Why People Are (Un)willing to Share Information with Online Advertisers

Pedro Giovanni Leon, Ashwini Rao, Florian Schaub, Abigail Marsh, Lorrie Faith Cranor, Norman Sadeh

May 2015
CMU-ISR-15-106

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Online advertisers track Internet users' activities to deliver relevant ads. To study how different online advertisers' data practices affect users' comfort with sharing their information, we conducted a between-subjects online study with 1,882 participants. We asked participants about their comfort with sharing commonly collected types of information in scenarios with varying data practices, and studied their reactions to a realistic example of a behavioral profile created by advertisers about users. We found that participants' willingness to share information with online advertisers is not only based on the sensitivity of the information, but also on the scope of collection and use, perceived necessity of collection, and perceived benefits or harms of disclosing specific data types. Qualitative data analysis revealed nuanced and contextualized reasons behind stated information-sharing preferences. Participants were particularly adverse to sharing information that they perceived as irrelevant for advertising. However, our results also reveal that—under the right circumstances—participants may be willing to share their data with advertisers to enhance the utility of shown ads.

This research was supported in part by the National Science Foundation under grants CNS 13-30596 (Towards Effective Web Privacy Notice & Choice: A Multi-Disciplinary Perspective) and DGE-0903659 (IGERT: Usable Privacy and Security).

Keywords: Privacy, User Preferences, Online Behavioral Advertising, OBA, Tracking, Data Retention, Do Not Track.

1 Introduction

Online advertising has developed into the de facto business model for sustaining free online services. Online Behavioral Advertising (OBA), also known as targeted advertising, aims to improve the effectiveness of online advertising by using data about users' online and offline activities to serve more relevant ads. Often advertising companies create detailed behavioral profiles of users [20]. To do so, they collect information that may include information provided by users to websites, such as email addresses, and information inferred from users' browsing activities, such as personal interests [4]. This data may be enriched with data from other sources such as third-party databases, public records and offline purchases [4]. Collected information may be shared with affiliates, marketing companies and data aggregators.

Internet users generally dislike being tracked [19]. Studies show that users have privacy concerns regarding OBA [24, 16] and creation of online behavioral profiles [20]. However, users' dislike for OBA is not absolute: users' dislike for OBA varies by type of data collected, retention period, scope of use and other OBA data practices [13]. In this work, we further investigate how OBA data practices influence users' willingness to share data with advertisers. We extend prior work in the following ways. First, prior work has largely focused on understanding how much people are willing to share data under different OBA practices [13], but not their reasons for doing so. We identify the reasons why users are willing or unwilling to share their data under different OBA practices. Second, we examine why users want to access and edit their behavioral profiles and how allowing access can benefit both users and advertisers.

We conducted a between-subjects online study on Amazon Mechanical Turk with 1,882 US participants. We studied seven data collection and sharing scenarios, three retention periods and ten information types. Our study included both fixed-response and open-ended questions to collect quantitative and qualitative data. Our results indicate that participants have several reasons for not sharing their information with advertisers. For example, participants were reluctant to share information that they felt was irrelevant for advertising or did not reflect their purchasing interests. However, our results also suggest that participants see benefits in sharing information with advertisers, for example, receiving relevant ads. Our results highlight how context of collection, purpose of collection, and attitudes towards targeted ads and privacy affect users' reasons for sharing information with online advertisers. Lastly, our results show that allowing access can benefit both users and advertisers.

The rest of this report is organized as follows. In Section 2, we give an overview of OBA practices and review related work. In Section 3, we discuss our study methodology including design, implementation and analysis. In Section 4, we discuss our results regarding why users are willing or unwilling to share information with advertisers and how allowing access can increase their willingness to share. Lastly, in Section 5, we consider the implications of our results and how they may influence public policy.

