Money Walks: A Human-Centric Study on the Economics of Personal Mobile Data

Jacopo Staiano[†], **Nuria Oliver**[‡], **Bruno Lepri**[§],
Rodrigo de Oliveira[‡], **Michele Caraviello**[¬], **Nicu Sebe**[†]
[†]University of Trento, [‡]Telefonica Research, [§]Fondazione Bruno Kessler, [¬]Telecom Italia

ABSTRACT
In the context of a myriad of mobile apps which collect personally identifiable information (PII) and a prospective market place of personal data, we investigate a user-centric monetary valuation of mobile PII. During a 6-week long user study in a living lab deployment with 60 participants, we collected their daily valuations of 4 categories of mobile PII (communication, *e.g.* phonecalls made/received, applications, *e.g.* time spent on different apps, location and media, *e.g.* photos taken) at three levels of complexity (individual data points, aggregated statistics and processed, *i.e.* meaningful interpretations of the data). In order to obtain honest valuations, we employ a reverse second price auction mechanism. Our findings show that the most sensitive and valued category of personal information is location. We report statistically significant associations between actual mobile usage, personal dispositions, and bidding behavior. Finally, we outline key implications for the design of mobile services and future markets of personal data.

Author Keywords
Privacy; Economics; Personal Mobile Data; Auction; Mobile Computing; Living Lab; Monetization.

ACM Classification Keywords
K.6.0. MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS: Economics

INTRODUCTION
The number of mobile phones actively in use worldwide today is about 5 billion, with millions of new subscribers every day[1]. Mobile phones allow for unobtrusive and cost-effective access to previously inaccessible sources of behavioral data such as location, communications (calls and text messages), photos, videos, apps and Internet access [30]. Hence, a result of the ever-increasing adoption of these devices is the availability of large amounts of *personal data* related to habits, routines, social interactions and interests [30, 34].

However, the ubiquitous collection of personal data raises unprecedented privacy challenges. Users typically have to make decisions concerning the disclosure of their personal information on the basis of a difficult tradeoff between data protection and the advantages stemming from data sharing. Perhaps more importantly, people are typically not involved in the life-cycle of their own personal data – as it is collected by websites and mobile phone apps, which results in a lack of understanding of who uses their data and for what.

Several researchers have proposed and investigated new user-centric models for personal data management, which enable individuals to have more control of their own data's life-cycle [39]. To this end, researchers and companies are developing repositories which implement medium-grained access control to different kinds of personally identifiable information (PII), such as *e.g.* passwords, social security numbers and health info [52], and more recently location [16, 21, 38] and personal data collected online by means of smartphones or wearable devices [16, 50].

Previous work has introduced the concept of *personal data markets* in which individuals sell their own personal data to entities interested in buying it [3]. Buyers are likely to be companies and researchers, while sellers are individuals who receive compensation for sharing their own data. Riederer *et al.* [42] have recently proposed a mechanism called *transactional* privacy, devised to maximize both the user's control of their own PII and the utility of a data-driven market.

In the context of prospective personal data markets that offer increased transparency and control, it is of great importance to understand the value that users put to their own PII. Recently, Carrascal *et al.* [8] used a refined Experience Sampling Method (rESM) [9] and a reverse second price auction to assess the monetary value that people assign to their PII shared online via websites – *e.g.* keywords used in a search engine, photos shared in a social network, etc. However, the authors focus only on web-browsing behaviors without taking into account behaviors and personal information that can be captured by mobile phones.

Taking Carrascal *et al.* [8] as an inspiration, in this paper we investigate the monetary value that people assign to different kinds of PII as collected by their mobile phone, including location and communication information. We carried out a comprehensive 6-week long study in a living lab environment with 60 participants and adopted a Day Reconstruction Method [27] along with a reverse second price auction mechanism in order to poll and collect honest monetary valuations.

[1]`http://www.ericsson.com/ericsson-mobility-report`

Ubicomp '14, September 13 - 17 2014, Seattle, WA, USA

http://dx.doi.org/10.1145/2632048.2632074

The work of JS and RdO has been performed while at Telefonica Research. RdO is currently affiliated with Google Inc., USA.

The main contributions of this paper are:

1. Quantitative valuations of mobile PII as collected by a 6-week long study conducted in the wild;
2. Qualitative feedback on the valuations provided by each participant as gathered by an End of Study (EoS) survey;
3. A segmentation of PII valuations and findings based on 4 categories of mobile PII (communications, location, media and apps), 3 levels of complexity (individual, processed, aggregated), and one level of temporal granularity (daily);
4. A set of key insights about people's sensitivities and valuations of mobile PII and implications for the design of mobile services that leverage mobile PII.

RELATED WORK

In recent years, researchers have analyzed the factors that can influence a person's disclosure behavior and economic valuation of personal information. Demographic characteristics, such as gender and age, have been found to affect disclosure attitudes and behavior. Several studies have identified gender differences concerning privacy concerns and consequent information disclosure behaviors: for example, women are generally more protective of their online privacy [18, 22]. Age also plays a role in information disclosure behaviors: in a study on Facebook usage, Christofides *et al.* [10] found that adolescents disclose more information.

Prior work has also emphasized the role of an individual's stable psychological attributes - *e.g.* personality traits - to explain information disclosure behavior. Korzaan *et al.* [29] explored the role of the Big5 personality traits [12] and found that Agreeableness – defined as being sympathetic, straightforward and selfless, has a significant influence on individual concerns for information privacy. Junglas *et al.* [26] and Amichai-Hamburger and Vinitzky [5] also used the Big5 personality traits and found that Agreeableness, Conscientiousness, and Openness affect a person's concerns for privacy. However, other studies targeting the influence of personality traits did not find significant correlations [45]. More recently, Quercia *et al.* [41] found weak correlations among Openness to Experience and, to a lesser extent, Extraversion and the disclosure attitudes on Facebook. In 2010, Lo [33] suggested that Locus of Control [44] could affect an individual's perception of risk when disclosing personal information: internals are more likely than externals to feel that they can control the risk of becoming privacy victims, hence they are more willing to disclose their personal information [55].

Individual differences are also found when providing economic valuations of personal data [2, 8]. For instance, some individuals may not be concerned about privacy and would allow access to their data in exchange for a few cents, whereas others may only consent if well paid. Recently, Aperjis and Huberman [6] proposed to introduce a realistic market for personal data that pays individuals for their data while taking into account their own privacy and risk attitudes.

