research
vijay@tid.es

Mauro Cherubini†
Telefónica Research
mauro.cherubini@acm.org

Rodrigo de Oliveira
Telefónica Research
oliveira@tid.es

ABSTRACT

Most online service providers offer free services to users and in part, these services collect and monetize personally identifiable information (PII), primarily via targeted advertisements. Against this backdrop of economic exploitation of PII, it is vital to understand the value that users put to their own PII. Although studies have tried to discover how users value their privacy, little is known about how users value their PII while *browsing*, or the exploitation of their PII. Extracting valuations of PII from users is non-trivial – surveys cannot be relied on as they do not gather information of the context where PII is being released, thus reducing validity of answers. In this work, we rely on refined Experience Sampling – a data collection method that probes users to valuate their PII at the time and place where it was generated in order to minimize retrospective recall and hence increase measurement validity. For obtaining an honest valuation of PII, we use a reverse second price auction. We developed a web browser plugin and had 168 users – living in Spain – install and use this plugin for 2 weeks in order to extract valuations of PII in different contexts.

We found that users value items of their online browsing history for about €7 ($\sim$ 10 USD), and they give higher valuations to their offline PII, such as age and address (about €25 or $\sim$ 36 USD). When it comes to PII shared in specific online services, users value information pertaining to financial transactions and social network interactions more than activities like search and shopping. No significant distinction was found between valuations of different quantities of PII (*e.g.* one vs. 10 search keywords), but deviation was found between types of PII (*e.g.* photos vs. keywords). Finally, the users' preferred goods for exchanging their PII included money and improvements in service, followed by getting more free services and targeted advertisements.

Categories and Subject Descriptors

H.1.1 [**Systems and Information theory**]: Value of information; J.4 [**Social and Behavioral sciences**]: Economics

*Research conducted while working at Telefónica Research.

†The author is currently working at Google Inc.

WWW 2013, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2035-1/13/05.

Keywords

Privacy; Economics of Personal Information; Experience Sampling; PII

1. INTRODUCTION

A large part of the Internet economy operates by being reliant on online advertisements. In recent years, targeted advertising has become an attractive offering where targeting is facilitated by the collection of large amounts of personally identifiable information (PII) of end-users. However, this collection comes at the cost of erosion of privacy of end-users. Web service providers are collecting more PII about the end-users, often *outside* the scope of their application (*e.g.*, search engines collecting browsing information via third party aggregators like Doubleclick *etc.* [36]). At the same time, users are becoming more aware of various privacy breaches [4, 38, 42], attracting the attention of regulatory bodies [41].

The economics of the online ecosystem can be summed up by the pithy adage 'if you are not the consumer, then you are the product', more specifically, the product being end-users' PII. In such an arrangement, it is easy for service providers to attach a value on each users' PII, based on the revenues they can extract. However, for users to perform a cost-benefit analysis, where the cost is loss of privacy, and the benefit is the service they obtain in return, it is important that they first know the value of *their* PII they are trading away.

There has been a lot of work on users valuating their information [9, 10, 18], and in general users' perceptions about privacy [2, 3, 14]. However, there has been surprisingly little to no work on valuating *web-browsing* information, even though it is known that privacy leakages can occur while web-browsing [26, 36]. In this paper, we focus on understanding the value that users attach to their own PII,[1] while web-browsing.

It is challenging to extract the value that users' put on their own PII. The valuation could change based on *context*.

[1]We focus on monetary value assigned by the user to their information, although one can imagine other notions of value and utility like satisfaction etc. We consider money as we are interested in the overall ecosystem of online services that partly hinges on monetizing PII. Secondly, money is a tangible concept and easier to arrive at as opposed to user happiness. We will consider other notions of value in future work.

For instance, the value that a user puts on the fact that she is searching for a restaurant can be different than when she is searching for cancer drugs. Even using the same keywords while searching, but in a different context, could lead to different valuations of the same PII (*e.g.* searching for leisure while at home or at work). Past work done in this domain has included valuating personal information (*e.g.*, weight, age [18]) as well as location information [10]. However they all rely on surveys that do not leverage contextual factors when the PII was generated and/or released.

In order to leverage these contextual factors, we rely on the refined Experience Sampling methodology (rESM) [7] (Sec. 3.1.1). This data collection approach probes users at appropriate times to obtain more reliable answers, as questions are presented to users in-context and hence minimizes retrospective recall and possible errors that come with such recall. We implemented rESM by means of a browser plugin (Secs. (3.3, 3.1.2)). Users get asked specific questions when they access different types of content/services (social networks, search engines, finance sites, etc.). We recruited 168 participants living in Spain with a diverse range of demographics (Sec. 3.2), and had them participate in our study for 2 weeks. We used a reverse second price auction to obtain an honest valuation for different types of PII (Sec. 3.1.4). We also use our methodology to obtain users' perceptions and awareness of the economic usage of their PII by online service providers (Sec. 4.2).

The major findings of this work are:

- Users value PII related to their *offline* identity (age, gender, address, economic status) at about €25 (∼ 36 USD), and this value does not change when the user is probed in different contexts (*e.g.* browsing search sites, webmail, etc.).
- Moreover, users value PII related to offline identity higher than PII related to browsing activity, which is about €7 (∼ 10 USD)[2].
- In terms of valuating service specific PII (*e.g.* photos uploaded to social networks, search keywords, online purchases, etc.), users gave higher valuations to interactions in online social networks (€12 or ∼ 17 USD) and finance web-sites (€15.5 or ∼ 22 USD), when compared to activities like search (€2 or ∼ 3 USD) and shopping (€5 or ∼ 7 USD).
- The majority of participants in our study were aware that their PII is being collected when web-browsing, and while they were positive about their PII being used to improve services, they were also negative that it could be monetized by service providers.
- Our results reveal that users prefer to trade their PII for monetary rewards or improved services more than trading it for additional free services or targeted advertisements.

2. RESEARCH CHALLENGES

The work presented herein aims at answering the following two research questions:

- **RQ1:** What monetary value do users assign to different types of PII[3] while being online?
- **RQ2:** What are the perceptions of users vis-a-vis their PII being monetized, improving existing services and for personalized advertisements?