2 Background and Related Work

Online Behavioral Advertising (OBA) or targeted-advertising may contribute significantly to advertising revenue because of higher click-through rates compared to non-targeted ads [2]. However, whether OBA is always effective is questionable. Farahat and Bailey found that when pre-existing consumer interest is considered, OBA may not benefit the advertiser [7]. Lambrecht and Tucker analyzed data from a travel website and found that general-audience, non-targeted ads performed better on average than targeted ads [11]. Tucker examined social advertising on Facebook and found that users responded more positively to social ads based on the standard Facebook algorithm than to targeted-ads based on Facebook social connections [23].

Advertisers can employ different data practices regarding collection, sharing, use etc. They may collect different types of data including demographic data (e.g.gender, date of birth, household income), psychographic data (e.g. personality type, interests, lifestyle) and behavioral data (e.g. loyalty, purchase history) [20]. Advertisers may share data with other entities such as affiliates and data brokers [6]. They may retain data for varying periods of time and sometimes infinitely [6]. Advertisers may use the collected data for the purpose of behavioral advertising and other purposes such as website analytics or marketing research. Some advertisers allow users to access profiles created about them [5].

Prior studies show that advertisers' data practices can influence users' willingness to share data for the purpose of OBA. For example, a study by Leon et al found that extent of sharing, purpose for which data is used and retention period can significantly impact users' willingness to share information [13]. Users' willingness to share data for OBA can depend on the data type. Joinson et al. found that participants were least willing to share sensitive financial data [10]. Similarly, Treiblmaier and Pollach found that users consider sharing credit card information for personalization most risky [22]. Leon et al. found that almost half of their participants were unwilling to share any type of data, and the rest would share non-sensitive information such as gender and hobbies [13]. Prior work has largely focused on understanding what kind and how much information people are willing to share under different OBA practices. In this work, we focus both on how much users are willing to share as well as their reasons for being willing or unwilling to share information for advertising purposes.

Another data practice that may influence user preferences is whether advertising companies allow users to access data collected about them. Users may have several concerns about the contents of their behavioral profiles [20]. Prior research on the benefits of allowing access to profiles has been limited. Prior studies have suggested that allowing users to access their data may be beneficial to both users and advertisers [20, 22]. However, a prior study by Leon et al. found that allowing access had minimal impact on users' willingness to share data. In this work, we investigate further benefits of access. To study benefits of access, we employ a more realistic setting. We show our participants a profile based on profiles found in the wild [20] and enriched with their own demographic information.

Many studies have found that users are generally concerned about OBA. Many do not want third parties to track and profile them online [9, 16]. Agarwal et al. found that users were particularly sensitive to being shown embarrassing ads as a result of OBA [1]. However, Ur et al. found that participants saw potential benefits of OBA for both users and companies, but they were concerned about the lack of transparency and control over OBA practices [24]. We also study user concerns and surprises regarding OBA, but focus on how their concerns and surprises vary by specific OBA data practices.

Benisch et al. [3] study people's willingness to disclose their location, including disclosing location to advertisers. They find that contextual factors, especially location and time, impact users' location sharing preferences. Mugan et al. [17] investigate how people's willingness to share their location, including sharing with advertisers, can be captured by a small number of privacy preference profiles. Lin et al. [14] extend this concept in the context of mobile apps for different types of information disclosures. Their privacy preference profiles take into account the purpose associated with the sharing of sensitive information with third parties, including advertising networks, analytics companies, and social networks. Our results can contribute to the creation of nuanced privacy preference profiles.

3 Methodology

To understand how advertisers' data practices influence an individual's willingness to share information for advertising purposes, we conducted a between-subjects study as an an online survey on Amazon Mechanical

Turk (MTurk).[1] We recruited 1,882 participants who were residents of the United States and who were at least 18 years of age. Compensation was $1.5 and average completion time was 22 minutes. Our study was approved by CMU's IRB. Below, we discuss our study design, survey implementation, and our analysis approach.