Previous research has shown that disclosure [28] and valuation [15, 23] depend on the kind of information to be released. Huberman *et al.* [23] reported that the valuation of some types of personal information, such as the subject's weight and the subject's age depends on the desirability of these types of information in a social context. Some empirical studies have attempted to quantify subjective privacy valuations of personal information in different contexts, such as personal information revealed online [20], access to location data [14], or removal from marketers' call lists [51]. These studies can be classified into two groups. The first and larger group includes studies that explicitly or implicitly measure the amount of money or benefit that a person considers to be enough to share her/his personal data, namely their *willingness to accept* (WTA) giving away his/her own data (see for example [14, 24]). The second and smaller group includes studies about tangible prices or intangible costs consumers are *willing to pay* (WTP) to protect their privacy (see for example, [1, 49]). In our paper, we do not deal with WTA vs WTP, but we focus on WTA for PII captured by mobile phones (communications, apps and media usage, locations).

A growing body of studies in the fields of ubiquitous and pervasive computing and human-computer interaction focuses on location sharing behavior and has highlighted the role played by the recipient of sharing (who can access the information), the purpose, the context, how the information is going to be used [7, 11, 32, 47, 54] and the level of granularity of the information shared [31]. Finally, studies have suggested the importance of analyzing people's actual behavior rather than attitudes expressed through questionnaires because often the actual behavior of people deviates from what they state [25].

Building upon previous work, in this paper we investigate the monetary value that people assign to different kinds of PII as collected by their mobile phone, including location and communication patterns. In particular, we carry out a comprehensive 6-week long study in a living lab environment with 60 participants and adopt a Day Reconstruction Method [27] and a reverse second price auction mechanism in order to poll and collect honest monetary valuations from our sample.

METHODOLOGY

Next, we describe the methodology followed during our 6-week study.

The Living Laboratory

The Living Laboratory where we carried out our study was launched in November of 2012 and it is a joint effort between industrial and academic research institutions. It consists of a group of more than 100 volunteers who carry an instrumented smartphone in exchange for a monthly credit bonus of voice, SMS and data access. The sensing system installed on the smartphones is based on the FunF[2] framework [4] and logs communication events, location, apps usage and photos shot. In addition, the members of the living lab participate in user-studies carried out by researchers. The goals of this living lab are to foster research on real-life behavioral analysis obtained by means of mobile devices, and to deploy and test prototype applications in a real-life scenario. One of the most important features of such a lab is its ecological validity, given that

[2] `http://funf.org`

the participants' behaviors and attitudes are sensed in the real world, as people live their everyday life, and not under artificial laboratory conditions.

All volunteers were recruited within the target group of young families with children, using a snowball sampling approach where existing study subjects recruit future subjects from among their acquaintances [19]. Upon agreeing to the terms of participation, the volunteers granted researchers legal access to their behavioral data as it is collected by their smartphones. Volunteers retain full rights over their personal data such that they can order deletion of personal information from the secure storage servers. Moreover, participants have the choice to participate or not in a given study. Upon joining the living lab, each participant fills out an initial questionnaire which collects their demographics, individual traits and dispositions (*e.g.* Big Five personality traits, trust disposition, Locus of Control, etc.) information.

Participants

A total of 60 volunteers from the living lab chose to participate in our mobile personal data monetization study. Participants' age ranged from 28 to 44 years old ($\mu = 38$, $\sigma = 3.4$). They held a variety of occupations and education levels, ranging from high school diplomas to PhD degrees. All were savvy Android users who had used the smartphones provided by the living lab since November 2012. Regarding their socio-economic status, the average personal net income amounted to €21169 per year ($\sigma = 5955$); while the average family net income amounted to € 36915 per year ($\sigma = 10961$). All participants lived in Italy and the vast majority were of Italian nationality.

Procedure

Our study ran for six weeks from October 28th, 2013 to December 11th, 2013. At the beginning of the study, participants were explained that the study consisted of three phases:

1. An initial questionnaire, which focused on their general perception of privacy and personal data;
2. A daily data collection phase that lasted 6 weeks where participants answered daily surveys to valuate their mobile personal data;
3. A final survey that aimed to clarify the results obtained and to collect qualitative feedback from participants.

Daily Surveys

Ad-hoc `java` code was developed and scheduled to run on a secure server each night in order to automatically generate personalized daily surveys for each participant. The survey questions were generated based on the mobile data collected during the previous day. Everyday, at 12PM, participants received an SMS reminding them to fill out their survey via a personalized URL (through a unique hash).

In order to test the live system and identify bugs, we ran a pilot for 10 days with a small set of volunteers who were not participants in the study. In addition, we allocated a *training* week prior to starting the actual study so participants would get accustomed to the survey/auction scheme.

COLLECTED DATA

Next we describe the data that we collected during the study.

Mobile Personal Data

We collected 4 categories of mobile personal data: (1) *communications*, in the form of calls made/received; (2) *locations*, collected by the device GPS sensor every $\sim$ 5 minutes; (3) *running applications*, sampled every 25 minutes; and (4) *media*, *i.e.* number and timestamp of pictures taken and obtained by monitoring the device file system. The sampling rates for the different categories of data were empirically determined in order to have good resolution without significantly impacting the device's battery life.

Moving from finer to coarser granularity, we probed participants about the following three levels of complexity for each category of data: (1) *individual*, encompassing individual data points (*e.g.* a call made/received, a picture taken, a specific GPS location); (2) *processed*, depicting higher level information derived from the sensed events (*e.g.* a given application has been running for N minutes, total distance traveled); and (3) *aggregated*, portraying cumulative event information (*e.g.* number of places visited, number of calls made/received).

For each data category and level of complexity, participants were asked to fill out daily surveys that asked them about data from the previous day for each category and for a specific level of complexity (up to 4 questions per day). For each question in the surveys, participants always had the option to opt-out and not sell that particular piece of information.

Next, we describe in detail the 4 categories and the 3 levels of complexity of mobile personal data that we collected in this study, which are summarized in Table 1.

Communications

Individual communication data was restricted to voice calls made/received; missed calls were discarded. The *processed* communication variable referred to the total duration of calls in the previous day, resulting in questions such as `"Yesterday, you spoke on the phone for a total of 52 minutes"`.

With respect to *aggregated* communications data, we alternated between two different aggregated variables on a weekly basis: on even weeks subjects were asked to monetize information about the total number of calls made/received during the previous day, while on odd weeks they were asked about call diversity, *i.e.* the number of different people that they talked to on the phone during the previous day. Examples of questions related to aggregate communications are `"Yesterday, you made/received 8 phone calls"`, or `"Yesterday, you spoke on the phone with 3 different persons"`.