In order to answer these questions, it is of great importance to consider a user-centric approach. Previous work addressed related questions and using techniques such as post-study surveys and diaries [6, 30]. These traditional methods could have drawbacks when trying to gather answers for the questions posed above. For instance, consider a web user Alice who browses the web on a daily basis. On a given day, Alice searches for symptoms pertaining to an illness she suspects she has. Alice then sends an email to her friend Bob about this illness. Some time later, she takes part in a traditional survey and/or diary study that aims at answering the aforementioned research questions. These techniques would most likely not collect accurate responses due to a number of reasons, including:

1. Retrospective recall: Self-report recall surveys and diaries suffer from recall and selective reporting biases [20, 8]. Alice may not be able to remember what she searched for some time ago, or what emails she exchanged. The greater the time that has passed since these actions occurred, the harder it is for Alice to remember and report them accurately in a survey/diary study.
2. Validity: Alice's valuation of the illness related keywords also depends on the context when she shares her PII (*e.g.* how, where and when she came up with the keywords in the first place). So even if Alice is later given keywords related to illness and asked to valuate, she might not remember or recreate the conditions she had when she came up with the keywords and may end up assiging an incorrect value to them.
3. Burden: Alice can be asked to note down her activities and assign values to different PII in a diary. This is however burdensome for Alice, who might even consider dropping out of the study.
4. Honesty: If Alice needs to valuate several different PII in a long survey or in a daily diary study, she can get disengaged and provide random values just for the sake of getting the job done.
5. Engagement: In order to address response fatigue and have Alice valuate information under a diverse set of conditions as accurately as possible, we need her to be motivated. Answering multiple survey questions could lead to a significant number of drop-outs if the study does not include an element of engagement.

Next, we present the methodology of our study describing *how* we tackled these challenges towards providing trustful results for our research questions.

[2]Equivalent to a Big Mac meal in Spain, circa 2011. Hence the title of this paper.

[3]A strict definition of PII does not include browsing behavior. However it has been shown that information leaked on the web via browsing can be combined to form PII [27], so we use PII to refer to all the information that a user can leave online, knowingly or unknowingly.

3. METHODOLOGY

3.1 Tackling Challenges

3.1.1 Users' Need for Recall

In order to address the challenge of users' *retrospective recall* for PII valuation, we use a refined version of the Experience Sampling Method (rESM). Experience Sampling involves asking participants to report on their experiences at specific points throughout the day. The method was originally developed in the psychology domain [5] and recently adapted successfully in many studies of Human-Computer Interaction [8, 19, 20, 29]. As Cherubini *et al.* highlighted [7], the main advantage of ESM is its ability to preserve the ecological validity of the measurements, defined by Hormuth *et al.* as "the occurrence and distribution of stimulus variables in the natural or customary habitat of an individual" [17]. This method is better than recall-based self-reporting techniques by "probing" the participant in close temporal proximity to when a relevant event was produced. One of the drawbacks of the method is that participants often are sampled at random times and therefore the probing might be invasive for many participants. This is why in recent years some researchers have proposed to refine the method by modeling the participants' context [7, 12], and this is what we use.

3.1.2 Validity of PII Valuations

As a means to perform rESM and further address the challenge of *validity* of valuations, we instrumented the web browser of participants with a plugin that was able to log the website they were browsing and probe them at the exact time a certain PII was being shared online. At a high level, the study operated as follows. First, participants installed the plugin and browsed as usual. Then the plugin would categorize every website the user would browse into one of the eight categories: Email, Entertainment, Finance, News, Search, Shopping, Social, and Health. These categories closely correspond to the eight popular categories that online ad-networks like Doubleclick[4] use, as we are interested in the monetary aspect of PII. In addition, the plugin was able to sense when the user was changing context and use this information to trigger a popup, which would have two goals: (i) collect the user's valuation of specific PII related to the category of the site the user is browsing, via an auction and (ii) inquire the user about perceptions of PII usage. Finally, the popup would send this data to a remote server for data analysis.

3.1.3 Engagement and User's Burden

With respect to preventing the user's *burden*, we adjusted the frequency of the popups triggered by the browser plugin and also allowed users to skip them if they wanted to. In order to provide users with an element of *engagement* to participate actively in the study, we created a real setting where participants could trade their PII for money based on their own valuations. More specifically, participants received the winning monetary value of every auction they won.

[4]Doubleclick has more than eight major categories and more than 600 subcategories. We chose eight as a good trade-off between obtaining detailed information without annoying the user, given that the rESM probing would increase linearly with the number of categories.

3.1.4 Honest PII Valuations

In order to persuade participants to provide an honest valuation of their PII, we relied on a reverse second price auction: given a set of k bids, pick the lowest bidder as the winner, and pay that person the amount equivalent to the second lowest bid. We chose this auction mechanism for the following reasons: (i) this mechanism has the strong property of being truth telling; the best strategy for participants in the auction is to be honest about their valuation [24], (ii) it has been used before for valuating location information [10], and (iii) it is relatively easy to explain.

We allowed positive amounts, including 0, with increments of one cent. We also gave the user a choice to not participate in the auctions at all. This was necessary to cover cases where users felt overwhelmed with participation and cases where users did not even want to disclose the fact that their PII was worth a very high amount. In order to reinforce the notion that the user would indeed part with their PII if they won, we had the user verify that they understood their data would be sold in a second popup. We ran an auction whenever we had 20 bids per category from 20 different users. We considered this amount of bids to provide an adequate tradeoff between a lower bound on the number of participants to create competitiveness and an upper bound on the number of participants bids that would be feasible to obtain within a reasonable amount of time. Multiple auctions were run during the study.

All winners of the auctions were notified by email with information including their winning bid, contextual information of the bid (date and time of bid, PII, website they were on). We reinforced the message that as they won, we would use their PII (showing the exact PII they bid on), for a period of 6 months. Likewise, we sent a similar email to the remainder participants, conveying that as they lost the bid, their PII would *not* be used. Only after the end of the study we informed participants that their PII was actually not going to be used for any commercial purposes. For all our communication with users, we used neutral language with regards to privacy, so as to not prime them one way or another, following the findings in Braunstein *et al.* [6].

3.2 Participants

Participants were recruited using a survey published via a major Web portal in Spain, that attracts a very diverse set of users. A total of 168 participants (93 male, 55%) installed the Firefox browser plugin and completed all requirements of the study. All participants were users of the Firefox browser and hence had it installed on their computer. Participants' age ranged between 18 and 58 years old ($\bar{x} = 31.83$, $s = 8.15$). With respect to their educational level, 1% did not finish primary school, 8% finished primary school, 14% did secondary school, 75% had a university graduate degree, and 2% a post-graduate degree. Socioeconomic status was also diverse: 28% of the sample said their annual gross salary to be lower than €10K, 25% said it was between €10K and 20K, for 22% it was in the range of €20K and 30K, 11% between €30K and 40K, and 10% reported earning more than €40K per year (4% preferred not answering this question). All participants lived in Spain and the vast majority were of Spanish nationality (94%). Each participant was given a gift card voucher worth €10 ($\sim$ 14 USD) for taking part in the whole study. Our ethical board and legal department approved the experiment. Participants were debriefed about what was being logged and instructed on how to temporarily disable or remove the plugin. Participants were free to leave the experiment at any time.

3.3 Apparatus: Browser plugin

In order to capture the browsing context of the users we developed a system consisting of two parts: a browser plugin – to be installed in participants' browsers – and a web server that communicated with the plugin, sending configuration information and receiving data from it.

Firefox Plugin: The plugin had three main tasks. First, it captured and stored all browsing activity of the user. This consisted of the url, time of page access, and a unique ID we assigned to each browser. This data was stored on the local machine and sent to the server at regular intervals.