3.1 Study Design

Our goal was to understand how advertisers' data practices would impact users' willingness to share if users were aware of such practices. Our main research questions were:

- Does the type of data collected matter? For example, does willingness vary by personal, financial or predictive information?
- Do variations in scope of collection and sharing affect users' willingness to share information? For example, are users more comfortable when their data is collected and used on a single website versus on multiple websites? Similarly, does a shorter retention period make them more comfortable?
- Is a limited purpose specification more conducive to sharing than a vague purpose?
- What do users think of the profiles that advertisers create about them, and do they see benefit in accessing their own profiles?

To answer these questions, we assigned each participant to one condition that described a particular data practice scenario. We described our scenarios in the context of a hypothetical news website (AllNews) where an advertising company (Best Ads or Facebook) would collect and use data from visitors. We considered seven scopes of collection and use, and three retention periods (one week, three months, one year), resulting in 21 conditions in total. We gradually increased the breadth of data collection and usage between subsequent scopes. Table 1 provides an overview of our scenarios.

In scenarios S1 and S2, Best Ads may only collect data and show targeted ads on the visited websites (AllNews). However in S2, Best Ads may also use the collected information for "other purposes." No restrictions for other purposes were mentioned. Scenarios S3–S6 extended Best Ads' collection of data and showing of targeted ads to include the AllNews website, as well as other news, entertainment, travel, and retail websites. S4 extended this scenario with the collection of information from a local department store to reflect the common practice of linking online activities with offline purchasing behavior. S4, further extended the use of collected information to include targeted ads, as well as offering targeted coupons from the local department store. Scenarios S5 and S6, extended S3 with collection for "other purposes," similar to S2, but with the larger collection and use scope defined in S3. However, S5 included the restriction that Best Ads may not share information with other parties, while in S6 no restrictions for other purposes were mentioned. Scenario S7 focused on Facebook's data practices. The scenario described that Facebook may collect information from the AllNews website and participants' Facebook page to show targeted ads only on Facebook.

While each participant saw only one of the scenarios, all participants were shown a realistic example of a behavioral profile that advertisers may create about users [20]. We were interested in what surprised or concerned participants about the profile, and perceived benefits of being able to access or edit the information in their profiles.

[1] www.mturk.com

	Scenario	Description
S1	AllNews	Best Ads may collect information and show targeted ads only on the AllNews website.
S2	OtherPurposes	S1 + Best Ads may use the collected information for other purposes.
S3	Websites	Best Ads may collect information and show targeted ads on the AllNews and other websites.
S4	Offline	S3 + Best Ads may collect information from a local department store and give targeted coupons for the store.
S5	Websites&OtherPurposes-NoShare	S3 + Best Ads may use collected information for other purposes, but not share it with other parties.
S6	Websites&OtherPurposes	S3 + Best Ads may use collected information for other purposes, no restrictions given.
S7	Facebook	Facebook may collect information on the AllNews website and users' Facebook page to show targeted ads on Facebook.

Table 1: Scenarios of the different conditions, each was tested with retention periods of one week, three months, and one year (21 conditions in total).

3.2 Survey Implementation

We told participants that the survey was intended to "understand how people experienced the Internet." Using a round-robin scheme, we assigned each participant to one of the 21 conditions that each included one scenario describing how an advertising company would collect and use data from visitors to a hypothetical news website.

The survey consisted of four sections. The first section asked for participant demographics, Internet use, and opinions about online advertising. To signal that the survey required more than minimal effort, we started with an open-ended question asking for their opinion about online advertising.

In the second section, we confronted participants with one scenario from Table 1. We first asked participants to visit the AllNews website, a static website we modeled after the CNN.com homepage with changed branding logos and text. Hyperlinks and forms were disabled. The main page of the AllNews website is shown in Appendix A. To verify that participants were following instructions, we required them to identify the title of a news article that appeared on the AllNews homepage, presented among four decoy titles.