Location

Individual location referred to a specific place visited by the participant in the previous day. Semantic information associated to GPS locations was derived via reverse geocoding using Yahoo Query Language. For individual locations, details on street, neighborhood and town were included

Category	Individual	Processed	Aggregated
Communications	A call event [*]	Total duration of calls	# of calls or diversity
Location	A place visited [*]	Total distance covered	# of places visited
Running Apps	App *X* running [*]	App *X* running for *N* minutes in the [**]	# of apps running
Media	A picture shot [*]	Pictures shot in the [**]	# of pictures shot

Table 1: Categories of personal data probed in the surveys. Include [*: at time hh:mm; **: night (12AM-6AM), morning (6AM-12PM), afternoon (12PM-18PM), evening (18PM-12AM)]. All questions referred to data collected the previous day.

in the question. For example, `"Yesterday, at 23:56 you were in Via Degli Orbi 4, Trento"`. The *processed* location variable referred to the total distance traveled in the previous day, resulting in questions such as `"Yesterday you covered a total distance of 13km"`.

Finally, location data was spatially clustered over the reference time-range using a threshold of 100 meters to generate the *aggregated* location question (*e.g.* `"Yesterday you have been in 23 different places"`).

Running Applications

With respect to running apps, the *individual* variable included the timestamp and the name of the app running in the foreground. *Processed* app information referred to the total number of minutes that a particular app was running over a specific time in the previous day, whereas *aggregated* app variables referred to the total number of different apps that the participant ran the previous day.

Examples of questions on app-related information for each level of complexity are `"Yesterday, at 10:23 you were using the Firefox Browser application"`, `"Yesterday night, the Google Talk application run on your device for 82 minutes"`, and `"Yesterday 9 applications were running on your device"`, respectively.

Media

Individual media asked participants about the fact that they shot a photo at a specific time (`"Yesterday, at 14:23, you shot one picture"`). For legal privacy reasons, the questions referring to individual media data could not include the actual picture they referred to. *Processed* media probed participants about their photo-taking activity during specific times of the day (*e.g.* `"Yesterday morning you took 4 pictures"`). Finally, the *aggregated* media variable referred the total number of pictures shot the previous day (*e.g.* `"Yesterday you took 9 pictures"`).

Individual Traits Data

As previously mentioned, upon joining the lab each participant filled out 4 questionnaires to collect information about their personality, locus of control, dispositional trust and self-disclosure behaviors.

The Big Five personality traits were measured by means of the BFMS [40] questionnaire, which is validated for the Italian language and covers the traditional dimensions of Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness [12]. Participants also provided information about their *Locus of Control* (LoC) [43], a psychological construct measuring whether causal attribution for subject behavior or beliefs is made to oneself or to external events and circumstances. The LoC measures whether the outcomes of a set of beliefs are dependent upon what the subject does (internal orientation) or upon events outside of her/his control (external orientation). LoC was measured by the Italian version of Craig's Locus of Control scale [17].

Moreover, we collected information about the participants' *dispositional trust*. Rotter [44] was among the first to discuss trust as a form of personality trait, defining interpersonal trust as a generalized expectancy that the words or promises of others can be relied on. In our study, we resort to Mayer and Davis's Trust Propensity Scale [35].

Finally, we targeted the *self-disclosure* attitudes of our subjects. Self-disclosure has been defined as any message about the self that an individual communicates to another one [13, 53]. We used Wheeless' scale [53] measuring five dimensions of emphself disclosure, namely (i) amount of disclosure, (ii) positive-negative nature of disclosure, (iii) consciously intended disclosure, (iv) honesty and accuracy of disclosure, and (v) general depth or intimacy of disclosure. Wheeless' scale has been utilized to measure self-disclosure in online communication and in interpersonal relationships [53].

Auctions of mobile PII

The personalized daily survey asked each participant to place a bid to sell one piece of their mobile personal information for each of the four categories of study (communications, location, apps and media), for a specific level of complexity (individual, processed, or aggregated) and for the previous day. The winner of each auction won the monetary value associated with that auction. In exchange, (s)he sold that particular piece of information to the Living Lab which could use it for whatever purpose it wanted.

In order to ensure a balanced sample, surveys were generated by rotating the different levels of complexity described above, such that each day participants placed bids in up to 4 auctions: one for each category of personal information and for a particular level of complexity (individual, aggregated or processed). Note that in the case a participant did not generate any data for a particular category, s(he) was still asked to provide a valuation to the fact that there was no data in that category, *e.g.* `"Yesterday you did not make any phone call"`.

The participants' bids entered a reverse second-price auction strategy, *i.e.*, the winner was the participant(s) who bid the lowest, and the prize was the second lowest bid. The choice

Question	mean	st_dev
Q1. I am concerned about the protection of the data collected by my smartphone	4.7	1.6
Q2. I trust the applications I install and run on my smartphone wrt how they use my data	3.7	1.5
Q3. I trust telco providers with respect to how they use my data	3.4	1.4
Q4. I always read the privacy terms and conditions for the applications I use	2.7	1.6
Q5. I know the legislation on mobile communication data protection	2.5	1.5

Table 2: Questions asked in the Initial questionnaire, and responses statistics. The 7-point likert scale used goes from *1-Totally Disagree* to *7-Totally Agree*.

of this auction mechanism was due to the following reasons: (1) the mechanism is truth telling given that the best strategy for the auction participants is to be honest about their valuation [36], (2) it is easy to explain and understand, and (3) it has successfully been used before to evaluate location information in [15] and Web-browsing information in [8]. Interventions, *i.e.* individual communications of auction outcomes to participants, took the form of e-mails sent every Thursday.

In order to evaluate possible effects of winning frequency on bidding behavior, we employed two different auction strategies for the first and second halves of the study. During the first 3 weeks (phase 1), we carried out weekly auctions on Wednesday, taking into account all bids that had been entered during the previous 7 days for each category. Therefore, in this phase, 12 weekly auctions took place with the daily bids for each category and level of complexity (4 categories x 3 levels of complexity). During the last 3 weeks of the study (phase 2), we switched to daily auctions; furthermore, the sample of bidding participants was split into 3 random subsets in order to increase their chances of winning, resulting in a total of 12 auctions per day.

Email interventions were always on Thursdays and therefore this change was transparent to participants. Interventions were sent to all participants, whether they had won auctions or not. In the case of winners, the intervention email included the specific piece of information that the participant had sold, the corresponding winning bid, and the amount won. In the case of losers, the intervention email simply communicated the participant that s(he) did not win any of their auctions. All emails were kept neutral for both winners and losers.

In total, 596 auctions were run during the entire study (36 in the first three weeks, 560 afterwards).

Pre- and Post- Study Questionnaires

As previously explained, at the beginning and at the end of the data collection participants were required to fill out initial and end-of-study (EoS) questionnaires. The initial questionnaire consisted of 5 questions (see Table 2) and was used to gather information about the participants' perception of privacy issues related to mobile personal data. From the responses provided to this survey, we notice that participants are concerned about mobile PII protection (Q1) but do not tend to read the Terms of Service (Q4) nor are aware of current legislation on data protection (Q5). Moreover, they do not seem to trust how neither application providers (Q2) nor telecom operators (Q3) use their data.