The second main task of the plugin was to categorize visited websites into one of the eight categories mentioned in Sec. 3.1.2. In order to do this, we relied on a hard-coded list of 1184 popular sites from different categories for Spain, gleaned from `alexa.com`. Although some popular sites like Facebook can host content pertaining to health or entertainment, we hard-coded it to Social.[5] For sites that were not on Alexa, we resolved them into categories in real-time by relying on an approach implemented in another browser plugin called Adnostic [40]. The basic idea is to perform a cosine similarity between the set of keywords present on the site the user visits and a massive corpus of words that are associated with specific categories. The category with the highest similarity score is used and the appropriate text is presented in the pop-up. Testing on individual unclassified and Alexa-classified websites gave a high level of accuracy (approx. 98% correct classification).

Third, the plugin presented the participants with two independent popups, as described earlier. The first popup displayed auction questions and the other displayed questions related to exploitation of PII. These were configured to be switched on or off from the server. From a UI perspective, the popup displayed the text of relevant auction question, with the type of PII in the auction in bold text, to highlight what is actually being traded in the auction. There was a box below the text where the user could enter an amount, and there was a radio button below the box where the user could select to not participate in the auction. Fig. 1 shows an example of a popup for category Social.

Server: We developed a simple, highly responsive webserver that the browser plugin would sync with at regular intervals. The server accepted data (bids, responses to questions) from the plugin and stored it in a database. The main function of the server was to run auctions. For each category of website, and for each type of PII (there were 4 types per category, as explained in Section 3.5, questions a1–a4), we set an auction to run once 20 bids were in. We pooled all these auctions, ran them once every morning, and sent out results to participants via emails. The entire process was automated.

[5]Such a monolithic categorization does have limitations; large service providers like Facebook or blogspot host content belonging to multiple categories. However, we consistently pick the first category as put out by Alexa. This ensures that we do not have any false positives – Facebook will always be categorized as Social. In addition, the questions we pose users (Table 1) for a certain category are always consistent; questions on Facebook are always related to Social. We leave a detailed categorization mechanism to future work.

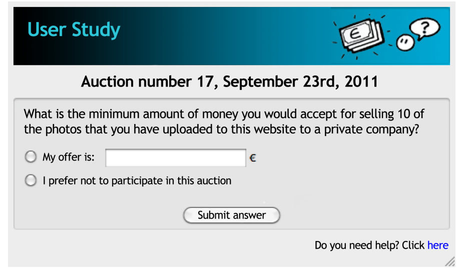

Figure 1: The auction popup. Each auction game was identified by a sequential number and a date. The participant had the option to either enter a bid or to not take part in the auction.

3.4 Procedure

We ran the study in September of 2011. After following our study advertisement on a famous Web portal and signing up for the study, participants were selected based on our unique filtering criteria – users of the Firefox web browser – and invited via email to participate in our study. We asked participants to fill a recruitment questionnaire which focused on demographics as well as their general Internet privacy knowledge and perception. We explained to participants that the study consisted of three phases: (1) an initial week where the popups were inactive, and their browsing behavior would be collected, (2) the actual study that lasted two weeks where popups were active, and (3) the final questionnaire.

Phase 1. During the first week, the plugin silently recorded the browsing behavior of participants (with their consent). The information captured during this phase was used to record a user's baseline browsing behavior. We used this information to make sure that our popups were not interfering with the normal browsing behavior of the participants. We extracted the frequency distribution across the visited sites for every user. We will refer to this as the user's *fingerprint*.

Phase 2. During the experiment, the plugin displayed popups when the participants were browsing the internet. The popups contained two kinds of questions: questions about valuing PII – used as the basis for the auctions – and questions on participants' perceptions and knowledge of PII. To avoid the popups being too invasive the plugin displayed at most one popup per category per day. Additionally, there was a *minimum* delay of 10 minutes between any two popups.

Phase 3. At the end of the experiment, we asked participants to fill in a post-study questionnaire that aimed to clarify the main results obtained during the study.

3.5 Measures

Table 1 summarizes the questions that we presented to the user during the entire study, along with their associated measures. We customized popup questions (a1–a4 and p1–p4, described below), according to the *context* they were asked in. Questions about PII unrelated to the website currently being visited are *context independent*. For instance, a question about a user's age while at a news website could

Table 1: Questions asked during the different phases of the study.

Code†	Question	Remarks	Type
r1	**Are you concerned about protection of your private data on the Internet?** [5- A lot … 1-Never]		5-point
r2	**Do you distrust the way the websites you visit use your data?** [5- I distrust of all of them … 1- I do not distrust]		5-point
r3	**Do you read the privacy policies of the web sites that you visit?** [5- Always … 1- Never]		5-point
r4	**How much do you know about current legislation about data protection?** [5- A lot … 1-Nothing]		5-point
a1	**What is the minimum amount of money you would accept for selling to a private company information about your age, gender, salary and address?**	Context independent	Numeric
a2	**What is the minimum amount of money you will accept for selling to a private company details about your presence on this webpage?**	Context dependent	Numeric
a3	**What is the minimum amount of money you would accept for selling * to a private company?**	Context dependent, category-specific	Numeric
a4	**What is the minimum amount of money you would accept for selling 10 * to a private company?**	Context dependent, category-specific	Numeric
p1	**Are you aware that the web site you are currently visiting might generate revenues from the information ** ?** [5- I was fully aware, 4- I did know, 3- I was not fully aware, 2- I figured but I was unsure, 1- I did not know]	Context dependent, category-customized**	5-point
p2	**How comfortable do you feel knowing that the web site you are visiting might generate revenues with the information you share?** [5- Very comfortable, 4- Comfortable, 3- I do not care, 2- Uncomfortable, 1- Very uncomfortable]	Context independent	5-point
p3	**If the company that uses this information does it in order to offer you a better service, how would you feel?** [5- Much better, 4- Better, 3- The same, 2-Worse, 1- Much worse]	Context independent	5-point
p4	**If the company that uses this information does it in order to present you with customized advertisements, how would you feel?** [5- Much better, 4- Better, 3- The same, 2-Worse, 1- Much worse]	Context independent	5-point
f1	**On day "X" you bid "Y" Euros for sharing "Z", and that was your lowest bid. What was your main motivation for bidding that low?** [Answers manually categorized as: 1- To win the bid, 2- Not important information, 3- Fair value, 4- Other] **On day "X" you bid "Y" Euros for sharing "Z", and that was your highest bid. What was your main motivation for bidding that high?** [Answers manually categorized as: 1- To win the bid, 2- To prevent selling, 3- Fair value, 4- Other]	Letters X, Y and Z were replaced by the actual bid date, bid value, and traded PII respectively	Text
f2	**I do not see a problem if *** generate revenues from my personal information as long as they:** [1. pay me some money, 2. improve their existing service, 3. provide more free services, 4. recommend me things that I like, 5. no need to change what they are currently doing, 6. other]	Category-customized***	Nominal
f3	**With respect to the services offered to you by ***, do you believe they have significant operational costs?** [1. Yes, 2. No, 3. I do not know] **With respect to the services offered to you by ***, do you believe they have significant revenues?** [1. Yes, 2. No, 3. I do not know]	Category-customized***	Nominal
f4	**In the case of having a market that you could sell your personal information (*e.g.,* clicks on a website, history of pages visited, contact details, bank account details, *etc.*), who would you trust to handle that information?** [Options to rank: Telecommunication company, Government, Bank, Insurance company, Yourself only, Other]		Ranking

† Codes refer to different phases of the study: 'r' stands for recruitment questionnaire; 'f' stands for final questionnaire, 'a' stands for auction game, and 'p' stands for perception of PII monetization.