Next, we asked participants to imagine that they were users of the AllNews website and provided a short explanation of how targeted ads work. Then we told them that the AllNews website had contracted with a company that was interested in showing users targeted ads and informed them about the scenario-specific data practices. The specific company and their data practices varied depending on scenario. We created scenarios based on real ad companies' data practices, including data collection, usage, and retention practices, as outlined in Section 3.1. Appendix B shows a sample scenario. We asked participants to read the given scenario thoroughly and then assessed their understanding with a follow-up question about the stated data practices. Participants who answered incorrectly were shown the correct answers, asked to read the scenario again, and then tested again.

We then collected participants' willingness to share the ten different types of personal information shown in Figure 1. Namely, IP address, operating system, email address, browsing activity (e.g., articles read, videos watched, pages visited), ZIP code, purchasing interests, income bracket, credit score bracket, gender, sexual orientation. These data types were chosen based on what advertising companies typically collect or

infer. For each item, participants rated their comfort with sharing that item by indicating their level of a agreement with the statement "I would be comfortable if *[the company]* collected or otherwise inferred the following information about me" on a five-point Likert scale ("Strongly Disagree" to "Strongly Agree"). We followed up with open-ended questions asking them to explain why they would or would not be comfortable with the advertiser collecting those data types. Next, we presented follow-up questions for at most four data types selected from the ones for which they indicated agreement or disagreement. Only four data types were randomly selected to avoid fatigue.

In the third part, we showed participants a realistic example of a profile. Rather than asking participants to access their own profiles kept by major online advertisers, which may fluctuate in content and may require registration, we showed them a sample profile (see Appendix C) created by combining data collected from real user profiles [20]. We contextualized the sample profile to each participant by dynamically adapting the first two categories in the profile (location, individual demographics) to the participant's IP address and their information provided at the beginning of the survey. We first asked the participants to select from a list two items that appeared in the sample profile to ensure that they read it. We then elicited participants' surprise and concerns about the profile's content and asked for perceived benefits (if any) of having access to their profile. We ended the survey with eight questions from the Internet Users' Information Privacy Concerns (IUIPC) instrument [15] to gauge participants' general privacy concerns.

3.3 Analysis Approach

We cleaned the data by removing participants from outside the US (39); that completed the survey in <5 minutes (3); were inconsistent in whether they visited news websites regularly (27); failed the AllNews website (120) or sample profile (13) content questions. We analyzed valid responses from 1,882 participants aged 18–79 (mean=34, σ=12.2). Half the participants were female. Participants predominantly indicated low (40%), medium (33%), or high (20%) Internet literacy, and few participants indicated no Internet literacy (7%). Participants exhibited a diverse range of occupations, and were well educated (31% some college, 10% Associate's degree, 35% Bachelor's degree, 15% Graduate degree). We did not observe any statistical differences between conditions concerning education, tech savviness, gender, age, or Internet literacy.

3.3.1 Quantitative Analysis

We performed Kruskal-Wallis rank sum tests for each of the ten assessed data types to determine for which data types the scenario and retention period had a statistically significant impact. We found that the scenario had a significant impact on willingness to share some types of information, but retention period did not.

We then performed binary logistic regressions on all data types using the scenarios and retention periods (to verify its null effect) as independent variables. The willingness to share questions served as dependent variables with "strongly agree" and "agree" responses binned as "agreement," and "neutral," "disagree," and "strongly disagree" responses binned as "non-agreement." In addition to scope and retention, our regression models controlled for participants' age and gender, and included indicator variables for privacy concerns, positive opinion of targeted ads, usage of ad blocking tools, positive opinion of the AllNews website, Facebook account, tech savviness, and whether participants answered correctly at least one of the scenario understanding questions.

3.3.2 Qualitative Analysis

We qualitatively analyzed multiple open-ended questions: participants' reasons for comfort/discomfort with sharing certain information types, surprise and concerns about the sample profile content, and perceived benefits of accessing their own profiles. All questions were shown dependent on a participant's answers to preceding Likert-scale questions. For instance, we asked about their reasons for being concerned if they indicated concern about the sample profile. In addition, participants assigned to conditions mentioning "other purposes," were asked what those might be, in order to understand whether they had positive or negative associations.