The EoS survey was designed to gather additional quantitative and qualitative information from our participants after the data collection was complete. In particular, we asked participants to put a value (under the same auction game constraints) on category-specific *bulk information – i.e.* all the data gathered in the study for each category. For instance, in the case of location information, a visualization of a participant's mobility data collected over the 6-weeks period was shown in the Web questionnaire (as depicted in Figure 1) and the participant was asked to assign it a monetary value. Furthermore, for each category, we asked participants about the minimum/maximum valuations given during the study, in order to understand the reasons why they gave these valuations. Table 3 contains all the questions of the EoS survey.

The EoS questionnaire was administered through a slightly modified version of the same Web application used for the daily surveys. The main difference are the visualizations of the collected data.

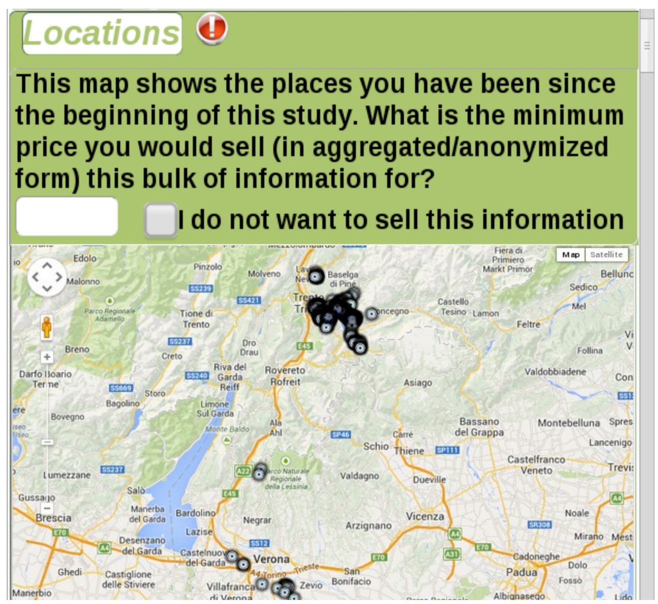

Figure 1: Location-specific bulk information question in the EoS survey.

DATA STATISTICS

The data used throughout this paper was collected from October 28th and December 11th 2013, inclusive. Data was not collected for the first 3 days of November, due to the All Saints festivities in Italy; hence, our data-set encompasses 43 days. A total of 2838 daily surveys were administered during this period. Statistics on bidding data and participation follow.

Question	Type
Q1. This {map‖chart} shows the information about {locations‖communications‖apps‖media} we collected during this study. What is the minimum amount of money you would accept to sell it in anonymized/aggregated form?	numeric
Q2. On day {dd/MM} you assigned a value of {*min-bid per category*} to the information [{*least valued info per category*}]. This was your minimum bid. Why?	multi-choice*
Q3. On day {dd/MM} you assigned a value of {*max-bid per category*} to the information [{*most valued info per category*}]. This was your maximum bid. Why?	multi-choice*
Q4. Imagine there was a market in which you could sell your personal information (e.g. information about people you called, places you've been, applications you've used, songs you've listened to, etc.). Who would you trust to handle your information? Please, order the following entities from most to least trusted.	rank**
Q5. The category {locations‖communications‖apps‖media} is the one that you refused to sell the most ({*percentage of opt-outs*}). Why?	free-text

Table 3: Questions asked in the EoS questionnaire. *included: *Fair value*, *Test/Mistake*, *Other (free text)*. For minimum-bid related questions additional options were *To win the auction*, *Info not important*; conversely, for maximum-bid related questions, the additional option was *To prevent selling*. **entities to be ranked included: *banks*, *government*, *insurance companies*, *telcos*, *yourself*.

Bids

Table 4 summarizes the bidding values for each personal data category and level of complexity. Figure 2 depicts median bid values each day for each category and level of complexity[3].

	Individual	Processed	Aggregated	Global
Location	[1, 3, 9]	[1, 2, 7]	[1, 3, 10]	[1, 3, 8]
Communications	[.95, 2, 5.96]	[.9, 2, 8]	[1, 2, 8]	[1, 2, 7]
Apps	[1, 2, 6]	[1, 2, 5]	[1, 1, 5]	[1, 2, 5]
Media	[.5, 1, 5]	[.5, 1, 3]	[.5, 1, 5]	[.5, 1, 4]

Table 4: [Q1,median,Q3] triplets for bid values (€) per category and level of complexity.

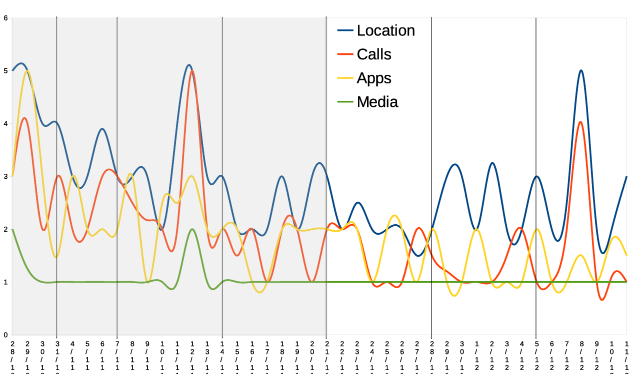

Figure 2: Daily median bid values (€) per category. Vertical lines indicate interventions. Shaded area indicates phase 1.

Participation

The participation rates for daily surveys is 79%. As mentioned earlier, users were granted opt-out options for each survey question by ticking a check-box which portrayed "`I do not want to sell this information`". Table 5 reports statistics of opt-out and distributions of valid responses (*i.e.* survey items for which participants did not opt-out and entered their bid) for each category.

[3]Note how the spatial gap between the first two interventions is smaller than between the rest of interventions because of the lack of data during 3 days in November.

Awards

The total amount won by participants in the form of auction awards was €262 which was paid in Amazon vouchers. Additionally, we selected the ten subjects with the highest response rate and ran a raffle to select the winner of a final prize of €100.

A total of 29 subjects won at least one auction during the study; the cardinality of the winning set ramped from 5 to 29 as an effect of the increased number of auctions run in the second phase of the study.

DATA ANALYSIS

The bidding data that was collected in the study is not normally distributed. Hence, we applied non-parametric analysis to test whether significant differences exist in the value distributions of different types of personal data. Thus, we report results using the Kruskal-Wallis test with a level of significance of $p < .05$.