* Customized per category: Email: "data about one of the contacts that you email more often"; Entertainment: "that you have visited this web Site"; Finance: "details about your last financial transaction"; News: "the last news or articles that you read"; Search: "the words that you used in your last search"; Shopping: "details about the last product or service that you bought online"; Social: "one of the photos that you have uploaded to this web site"; and Health: "details about the last time you were sick". Auction question a4 is similar to a3, but increases the number of PII items by 10.

** Customized per category: Social: "you share with your friends"; Entertainment: "you share when you fill its forms"; Health: "you are looking for here"; Search: "your search history"; Finance: "about your finance might be shared with other companies"; Email: "the content of your email messages"; Shopping: "your shopping behavior"; News: "your news reading history".

*** Customized for 3 categories: Social: "online social networks"; Email: "e-mail providers"; Search: "search engines (e.g. Google, Yahoo, Bing, etc.)".

be considered context independent. Conversely, questions about PII that *are* related to the current website are called *context dependent*. For instance, financial transactions on a banking website. Additionally, the content of some questions was customized according to the category of the website they refer to, as explained in Table 1.

Recruitment (r1–r4). Questions in the recruitment questionnaire aimed to gauge participants' knowledge of privacy related issues.

Privacy Valuation (a1–a4). These questions were presented to participants with a plugin popup during the auctions, and asked them to bid the *minimum* value they would accept to sell their PII. We were deliberately vague about *how* we were going to use their PII for two reasons: (i) to realistically reflect the conditions that exist today, as there is little knowledge of how one's PII is being used (targeted advertisements, price discrimination [31]), and (ii) not to bias the user by providing a specific use case of their PII; for instance using PII for behavioral targeting can be construed positively or negatively.

Question a1 is context independent. Its purpose was to assess the validity of our measures by contrasting with results from a2. Indeed, a2 and a3/a4 were context dependent, but while the former asks about the same PII item across categories, the latter is customized for each category of websites. Our goal was not to produce generalized estimates of context valuation but rather to understand whether online context had an influence on the valuation that people attach to certain types of PII. Furthermore, we chose to ask a2 as this is the information that most entities engaged in large scale tracking across the web (like DoubleClick) have access to, and hence can monetize. These are often referred to as 'third' parties. Questions a3-a4 are category specific and in most cases, this PII is available only to the service provider actually providing that service (*e.g.,* photos on social networks, financial transactions, online purchase history, *etc.*). These are referred to as publishers or 'first' parties.

Privacy Perception (p1–p4). These questions were also presented with a plugin popup, and were designed to understand if users are aware of monetization of their PII by online entities.

Exit (f1–f4). These questions were asked in the final questionnaire in order to clarify results obtained during the study.

3.6 Statistical Analysis

Nonparametric analysis was applied to the data considering the ordinal nature of some observed variables and that continuous variables did not follow the normal distribution. Given that participants browsed web pages in their natural environment without being forced to visit sites from all categories mapped in our study – thus preserving ecological validity, our sample had several missing values across categories. Removing subjects that did not provide information for all categories – as they did not browse all types of web pages – would significantly reduce the generalization power of our results and yield unrealistic findings based on the assumption that everybody browses web pages from all categories considered in this study. Therefore we opted for *not* using related sample analysis. Hence differences between median bid values (or Likert scale measures) across categories were tested using the Kruskal-Wallis test and the Mann-Whitney test whenever appropriate. Associations between ordinal/interval variables were assessed using the Spearman's Rho test. Comparisons between related sample distributions of dichotomous variables were performed using both the Cochran's Q test and the McNemar test. The level of significance was taken as $p < .05$.

4. AUCTION AND SURVEY RESULTS

We summarize the main results obtained towards addressing our two research questions. Our results are mainly reported in Euros, and at the time the conversion was approx. €1 gives 1.42 USD.

4.1 Results for RQ1: Monetary value of PII

Effect of pop-ups on browsing behavior. We used the $L2$ distance between participants' first week's "baseline" fingerprints and their fingerprints for the second week of the study after pop-ups were turned on and found little differences (165 users had less than 5% difference). Specifically, only three users (2% of the sample) had high browsing behavior deviation and reported being on vacation during the second week, thus explaining why they used their browser sparsely. These findings indicate that users *did not* deviate from their 'normal' browsing behavior when participating in the study.

Representativeness of categories. We look into the bidding behavior of the whole sample ($N = 168$) while browsing websites as they map to each of the 8 categories and also in relation to the nature of the information being sold (see questions a1-a4 in Table 1). Overall, participants visited websites from all of the eight categories, HEALTH being the least visited category (SEARCH: 82%, ENTERTAINMENT: 82%, SOCIAL: 78%, NEWS: 76%, FINANCE: 75%, SHOPPING: 75%, EMAIL: 64%, HEALTH: 2%). Given the lack of representativeness for the number of subjects visiting health related web pages, we therefore consider only seven categories when comparing participants' bids and other relevant measures across categories.

Bids on context independent PII. With respect to selling their PII that is related to their offline identity (*i.e.*, age, gender, address and salary; see question a1 in Table 1), we found no significant difference among participants' median bid values across categories ($p = .702$). Note that this result was somewhat expected as question a1 was context independent – no mention was made to selling the participants' PI to an entity related to the website they were browsing. The overall median bid value across categories was €25.

Bids on context dependent PII. When probed about selling clicks they performed on a given web page (see question a2 in Table 1), which represents their browsing behavior, participants' median bids were again not significantly different across categories ($p = .569$). In this case, the overall median bid value was €7. Median bid values for highly category specific PII – as captured by questions a3 and a4 in Table 1 – revealed significant differences across categories ($p < .001$). The highest median bid values (in euros) were from categories FINANCE ($\tilde{x} = 15.5$), SOCIAL ($\tilde{x} = 12$), and EMAIL ($\tilde{x} = 6$), with FINANCE similar to the latter two categories ($p = .31$ and $p = .09$ respectively) and significantly different from the remaining categories (SHOPPING = 5, NEWS = 2, ENTERTAINMENT = 2, SEARCH = 2; $p < .001$). Table 2 summarizes the most relevant descriptive statistics of median bid values per category.