For each open-ended question, we randomly selected a 10% sample of the respective responses, drawn evenly from all 21 conditions, for qualitative data analysis. Considering the large total of participants, this provided us with a sufficiently large sample per question for qualitative data analysis (138–199 responses per question). Due to random sampling within each condition we were confident that the selected responses are representative of the whole dataset, which we confirmed with cursory inspection of the remaining responses. For the "other purposes" responses, all 792 responses were coded as *positive*, *negative*, or *ambiguous* in order to enable integration in the regression models.

For each open-ended question, two researchers independently evaluated the same subset of responses to derive relevant codes from which a question-specific coding taxonomy was jointly developed. Next, they coded the full sample of responses. Initial inter-coder reliability was evaluated with Cohen's Kappa coefficient. Coding disagreements were subsequently resolved on a per-statement basis in an iterative process between the two coders, resulting in fully reconciled response annotations for each open-ended question, which were used in subsequent analysis. In total, 2,245 statements were examined as part of qualitative data analysis, resulting in 2,919 assigned codes.

3.4 Limitations

Our analysis is based on self-reported data regarding participants' sharing comfort. We created a browsing simulation scenario, asking participants to visit a news website to emulate a real Internet experience, while not a perfect substitute for behavioral data, the fact that participants reacted differently to different scenarios and showed to be invested in the study by providing rich qualitative data, suggests that users provided with effective notice may behave accordingly to their stated preferences. Furthermore, our results provide clear indications of which data types web users are willing to receive ads based on and which ones are considered too sensitive or irrelevant for advertising use.

MTurk users have known demographic differences compared with the general Internet population; however, it has been demonstrated that MTurk participants behave similarly in studies as subjects recruited from other sources [18].

4 Results

First we discuss how different factors affected participants' comfort with sharing different types of data. We provide both statistical and qualitative evidence of participants' sharing preferences. We then discuss participants' reaction to an exemplary behavioral profile contextualized to their individual demographics.

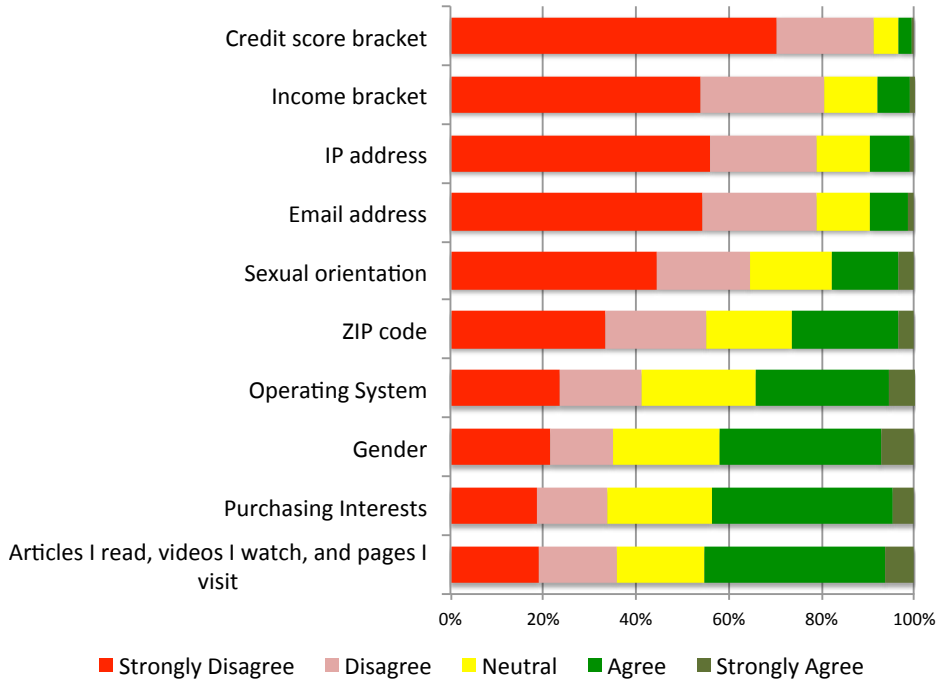

Figure 1: Participants' responses to the statement, "I would be comfortable if *[Best Ads / Facebook]* collected or otherwise inferred the following information about me." The green shades represent sharing comfort, while the red shades indicate discomfort.