Furthermore, we carried out correlation analyses to investigate whether associations between mobile phone usage patterns, demographics, subjects' predispositions, traits and auction behavior exist. For these analyses we employed the non-parametric Spearman's Rho method with a level of significance of $p < .05$.

Bids

We investigate first daily bids and specifically whether significant differences exist between (1) the categories and (2) levels of complexity within each category of mobile personal data we collected.

Between-Category Study

Significant differences in bid distributions were found between all data categories, with the only exception of Communications and Apps.

The lack of statistically significant differences between Communications and Apps could be partially explained by the fact that most of the apps installed and used by participants in the study are communication apps. In terms of both running time and installations, ∼50% of the top 20 apps are

Category	1st quartile	median	3rd quartile	mean	st_dev	Opt-out (avg.%)	Opt-out (median%)
Location	34.3	69.8	84.3	58.2	33.1	17.7	2.63
Communications	55.8	74.4	88.4	64.8	29.9	5.01	0
Running Apps	40.7	65.1	81.4	58	30.8	7.59	0
Media	62.8	76.7	90.7	66.4	32.8	9.25	0

Table 5: Distribution statistics of valid bid responses per category. Values reported in percentages. Last columns portray the opt-out statistics per category.

messaging apps (WhatsApp and similar), email (Gmail, Hotmail, Y!Mail), voice-over-IP clients (Skype, Viber) and social networking clients (Facebook). We thus hypothesize that the distinction between Communication and Apps might be blurred. We leave the validation of such a hypothesis to future work. Nonetheless, the finding that participants seem to perceive, and consequently valuate, communications provided by a telco company and those provided by mobile apps in a similar manner, is intriguing and worth investigating.

Within-Category Study

Next we analyzed the differences in the distribution of bids within the different levels of complexity of mobile personal data. In other words, we looked if bid distributions within a given mobile data category showed significant differences for individual, aggregated, and processed information.

Applications. Significant differences emerged between individual and aggregated information ($p = .0108$), and between aggregated and processed information ($p = .039$). In particular, aggregated information about running applications (e.g. *yesterday 7 applications were running on your device*) was valued less ($\tilde{x} = €1$) than individual (e.g. *yesterday the Gmail application was running on your device*) or processed (e.g. *yesterday the Gmail application ran for 120 minutes on your device*) information ($\tilde{x} = €2$). No significant difference was found between monetary valuations of individual and processed information on running applications ($p = 0.659$).

Media. Within the Media category, a significant difference in bid distributions was found ($p = .046$) between aggregated (e.g. *yesterday you shot 8 pictures*) and processed (e.g. *yesterday night you shot 3 pictures*) information. While for both information types the median bid value is $\tilde{x} = €1$, a significant difference exists in terms of dispersion: the quartile coefficient of dispersion (*i.e.* the ratio between difference and sum of the 3rd and 1st quartiles) is, respectively, $qcod_{agg} = .81$ and $qcod_{pro} = .71$.

Communications and Locations. No significant differences were found in within-category analyses for Communications and Locations. In other words, participants valued similarly the communication and location data with each of the 3 levels of complexity.

Impact of the Change in Auction Strategy

As described earlier, in the middle of the study we increased the frequency of auctions from weekly (phase 1) to daily (phase 2). This change was transparent to participants and the frequency of email interventions was kept constant – every Thursday. We designed these two phases to assess if the probability of winning had an effect on bidding behavior. Indeed, we observe significant differences in bid distributions between the two phases for all categories: locations ($p = .02$), communications ($p = .01$), apps ($p = .001$) and media ($p = .005$). Moreover, we find that mobile PII valuations drop for all categories in the second phase, as more participants won the auctions to monetize their data.

The Value of Bulk Information

The monetary valuations gathered in the final questionnaire for bulk information (*i.e.* all the data collected in the 6-weeks presented in aggregated/anonymized form) are summarized in Table 6. Since participants could opt-out, we also report opt-out percentages for bulk information.

Comparing with daily bids (see Table 4), the median bids for bulk information are one order of magnitude larger than the median individual bids, except for the media category. Mean opt-out percentages are similar except for the apps category. The value ranking obtained from daily bids (Location > Communications > Apps > Media) is different from that obtained in bulk bids (Location > Apps > Communications > Media). In particular, application-related bulk data is valued significantly higher than communications-related bulk information.

	Location	Communications	Apps	Media
mean	588.1	51.1	170.4	25.1
median	22.5	15	20	5
opt-out (%)	16.67	3.34	0	8.34

Table 6: Median/mean values (€) for bulk bids, and corresponding opt-out percentages.

Relationship between Bids and Daily Behaviors

In order to assess whether significant effects exist between mobile phone usage patterns and bidding behavior, we first computed daily behavioral variables from the sensed data. Table 7 depicts the variables that we extracted with a daily granularity and for each participant. With respect to *location* data, information about the number of places visited was derived under the assumption that two locations would correspond to different places if the distance between them was larger than a threshold set to 100 meters. The radius of gyration corresponds to the radius of the smallest circle encompassing all location points registered each day.

For all these behavioral variables, we computed higher-order features corresponding to their statistical behavior over the 6-weeks period: mean, median, standard deviation, coefficient of variation (ratio of the standard deviation to the mean) and the quartile coefficient of dispersion. The last two features capture dispersion effects.

Furthermore, for each participant and data category, we computed mean, median, and standard deviation of their bids.

Category	Daily Behavioral Variables
Location	Distance *total/mean/median/std* Speed *mean/median/std* Radius of Gyration Number of Places Visited
Communications	Calls Duration *total/mean/median/std* Calls Diversity Calls Total
Applications	Total Apps Running Total Apps Running Time
Media	Total Pics shot

Table 7: Daily behavioral variables computed from mobile phone usage data.

Daily Bids

We studied all correlations found between daily behavioral variables and bids in each category.

We found a positive correlation between the mean location bid value and the median of daily distance traveled ($R = .294, p = .024$). That is, the larger the daily distance traveled, the higher the valuations of location information. With respect to applications, there are several statistically significant correlations. In particular, the total app running time is negatively correlated with the median app bid value ($R = -.26, p = .048$), meaning that the more time a participant spent using mobile apps, the lower the median valuations of app information. No significant correlation was found between communication and photo-taking behavioral features and bids on the communications and media categories.

Bulk Bids

There were a number of significant correlations between bids on bulk information and daily behaviors. Below we summarize the most notable correlations that we found.