Effectiveness of the auction: Categorization of the participants' free text responses to why they bid so low/high in their lowest/highest bids (in euros) was categorized manually with an acceptable inter-rater reliability (lowest bid: $K = .77, p < .001$; highest bid: $K = .78, p < .001$). Even considering the extreme case of each participant's lowest bid, only 15% explained that they bid that low in an attempt to win the auction. The majority said it was because the information was not important ($50 - 51\%$) or they thought it was a fair value ($8\% - 10\%$), or due to some other reason ($25\% - 26\%$). On the other hand, highest bids were mainly due to prevent selling important information ($53\% - 58\%$), although also being explained as a fair value ($16\% - 22\%$) or due to some other reason (22%). Note that there are subtle differences between a 'fair' value assigned to information, and very high/low values assigned because information was very important or not important at all. Fair value indicates a more reasoned approach while bidding very high values indicates focus on the outcome (no selling under any circumstance). Bidding very low values indicates nonchalance; value of information is so low that it isn't worth reasoning about. The fact that only 18 zero – 11 winings bids – bids were placed during the whole auction period is an indication that participants were not bidding just for the sake of winning. Very few participants ($3\% - 4\%$) explained

Table 2: Median bid values per category calculated from participants' median bids in each category (1st and 3rd quartiles shown between brackets

Questions	Email	Entertainment	Finance	News	Search	Shop	Social	All Categories	*p*-value
a1	24.5 [1.6, 97.4]	26.5 [3, 115]	20.2 [3.4, 100]	25 [4, 150]	20 [2.5, 150]	10 [2, 100.2]	15 [3.5, 60]	25 [5.5, 151]	.702
a2	5 [1, 25]	5 [0.9, 20]	3 [1, 20]	5 [1, 43.5]	4 [0.7, 20]	5.2 [1, 30]	7.1 [1, 25]	7 [1, 38]	.569
avg(a3, a4)	6 [2, 89]	2 [1, 14.3]	15.5 [3.8, 229.5]	2 [0, 13.5]	2 [1, 12.8]	5 [1, 20.5]	12 [2, 81.5]	5.5 [1, 39.3]	$< .001$

their highest bid as a strategy to win the auction. These results could indicate that the rules of the reverse second price auction were understood. Overall, the results indicate that the auction scheme is indeed effective for truth telling, given that the majority gave reasons of fair value or worth of information for bids instead of trying to game the system.

Bulk PII effect. We verified no significant difference between the median bid value for all categories in question a3 ($\tilde{x}_{a3} = 5$) and in question a4 ($\tilde{x}_{a4} = 5$, $p = .59$). This finding indicates that the amount of information being sold was not a factor for participants when placing their bids, as they valued one piece of information (question a3) and 10 pieces of information of the same type (question a4), to be used for the same period of time if sold, in a similar way.

Winning bids and pay-outs. Considering the 40 subjects that won at least one auction, their median winning bid was of 5 cents of Euro ($min = 0, \bar{x} = 0.19, max = 2.29$). Even though we allowed a bid of 0 as a valid bid, only seven winners bid 0 on 11 occasions, out of 4000+ bids. The other winners' bids were strictly positive. Finally, as we used the reverse second price auction, the median payout was actually 45 cents of Euro ($min = 0.01, \bar{x} = 0.65, max = 5.69$). We describe this result for completeness.

Trading PII for alternative goods. At the end of the study, we wanted to understand if there were preferred goods participants would be willing to trade their PII for, and if the preferred goods would be different across the most popular categories, *i.e.* SOCIAL, EMAIL and SEARCH (question f2 from Table 1). According to our results, participants' first choice was to either exchange their PII for money (32%−37% of participants) or have improvements in services they are currently using (33% − 37%). The second choice was to receive more free services (14% − 18%). No significant differences were found between distributions of each strategy across the three categories (money: $Q = 2.000, p = .37$, better services: $Q = 1.042, p = .59$, free services: $Q = 1.805, p = .41$). Interestingly, receiving PII-based recommendations was the third option for social networks (7%), but rather the fourth for EMAIL (2%) and SEARCH (4%), with a significant difference between them ($Q = 6.167, p = .046$). Fig. 2 shows a graph comparing the participants' preferred monetization strategies across the three categories.

Relationship between bids, demographics, and privacy. We next looked into significant associations between variables captured in the recruitment questionnaire and the participants' bids. Our findings reveal a medium negative correlation between participants' age and their median bid values for question SOCIAL-a3 ($n = 64$, $\rho = -.276$, $p = .03$). Similarly, age is negatively correlated to the combination of questions SOCIAL-a3 and SOCIAL-a4 ($n = 69$, $\rho = -.287$, $p = .02$), thus providing evidence that the older people are, the lower they tend to bid on photos they share online. Furthermore, we found a medium negative association between

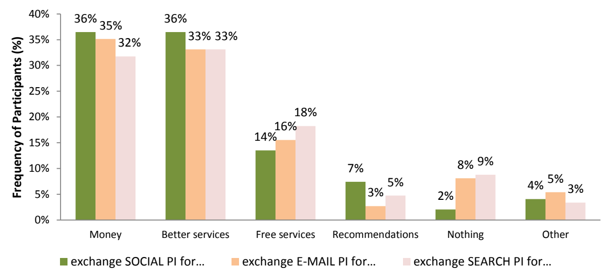

Figure 2: Participants' preferred goods in exchange of PI to online social networks, e-mail providers, and search engines.

gender (1 = male, 2 = female) and median bids for question EMAIL-a3 ($n = 45$, $\rho = -.333$, $p = .03$). This result indicates that men might value information related to their email contacts more than women. Correlations between income levels and bid values were not significant. Finally, we found medium negative correlations between participants' education level and their median bid values for question a2 in most categories (ENTERTAINMENT: $\rho = -.277$, FINANCE: $\rho = -.282$, SEARCH: $\rho = -.235$, SHOPPING: $\rho = -.32$).

We also correlated bid values with responses provided to privacy-relevant questions in the recruitment questionnaire. Positive correlations were found between being worried about online data protection and higher bids on context independent PII (question a1, ENTERTAINMENT: $\rho = .252$, FINANCE: $\rho = .278$, SEARCH: $\rho = .23$).

4.2 Results for RQ2: Perceptions around usage of PII

Results presented in this subsection contribute to the understanding of how users' perceive the economic usage of their PII by online service providers. Note that we considered only the first answers that participants gave to questions p1–p4 per category. This decision guaranteed that their initial opinion would be taken into account instead of a potentially biased opinion due to the effect of long exposure to the study.

Knowledge of PII-based monetization. Participants were aware that PII shared on a particular web site could be used to generate revenue (question p1, $\tilde{x} = 4$, $q1 = 2$, $q3 = 4$). Moreover, no significant difference was found between median ratings across categories ($p = .107$). This finding suggests that knowledge of PI-based monetization is related to Internet services in general and not to a particular set of services.