Mobility information was positively correlated with bids on bulk *location*, *communication* and *application* information. In particular, with the median of the i) radius of gyration ($R = .46, p = .0008$ for loc.; $R = .37, p = .005$ for comm.; $R = .34, p = .009$ for apps); and ii) daily mean speed ($R = .29, p = .04$ for loc.; $R = .39, p = .002$ for comm.; $R = .29, p = .029$ for apps). Location and application data was also positively correlated with the median of the daily mean distance traveled ($R = .39, p = .005$ for loc.; $R = .28, p = .031$ for apps) whereas communication bids were also positively correlated with the median of the i) total distance traveled ($R = .314, p = .018$) and ii) number of places visited ($R = .336, p = .011$).

We also found statistically significant negative correlations of bulk location, communication and application bids with the coefficients of variation of mobility variables.

These correlations imply that the larger the daily distance traveled, the higher the valuation of location, communication and application bulk bids. Conversely, the higher the variation in the patterns of mobility of a person, the lower his/her valuation of location, communication and app bulk information. Note that bulk communication bids were not correlated with communication variables.

In addition, bulk application bids are negatively correlated with the cumulative sum of daily unique total apps ($R = -.37, p = .003$) and with the median ($R = -.28, p = .029$) and mean ($R = -.26, p = .04$) of total apps running daily.

Finally, bulk media bids are correlated with the cumulative sum of daily unique total apps ($R = -.29, p = .03$).

Relationship between Bids, Demographics, Traits and Dispositions

Daily Bids

In the case of daily bids, we did not find any meaningful statistically significant correlation between bids and our participants' demographics or personality.

There were statistically significant correlations with *self-disclosure* variables that could be explained by the relevance of privacy aspects for all types of self-disclosure [37]. In particular, the Intentional/Unintentional factor in self-disclosure is positively correlated with bids in three categories (communication, applications and media): (1) mean ($R = .258, p = .048$), median ($R = .291, p = .02$) and standard deviation ($R = .323, p = .012$) in communication bids, (2) median application bid value ($R = .26, p = .04$), and (3) median ($R = .30, p = .02$), mean ($R = .27, p = .041$), and standard deviation ($R = .305, p = .019$) of media bids.

Bulk Bids

Bulk location bids are found to be negatively correlated with Creativity ($R = -.375, p = .007$), while having positive correlations with the Intentional/Unintentional factor in self-disclosure ($R = .295, p = .039$) and Agreeableness ($R = .31, p = .027$). Interestingly, a positive correlation exists between bulk location bids and personal income ($R = .32, p = .02$). Furthermore, bulk communication information positively correlates with Agreeableness ($R = .31, p = .018$), and with the Intentional/Unintentional factor in self-disclosure ($R = .34, p = .009$).

INSIGHTS FROM THE EOS SURVEY

In the final survey, we asked our participants about particular bids they made during the 6-week data collection phase, and gave them the opportunity to express their views and concerns in free-form text (see Table 3 for details).

Trust

As seen in Table 3, Q4 asked our participants about their trust preferences with respect to 5 different entities who could be the safekeepers of their personal data: themselves, banks, telcos, governments and insurance companies. From the trust rankings provided by our participants, we computed a *trust score* for each entity by assigning a 1 to 5 value according to its rank and subsequently normalizing by the number of respondents. The final ranking that we obtained was: *yourself* (.997), *banks* (.537), *telcos* (.513), *government* (.49), and *insurance companies* (.46).

This result is aligned with the initial survey answers (Q2 and Q3 in Table 2) where participants conveyed that they do not trust telco operators or app providers with how they use their data.

In sum, overwhelmingly our participants trust themselves with their personal data more than any other entity, followed by banks and telcos. Insurance companies were the least trusted party. A similar question was also asked by Carrascal *et al.* [8] obtaining similar results: the most trusted entity for a subject was the subject himself and the least trusted entities were the insurance companies. Interestingly, in our study, conducted in Italy, government was the second *least* trusted entity while in Carrascal *et al.* [8], conducted in Spain, the government was the second *most* trusted entity.

Lowest/Highest Bids per Category

When analyzing the lowest/highest bids per category, we found that 70% of the highest bids for all categories took place in the first phase of the study (during the first three weeks). Adding more auctions (as it happened in the second phase of the study) led to lower bids.

In the communications category, 61% of the time participants entered a low bid to win and sell the associated communications information. This was significantly higher than for any other category. For all other categories, the most common reason reported for entering the low bid was that the information was not important. This finding suggests that participants found communication data to be the most desirable to sell.

Conversely, location was the most sensitive category of information as 25% of the time participants entered a high location bid in order to avoid selling the information. This was significantly higher than for the other categories (5% for communications, 3% for apps and 6% for media).

Insights about Opt-out Choices

Location was the category of data for which subjects opted-out the most (56%), followed by media (24%), apps (18%) and communications (2%). In the free-text explanations provided by our subjects it is clear that location is deemed to be the most sensitive category of information, *e.g.*:

> *"I don't like the idea of being geo-localized."*
> *"This kind of information is too detailed and too personal."*

Interesting explanations were also provided to justify the choice of not selling apps information, including that from apps usage is possible to infer information related to interests, opinions (expecially political opinions), and tastes:

> *"From the usage of some applications it is possible infer information such as political orientation and other opinions and interests."*

DISCUSSION AND IMPLICATIONS

From the previously described analyses we can draw six insights related to mobile personal data:

1. The Value of Bulk Mobile PII: Carrascal *et al.* [8] have reported higher values in their study on valuation of personal Web-browsing information than the ones we obtained in our study. The overall median bid value in our study was $\tilde{x} =$ €2 while Carrascal *et al.* reported an overall median bid value equal to $\tilde{x} =$ €7 when they took in account context-dependent personal information. There are a few methodological differences between both studies which might explain the differences in bid values. In particular, [8] asked participants to provide a valuation of personal information captured while browsing the Web *in-situ* using a rSEM methodology. Instead, we employed a DRM methodology querying participants about ther mobile PII from the previous day. From the valuations obtained in [8] and our study, it seems that individual pieces of PII are not as valuable when queried *out-of-context* –such as in our study– than *in context* –such as in [8].

Conversely, bulk mobile PII was valued higher in our study than in [8] and significantly higher than individual PII. As shown in Tables 4 and 6, bulk information was valued an order of magnitude higher than individual data except for information in the media category. This finding is probably due to the power of the visualizations in the EoS survey, particularly for location and apps data.

One hypothesis for this higher valuation is that participants realized how bulk data conveyed information about their lifestyle and habits and therefore considered it to be more valuable than daily items. Recently, Tang *et al.* have shown the impact of different visualization types (text-, map-, and time-based) on social sharing of location data [46].

This result has a direct consequence for the design of trading mobile PII and highlights an asymmetry between buyers and sellers: for buyers, it would be more profitable to implement mechanisms to trade single pieces of information –that they could later aggregate. For sellers, however, it would be more advantageous to sell bulks of information.

2. Location, location, location: As shown in Tables 4 and 6, location information received the highest valuation for all levels of complexity and was the most opted-out category of mobile PII. Bulk location information was very highly valued, probably due to the powerful effect of the map visualization in the EoS survey. Several participants also expressed that they did not want to be geolocalized and considered location information to be highly sensitive and personal.

Moreover, we found statistically significant correlations between mobility behaviors (*e.g.* mean daily distance traveled, daily radius of gyration, etc.) and valuations of personal data. Not all users value their personal data equally: the more someone travels on a daily basis, the more s/he values not only her/his location information but also her/his communication and application information.

Regarding this relation, previous works who focused on location information have presented contrasting results [14, 15]; as we probe participants daily about fine-grained personal data they have just produced, our approach substantially differs from these survey-based studies, and it is thus difficult to directly compare with these works. Generally, our results seem to support the findings presented in [15].

These insights may have an impact on the design of commercial location-sharing applications. While users of such applications might consent at install time to share their location with the app, our work suggests that when explicitly asked about either individual or bulk location data, $\sim 17\%$ of users decide not to share their location information. In addition, mobility behaviors will influence the valuations of PII.

Tsai *et al.* [48] conducted an online survey with more than 500 American subjects to evaluate their own perceptions of the likelihood of several location-sharing scenarios along with the magnitude of the benefit or harm of each scenario (*e.g.* being stalked or finding people in an emergency). The majority of the participants found the risks of using location-sharing technologies to be higher than the benefits. However, today a significant number of very popular mobile apps such as Foursquare and Facebook Places make use of location data. These popular commercial location sharing apps seem to mitigate users' privacy concerns by allowing them to selectively report their location using check-in funtionalities instead of tracking them automatically.

Based on our findings and given our participants concerns and high valuations of bulk location information, we believe that further user-centric studies on sharing and monetary valuation of location data are needed.

3. Socio-demographic characteristics do not matter, behavior does: When we correlated bid values against socio-demographic characteristics, we did not find significant correlations. This result is in contrast to previous work that found socio-demographic (mainly sex and age) differences in privacy concerns and consequent information disclosure behaviors [10, 18]. However, these previous studies were focused mainly on online information and on disclosure attitudes and privacy concerns than on monetary valuation of personal data. Carrascal *et al.* [8], instead, found results in line with ours (no significant correlations) except for a surprising low valuation of online information from older users.

On the other hand, we found statistically significant correlations between behavior (particularly mobility and app usage) and valuations of bids. From our findings it seems that personal differences in valuations of mobile PII are associated with behavioral differences rather than demographic differences. In particular, the larger the daily distance traveled and radius of gyration, the higher the valuation of PII. Conversely, the more apps a person used, the lower the valuation of PII. A potential reason for this correlation is due to the fact that savvy app users have accepted that mobile apps collect their mobile PII in order to provide their service and hence value their mobile PII less.

4. Intentional self-disclosure leads to higher bids: We found a positive correlation between the Intentional/Unintentional dimension of self-disclosure and the median values of the bids. This result could be explained by the fact that people with more intentional control about disclosing their own personal information, may be more aware of their personal data and hence also value it more from a monetary point of view.

Interestingly, we did not find significant correlations between bid values and other traits with the exception of Agreeablenness (with bulk location and communication bids). Previous studies on the influence played by individual traits (usually personality traits and LoC) on privacy dispositions and privacy-related behaviors have provided contrasting evidence: some of them found small correlations [33, 41], while Schrammel *et al.* found no correlations [45]. Hence, our results require additional investigations in order to clarify which are, if any, the dispositions and individual characteristics to take in account when a buyer makes a monetary offer for personal data.

5. Trust: From our study and from Carrascal *et al.* [8], it clearly emerges that individuals mainly trust themselves to handle their own personal data. This result suggests the adoption of a decentralized and *user-centric* architecture for personal data management.

Recently, several research groups have started to design and build personal data repositories which enable people to control, collect, delete, share, and sell personal data [16, 38], and whose value to users is supported by our findings.

6. Unusual days lead to higher bids: During our study there were two unusual days: December 8th (Immaculate Conception Holiday) and November 11th (a day with extremely strong winds which caused multiple road blocks and accidents). As can be seen in Figure 2, the median bids for all categories in these two days were significantly higher than for the rest of the days in the study. Perhaps not surprisingly, participants in our study value their PII higher in days that are unusual when compared to typical days.

This result suggests that not all PII even within the same category and level of complexity is valued equally by our participants, which has a direct implication for personal data markets and for services that monetize mobile personal data.

CONCLUSION

We have investigated the monetary value that people assign to their PII as it is collected by their mobile phone. In particular, we have taken into account four categories of PII (location, communication, apps and media) with three levels of complexity (individual, aggregated and processed). We have carried out a comprehensive 6-week long study in a living lab environment with 60 participants adopting a Day Reconstruction Method along with a reverse second price auction mechanism to collect honest monetary valuations.

We have found that location is the most valued category of PII and that bulk information is valued much higher than individual information (except for the media category). We have identified individual differences in bidding behaviors which are not correlated with socio-demographic traits, but are correlated with behavior (mobility and app usage) and intentional self-disclosure.

Finally, we have found that participants trust themselves with their PII above banks, telcos and insurance companies and that unusual days are perceived as *more valuable* than typical days.