Comfort with PII-based monetization. In question p2, participants revealed how comfortable they were with web sites extracting revenue out of their PII. With a median rating of 2 ($q1 = 2$, $q3 = 3$), they reported being uncomfortable with it, and this feeling was shared across categories as no significant difference between participants' median ratings per category could be found ($p = .429$). From this finding, we conclude that the act of monetizing from users' PII is what generally makes people uncomfortable, and not the type of online service providers the revenue will go to (*e.g.*, finance, search, *etc.*).

Improving services with PII. Although not comfortable with their PII being monetized, participants pointed out that they would like online companies to improve their web services using their PII (question p3, $\tilde{x} = 4$, $q1 = 3$, $q3 = 4$). No significant difference was found between participants' median ratings across categories ($p = .869$). This finding is consistent with results presented in Fig. 2 about money and improved services being the participants' preferred PII monetization strategies.

PII-based publicity/ads. Finally, subjects were indifferent with regards to online service providers making personalized publicity/ads by using their PII (question p4, $\tilde{x} = 3$, $q1 = 3$, $q3 = 4$). Once again no significant difference could be identified between participants' median ratings across categories ($p = .686$). This finding suggests that leveraging users' PII to provide them with personalized ads generally have neither a negative nor a positive impact on people.

Perception of costs and revenues. Participants were more confident about revenues than costs of providing social network, email, and search services (answered "do not know": 3% vs. 29%, 10% vs. 24%, 6% vs. 21% respectively). In general, most participants agree that these service providers have high revenues (93%, 69%, and 89% respectively) and high costs (43%, 45%, and 53% respectively), but the perception of revenues is significantly higher than costs ($p < .001$ for each of the three categories). Finally, more participants perceived search services to have significant costs compared to email (68% $n = 117$ vs. 58% $n = 113$, $p = .02$), while more participants perceived social network and search services to have significant earnings compared to email (97% $n = 143$ vs. 77% $n = 133$, $p < 001$; 94% $n = 139$ vs. 77% $n = 133$, $p < .001$ respectively). These results reveal that users might consider social network services to be more profitable than search or email services.

5. DISCUSSION

The conclusions that can be derived from our results (Sec. 4) are:

Users value offline PII more and online PII less: If we consider the results for a1, the question on valuating offline PII (Sec. 4), users consistently bid high values for their offline PII like age, gender, address and financial status; pieces of PII that form their offline identity, to trade with online entities. Likewise, users attach lower value to a2, a3 and a4, PII that mostly has to do with their online behavior (a2 is exclusively about browsing history, the other two are about online transactions). Digging deeper, we also note that users tend to value category-specific PII (a3 and a4) on FINANCE and SOCIAL, categories that are more explicitly intertwined with one's offline identity, more than SEARCH and NEWS.

We conjecture that the difference in valuation exists because of lack of awareness. Offline PII is easier to valuate as it is more explicit. It is harder to understand the implications of having your PII continuously tracked, data-mined, and linked to an offline identity [13, 37]. As a consequence, users value such PII less.

Higher valuations than previous studies: Previous studies on valuation of privacy or personal information have reported lower values for various PII than what we encounter in our results [23]. This could be for two reasons. First, we use experience sampling that puts emphasis on valuating PII during web-browsing at the appropriate time, and second, specific properties of the demographics (Spanish citizens) could play a role. We note here that the regulatory framework surrounding privacy in EU is much stricter than in other parts of the world and this could affect the norms related to privacy and personal information of users. We also note that cultural norms can play a role. Addressing these concerns is beyond the scope of this work.

High variation of PII valuation: Bid values show great variance, as it can be seen in the difference between the first and third quartiles of the seven categories (see Table 1). We envision two hypotheses that could explain our participants' behavior that led to this observation: either (i) different PII can largely differ in terms of bidding prices, or (ii) people indeed have very little idea about how much their PII should be valued. Future work will look into this.

Users do not distinguish between quantity of PII, but type: We compared the median bid values for a3 and a4 across categories and found no significant difference. These two auction questions differ only in quantity of information being traded, with the type of PII and the context remaining the same. As reported earlier, there are significant differences between type (FINANCE and SOCIAL being higher than SEARCH, SHOPPING etc.)

We correlated the values with demographic information as well as the responses to the privacy related questions (r1-r4). We found weak to no correlation. A possible conjecture can be on the lines of what is reported in [9], that users factor in diminishing returns of more information in their valuation – although we have no evidence to support or refute this conjecture.

Older users less concerned about online PII: When we correlated bid values against demographics, a high (negative) correlation occurred between age and category specific PII on SOCIAL, ENTERTAINMENT and NEWS, and more so while valuating bulk information (a4). For SOCIAL, this can be linked to the fact most older users do not use online social networks, let alone upload photos to online social networks.[6] This result is in contrast to previous work that stated that older users are generally more concerned about their privacy, while being online [33].

Users do not like monetization of their PII: Users are negative when it comes to their PII being used for monetization by entities (question p2), despite knowing that online entities collect and use their PII for monetization (p1). In addition, they prefer their PII to be used for improving the services they are offered (ap3), across all categories. On the one hand, these results are expected – the former deals with monetization of a good (PII) that users probably perceive as theirs, while users view the latter as a positive outcome of their PII being exploited. In order to understand why users are negative about their PII being monetized, one

[6] http://www.comscoredatamine.com/2010/09/visitor-demographics-to-facebook-com/

can posit that most users are not aware of the functioning of the online ecosystem in place – they do not perceive that the services they get for 'free' (storage in Gmail, Bing search, Facebook etc.) actually are expensive (large datacenters, equipment and bandwidth costs) and while users are aware of their PII being monetized, they are possibly not aware that large parts of that monetization goes towards providing them with a 'free' service.

However, when we look at the results from the post-study questionnaire (question f3), we find that users indeed seem to be aware of the costs and revenues of different services with most users assigning *higher* revenues than costs for services. Taken together, users' negative attitude about monetization of their PII by services can be due to a feeling of unfairness.

Users are indifferent when it comes to the use of the PII to send them personalized ads (p4), again across categories. This is somewhat in contrast to results in [30] where the authors report that 64% of the survey respondents (all Americans) find behavioral targeting invasive. The differences between our results and theirs can be due to cultural differences (our sample consists mainly of people from Spain) and/or methodological differences – we used experience sampling to capture the context, while the results reported in [30] were gathered via traditional surveys.

6. IMPLICATIONS FOR DESIGN AND FUTURE RESEARCH

Our study has direct implications on the monetization of personal information online. As the focus of the study has been towards understanding the *economic* aspects of PII, we believe the findings can help in the future research topics and new offerings. We propose three major implications.

6.1 Incentives for adoption of privacy solutions

A prominent reason for the failure of adoption of most privacy solutions are the lack of proper incentives (economic or otherwise) for various parties to support the adoption [32]. Consider online privacy; on one side there are online service providers who have stated that they want to move up to the 'creepy' line [34] on accessing and using PII, while on the other side users are resorting to unilateral measures like anti-tracking plugins etc. to prevent data collection, hence deterring service providers from supporting such privacy preserving measures.