REFERENCES

1. A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 2:24–30, 2005.
2. A. Acquisti, L. John, and G. Loewenstein. What is privacy worth. In *Twenty first workshop on information systems and economics (WISE)*, pages 14–15, 2009.
3. E. Adar and B. Huberman. A market for secrets. *First Monday*, 6(8), 2001.
4. N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland. Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659, 2011.
5. Y. Amichai-Hamburger and G. Vinitzky. Social network use and personality. *Computers in Human Behavior*, 26(6):1289–1295, 2010.
6. C. Aperjis and B. A. Huberman. A market for unbiased private data: Paying individuals according to their privacy attitudes. *arXiv preprint arXiv:1205.0030*, 2012.
7. L. Barkhuus, B. Brown, M. Bell, S. Sherwood, M. Hall, and M. Chalmers. From awareness to repartee: sharing location within social groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 497–506. ACM, 2008.
8. J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira. Your browsing behavior for a big mac: Economics of personal information online. In *Proceedings of the 22nd international conference on World Wide Web*, pages 189–200, 2013.
9. M. Cherubini and N. Oliver. A refined experience sampling method to capture mobile user experience. *CoRR*, abs/0906.4125, 2009.
10. E. Christofides, A. Muise, and S. Desmarais. Hey mom, whats on your facebook? comparing facebook disclosure and privacy in adolescents and adults. *Social Psychological and Personality Science*, 3(1):48–54, 2012.
11. S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 81–90. ACM, 2005.
12. P. T. Costa and R. R. McCrae. The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, 2:179–198, 2008.
13. P. Cozby. Self-disclosure: A literature review. *Psychological Bulletin*, 79(2):73–91, 1973.
14. 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*, pages 109–118. ACM, 2006.
15. G. Danezis, S. Lewis, and R. J. Anderson. How much is location privacy worth? In *WEIS*, volume 5. Citeseer, 2005.
16. Y.-A. de Montjoye, S. S. Wang, A. Pentland, D. T. T. Anh, A. Datta, et al. On the trusted use of large-scale personal data. *IEEE Data Eng. Bull.*, 35(4):5–8, 2012.
17. T. Farma and I. Cortivonis. Un questionario sul ”locus of control”: suo utilizzo nel contesto italiano (a questionnaire on the locus of control: its use in the italian context. *Ricerca in Psicoterapia*, 2, 2000.
18. J. Fogel and E. Nehmad. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25(1):153–160, 2009.
19. L. Goodman. Snowball sampling. *Annals of Mathematical Statistics*, 32:148–170, 1961.
20. I.-H. Hann, K.-L. Hui, S.-Y. T. Lee, and I. P. Png. Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems*, 24(2):13–42, 2007.
21. J. I. Hong and J. A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services*, MobiSys ’04, pages 177–189. ACM, 2004.
22. M. G. Hoy and G. Milne. Gender differences in privacy-related measures for young adult facebook users. *Journal of Interactive Advertising*, 10(2):28–45, 2010.
23. B. A. Huberman, E. Adar, and L. R. Fine. Valuating privacy. *Security & Privacy, IEEE*, 3(5):22–25, 2005.
24. K.-L. Hui, H. H. Teo, and S.-Y. T. Lee. The value of privacy assurance: An exploratory field experiment. *MIS Quarterly*, 31(1):19–33, 2007.
25. 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.
26. I. A. Junglas, N. A. Johnson, and C. Spitzmüller. Personality traits and concern for privacy: an empirical study in the context of location-based services. *European Journal of Information Systems*, 17(4):387–402, 2008.
27. D. Kahneman, A. Krueger, D. Schkade, N. Schwarz, and A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306:1776–1780, 2004.
28. B. P. Knijnenburg, A. Kobsa, and H. Jin. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, 71(12):1144–1162, 2013.
29. M. Korzaan, N. Brooks, and T. Greer. Demystifying personality and privacy: An empirical investigation into antecedents of concerns for information privacy. *Journal of Behavioral Studies in Business*, 1, 2009.

30. N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A survey of mobile phone sensing. *IEEE Communications Magazine*, 48(9):140–150, 2010.

31. J. Lin, G. Xiang, J. I. Hong, and N. Sadeh. Modeling people's place naming preferences in location sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pages 75–84. ACM, 2010.

32. J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2409–2418. ACM, 2011.

33. J. Lo. Privacy concern, locus of control, and salience in a trust-risk model of information disclosure on social networking sites. In *AMCIS*, page 110, 2010.

34. A. Madan, M. Cebrián, S. T. Moturu, K. Farrahi, and A. Pentland. Sensing the "health state" of a community. *IEEE Pervasive Computing*, 11(4):36–45, 2012.

35. R. Mayer and J. Davis. The effect of the performance appraisal system on trust for management: a field quasi-experiment. *Journal of Applied Psychology*, 84:123–136, 1999.

36. R. Mc Afee and J. Mc Millan. Auctions and bidding. *Journal of Economic Literature*, 25:699–738, 1987.

37. G. Mesch and G. Beker. Are norms of disclosure of online and offline personal information associated with the disclosure of personal information online? *Human Communication Research*, 36:570–592, 2010.

38. M. Mun, S. Hao, N. Mishra, K. Shilton, J. Burke, D. Estrin, M. Hansen, and R. Govindan. Personal data vaults: A locus of control for personal data streams. In *Proceedings of the 6th International Conference*, Co-NEXT '10, pages 1–12, 2010.

39. A. Pentland. Society's nervous system: Building effective government, energy, and public health systems. *IEEE Computer*, 45(1):31–38, 2012.

40. M. Perugini and L. Di Blas. The big five marker scales (bfms) and the italina ab5c taxonomy: Analyses from an emic-etic perspective. In B. de Raad B. and M. Perugini, editors, *Big Five Assessment*. Gottingen: Hogrefe and Huber Publishers, 2002.

41. D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski, and J. Crowcroft. The personality of popular facebook users. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 955–964. ACM, 2012.

42. C. Riederer, V. Erramilli, A. Chaintreau, B. Krishnamurthy, and P. Rodriguez. For sale : Your data: By : You. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, HotNets-X, pages 13:1–13:6, New York, NY, USA, 2011. ACM.

43. J. Rotter. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1965.

44. J. B. Rotter. A new scale for the measurement of interpersonal trust. *Journal of personality*, 35(4):651–665, 1967.

45. J. Schrammel, C. Köffel, and M. Tscheligi. Personality traits, usage patterns and information disclosure in online communities. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, pages 169–174. British Computer Society, 2009.

46. K. P. Tang, J. I. Hong, and D. P. Siewiorek. Understanding how visual representations of location feeds affect end-user privacy concerns. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 207–216. ACM, 2011.

47. E. Toch and I. Levi. Locality and privacy in people-nearby applications. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 539–548. ACM, 2013.

48. J. Tsai, P. Kelley, L. Cranor, and N. Sadeh. Location sharing technologies: Privacy risks and controls. *I/S: A Journal of Law and Policy for the Information Society*, 6(2):119–151, 2010.

49. J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254–268, 2011.

50. M. Van Kleek, D. A. Smith, N. Shadbolt, and m. schraefel. A decentralized architecture for consolidating personal information ecosystems: The webbox. In *PIM*, pages 177–189, 2012.

51. H. Varian, F. Wallenberg, and G. Woroch. The demographics of the do-not-call list. *IEEE Security & Privacy*, 3:34–39, 2005.

52. R. Want, T. Pering, G. Danneels, M. Kumar, M. Sundar, and J. Light. The personal server: Changing the way we think about ubiquitous computing. In *In Proceedings of 4th International Conference on Ubiquitous Computing*, pages 194–209, 2002.

53. L. Wheeless and J. Grotz. Conceptualization and measurement of reported self-disclosure. *Human Communication Research*, 2:338–346, 1976.

54. J. Wiese, P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 197–206. ACM, 2011.

55. A. L. Young and A. Quan-Haase. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of the fourth international conference on Communities and technologies*, pages 265–274. ACM, 2009.