Recent privacy preserving solutions have been designed to preserve privacy of the users as well as provide means for online service providers to access and monetize PII via targeted ads, thereby preserving the business models of these providers [15, 40]. Based on our findings (Fig. 2) these solutions can have a better chance of adoption if they incorporated some form of economic incentives, by way of monetary compensation to the end-user. Such economic incentives based solutions have been proposed as well [28, 35], with some start-ups going for such a model[7].

The results in this paper provide the first empirical foundation for economic incentives by demonstrating how users value different types of PII for a variety of actions performed while online. The prices can be taken to be the reserve prices[8] that users will be willing to accept to part with their PII. Likewise, we have seen that different types of PII have different valuations (e.g. photos in social networks *vs.* online purchase history). These differences can be used by service providers to strategically target different types of PII. The findings in our paper can be used as inputs to drive models to better understand the ecosystem. For instance, a recent proposal to address privacy breaches using insurance can benefit from our analysis to set premiums [16].

In addition, other types of incentives can also help drive adoption of privacy preserving solutions. If we consider Fig. 2, users also prefer improved services that use their PII. If service providers can convince users that there have been improvements to the respective services and which PII bits went into the improvements, users may be less concerned about their privacy.

We asked participants of our study about who they would trust to handle their PII in the case that an entity enables economic transactions around their PII, in the post-questionnaire. Users trusted themselves more than any other entity ($\bar{x} = 5.2$, $\tilde{x} = 6$, $q1 = 6$, $q3 = 6$, 6-point scale). Government was the second most trusted entity ($\bar{x} = 3.8$, $\tilde{x} = 4$, $q1 = 2$, $q3 = 5$), followed by banks ($\bar{x} = 3.5$, $\tilde{x} = 4$, $q1 = 3$, $q3 = 4$) and telecommunication companies ($\bar{x} = 3.4$, $\tilde{x} = 3$, $q1 = 2$, $q3 = 4$) tied in the third place ($Z = -.299, p = .77$). Finally, insurance companies were considered the least trustful entities for handling people's PII ($\bar{x} = 3.1$, $\tilde{x} = 3$, $q1 = 2$, $q3 = 4$). Trusting oneself with one's PII could point to a totally decentralized architecture for a privacy solution. However, more work needs to be done to verify if users can undertake the burden of dealing with all the transactions around their PII themselves.

6.2 Transparency on monetization of PII

One of the findings reported in Sec. 4.2, is that while users have knowledge of their PII being collected, they are not comfortable about their PII being monetized. This lack of awareness also plays out in valuations – while offline PII and certain types of online PII like photos and financial transactions have high valuations, presence of the user on different sites were valued very low. This is interesting as a behavioral profile can be constructed just by tracking users across sites (via cookies etc.) and this profile can be used to identify users and be monetized [11]. We believe that most privacy concerns that arise are due to lack of awareness of precisely this fact – that PII is being monetized (participants knew their PII could be monetized by entertainment and search related websites, but not for the other categories).

The findings reported in this paper indicate that if online service providers are explicit and up front about the fact that they provide a service (email, video streaming, a social network, etc.) for free and in return collect and monetize PII, along with details on the specific types of PII they collect, the privacy concerns of most users will be tempered. Long privacy policies written in complicated legalese that are seldom effective [21], can be dispensed with.

For example, we can think about agreements that could expose the amount of money required to run the service the user is signing up for and how the revenues generated by exploiting PII help cover those costs. This implication is further strengthened when we factor in that majority of the users perceive that service providers have higher revenues than costs (Sec. 4.2), hence being transparent about costs can help educate users. Additionally, we can have alterna-

[7] `www.personal.com`

[8] http://en.wikipedia.org/wiki/Reservation_price

tive business models where the user has the *option* to pay for the service that s/he is signing up for either with his/her PII or with real money.

6.3 Bulk data mechanism

A final implication for design is related to the indifference in valuation for bulk quantity of data. Specifically, participants assigned a similar value to a certain piece of PII as to 10 pieces of the same information. This has a direct consequence for the design of trading PII. In fact, it does not make sense to implement mechanisms for the trade of a single piece of information. Rather, it makes more sense –according to these results– to design solutions that would allow interested users to trade a bulk amount of PII. For instance, such a mechanism could be presented during registration to a new service and extended for bulk amounts of PII that the user will be sharing throughout the use of the service. The effect of such a design could be two fold: on one hand it would minimize the user's effort and mental load, while on the other hand it would maximize the effectiveness of the service provider's budget expenditure.

7. RELATED WORK

Previous research has shown that valuation can depend on the type of information release. For instance, Huberman [18] reported that valuation of certain bits of PII like weight and age depends on the desirability of those bits of information in a social context. Likewise, valuation of location information has been found to depend on factors like the distance traveled by the user and other factors [9, 10]. Our work differs in multiple regards. First, we focus on *web browsing information* of users that is of economic interest to online services (*e.g.*, search providers, social networks) and such information raises privacy concerns [25, 27]. Second, we study the effects of demographic information like age, gender, education levels and socio-economic factors on valuation of one's PII. While the aforementioned works used mostly surveys to figure out different valuations, we use a methodology based on experience sampling to capture PII context and obtain valuations in-situ. Finally, whereas previous works used hypothetical payments to determine PII valuation [10], we use actual payments, hoping to obtain a more accurate value and have user engagement.

Another body of work that is related has to do with studying the dichotomy that exists between willingness to pay (WTP) to buy privacy protection and willingness to accept (WTA) to reveal PII. A difference between WTP and WTA can be indicative of an *endowment* effect [39]: people can place a higher value on an object that they own, in this case PII. In our paper we do not deal with WTP vs WTA explicitly, instead we focus on extracting WTA for web-browsing, while leveraging contextual factors when PII is generated and/or released.

A majority of the work done on understanding the awareness levels of users in terms of how their PII is exploited and related privacy concerns has focused on how the actual behavior of people deviates from what they state. This deviation has been noted by Jensen & Potts [22] who also found that there is a difference between reported knowledge and reality; in general people do not seem to know as much about privacy protection measures as they state. They also report that surveys as a method should not be taken as indicative of users' actual behavior. Acquisti studies the reasons that affect people's behavior vis-a-vis privacy and reports bounded rationality as well as the practice of hyperbolic discounting [1]; assigning a higher value to actions involving immediate gratification than those actions leading to long-term protection. In this work, we focus on understanding people's knowledge and perception of how their PII is exploited from an economic view-point, and use experience sampling to capture the behavior and context and as a result, do not suffer from the limitations seen in survey-based studies.

8. CONCLUSIONS

Our paper deals with the economic value that users assign to PII. Previous literature has focused on different types of PII, but not web-browsing behavior, which is the focus of this work. Previous work has also shown that privacy valuation is a difficult problem, as is affected by a number of technical, legal, social and psychological factors that lead to inconsistencies between what people say and what they actually do. We attempt to overcome these issues through the use of the refined Experience Sampling Method and a truth-telling auction mechanism that incentivizes users to participate honestly.

We found that users give more importance to PII related to their offline identities than to PII that is related to their online behavior. They mostly do not care about the quantity of PII released but they do care about its type. Users tolerate the use of their personal information for improving service, they do not like their information to be used to generate revenues. Users also preferred trading in their PII for money or improved services, and targeted advertising, in this order. We hope the results in this paper can guide future privacy research and solutions.

9. ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for constructive feedback. The authors would also like to express gratitude to Jonathan Smith, David Wetherall and Lorrie Cranor for providing excellent feedback on the previous versions of the paper. Finally, the authors would also like to thank Dr. Josep Blat for facilitating collaboration in the development to this study and Edgar Olivares for his assistance in data coding.

10. REFERENCES

[1] A. Acquisti. Privacy and Security of Personal Information: Technological Solutions and Economic Incentives. *The Economics of Information Security, J. Camp and R. Lewis (eds)*, pages 165–178, 2004.

[2] A. Acquisti and J. Grossklags. Privacy and Rationality in Individual Decision Making. *Security & Privacy, IEEE*, 3, 2005.

[3] A. Acquisti and L. John. What is privacy worth. Technical report, Heinz College, Carnegie Mellon University, 2009.

[4] J. Angwin and S. Stecklow. Scrapers dig deep for data on web. http://on.wsj.com/P17bnx, 10 2010.

[5] L. F. Barrett and D. J. Barrett. An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19(2):175–185, Summer 2001.

[6] A. Braunstein, L. Granka, and J. Staddon. Indirect Content Privacy Surveys: Measuring Privacy Without

Asking About It. In *Symposium on Usable Privacy and Security (SOUPS)*, 2011.

[7] M. Cherubini and N. Oliver. A refined experience sampling method to capture mobile user experience. In *Workshop of Mobile User Experience Research part of CHI'2009*, 2009.

[8] S. Consolvo and M. Walker. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing*, 2(2):24–31, Apr. 2003.

[9] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis. A study on the value of location privacy. In *Proceedings of the 5th ACM workshop on Privacy in electronic society - WPES '06*, page 109, New York, New York, USA, 2006. ACM Press.

[10] G. Danezis, S. Lewis, and R. Anderson. How much is location privacy worth? In *Online Proceedings of the Workshop on the Economics of Information Security Series (WEIS) 2005*, 2005.

[11] C. Dwyer. Behavioral targeting: A case study of consumer tracking on levis.com. *Risk Analysis*, 2009.

[12] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *ACM MobiSys '07*, 2007.

[13] S. Goel, A. Broder, E. Gabrilovich, and B. Pang. Anatomy of the long tail: Ordinary people with extraordinary tastes. In *in WSDM*, 2010.

[14] J. Grossklags and A. Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *Workshop on Economics of Information Security*. Citeseer, 2007.

[15] S. Guha, B. Cheng, and P. Francis. Privad: Practical Privacy in Online Advertising. In *NSDI*, 2011.

[16] S. Guha and S. Kandula. Act for affordable data care. In *HotNets*, pages 103–108, 2012.

[17] S. E. Hormuth. The sampling of experiences in situ. *Journal of Personality*, 54(1):262–293, 1986.

[18] B. Huberman, E. Adar, and L. Fine. Valuating privacy. *Security & Privacy, IEEE*, 3(5):22–25, 2005.

[19] G. Iachello, K. N. Truong, G. D. Abowd, G. R. Hayes, and M. Stevens. Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. *ACM SIGCHI CHI '06*, 2006.

[20] S. S. Intille, J. Rondoni, C. Kukla, I. Ancona, and L. Bao. A context-aware experience sampling tool. In *CHI '03 extended abstracts on Human factors in computing systems*, CHI EA '03, pages 972–973, New York, NY, USA, 2003. ACM.

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

[22] C. Jensen, C. Potts, and C. Jensen. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, 2005.

[23] N. Jentsch, S. Preibusch, and A. Harasser. Study on monetising privacy: An economic model for pricing personal information. *ENISA*, Feb. 2012.

[24] J. Kenneth and L. Hurwicz, editors. *Studies in resource allocation processes*, chapter On informationally decentralized systems. Cambridge University Press, 1978.

[25] B. Krishnamurthy, D. Malandrino, and C. E. Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *SOUPS*, 2007.

[26] B. Krishnamurthy, K. Naryshkin, and W. Craig. Privacy leakage vs. protection measures: the growing disconnect. *WISP*, May,2011.

[27] B. Krishnamurthy and C. Wills. Privacy diffusion on the web: a longitudinal perspective. In *WWW 2009*, 2009.

[28] K. C. Laudon. Markets and privacy. In *ICIS*, 1993.

[29] C. Mancini, K. Thomas, Y. Rogers, B. Price, L. Jedrzejczyk, A. Bandara, A. Joinson, and B. Nuseibeh. From spaces to places: emerging contexts in mobile privacy. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 1–10. ACM, 2009.

[30] A. M. McDonald and L. F. Cranor. Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*, 2010.

[31] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting price and search discrimination on the internet. In *HotNets*, 2012.

[32] A. Narayanan, V. Toubiana, S. Barocas, H. Nissenbaum, and D. Boneh. A critical look at decentralized personal data architectures. *CoRR*, abs/1202.4503, 2012.

[33] C. Paine, U.-D. Reips, S. Stieger, A. Joinson, and T. Buchanan. Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *Int. J. Hum.-Comput. Stud.*, 65:526–536, June 2007.

[34] S. Richmond. Eric schmidt: Google gets close to 'the creepy line'. http://tgr.ph/dp8BaH, 10 2010.

[35] C. Riederer, V. Erramilli, A. Chaintreau, P. Rodriguez, and B. Krishnamurty. For sale: Your data, by: you. In *ACM Hotnets*, 2011.

[36] F. Roesner, T. Kohno, and D. Wetherall. Detecting and defending against third-party tracking on the web. In *NSDI'12: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 12–12, 2012.

[37] E. Steel. A web pioneer profiles users by name. http://on.wsj.com/riy5Qo, 10 2010.

[38] E. Steel and G. A. Fowler. Facebook in privacy breach. http://on.wsj.com/buDn6d, 10 2010.

[39] R. Thaler. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1):39–60, 1980.

[40] V. Toubiana, H. Nissenbaum, A. Narayanan, S. Barocas, and D. Boneh. Adnostic: Privacy preserving targeted advertising. In *NDSS*, 2009.

[41] J. Valentino-Devries and E. Steel. White house to push privacy bill. http://on.wsj.com/hlSR7E, 03 2011.

[42] WSJ. Wsj: What they know series. http://online.wsj.com/public/page/what-they-know-digital-privacy.html, 04 2012.