4.1 Sharing Preferences

Overall, almost half the participants (45%) were comfortable sharing information with advertising companies. They were most comfortable sharing the pages visited, articles read, and videos watched on the news website (45%), the products they might be interested in purchasing (44%), gender (42%), computer's operating system (35%), ZIP code from where they access the Internet (26%), as well as sexual orientation (17%). Only a small fraction of participants were comfortable sharing their email address (9.7%), IP Address (9.6%), income bracket (7.9%), or credit score (3.3%). Figure 1 shows overall willingness to share.

Participants' sharing comfort did not only depend on the sensitivity of information, but also the scope of collection and use, necessity of collection, and perceived benefits and harms of disclosure. Furthermore, personal attitudes such as trust in the visited website, the opinion of targeted ads, and privacy concerns had a strong effect on willingness to share.

We discuss how the elements in the tested scenarios impacted participants' comfort with sharing different data types. First we present quantitative results, followed by qualitative results that provide further insights into participants' decision making process. Overall, our results help to explain why participants might not be willing to share even apparently innocuous information under some circumstances, but why they might be willing to share arguably more personal information under other circumstances. Finally, using results from our regression models, we discuss how personal attitudes towards targeted ads, trust perceptions, and privacy concerns affected participants' sharing comfort.

4.1.1 Factors Affecting Sharing Comfort

Statistical analysis provided evidence that participants took into account the scope of collection and purpose of use to make information sharing decisions. Non-parametric analyses of variance showed general differences between scenarios for six information types: interactions with the AllNews website, purchasing interests, gender, ZIP code, sexual orientation, and email address. However, retention period was not a significant factor in predicting willingness to share for any information type. We validated these results with our regression models, which also allowed us to investigate the particular direction of the effects. The regression model results are provided in Appendix D. Participants' willingness to share their online interactions and purchasing interests decreased as the scope of collection and use increased, while their willingness to share their gender, email address, ZIP code, and sexual orientation was more nuanced.

Online interactions. Participants' comfort sharing their online interactions was similar (49%-53%) in scenarios where this data type was exclusively used for targeted ads either on the first-party website (S1) or other visited websites (S3), as well as when it was used for targeted ads and other purposes only on the first-party website (S2). Participants were statistically less comfortable sharing (βs from -0.62 to -0.55, p-value < 0.001) this data type (34%-41%) in scenarios where the information would be linked with offline data to receive coupons (S4), used on other websites for unspecified purposes (S5 & S6) and shared with Facebook (S7).

Purchasing interests. Participants' comfort sharing their purchasing interests was similar (45%-51%) in scenarios where this data type was exclusively used for targeted ads either on the first-party website or other visited websites (S1 & S3, 46%), used for targeted ads and other purposes only on the first-party website (S2, 51%), and even when linked with offline data to receive coupons (S4, 45%). However, participants were statistically less comfortable sharing this data type (38%) in scenarios where the information was going to be used on other websites also for other unspecified purposes (β=-0.44, p-value=0.03) and shared with Facebook (β=-0.35, p-value=0.07).

Gender. None of the scenarios (35%-47%) were statistically different (α = 5%) in our regression model from the baseline scenario (S1, 40%). However, a larger fraction of participants was comfortable sharing their gender with Facebook (S7, 47%) and when used for targeted ads and other purposes only on the first-party website (S2, 47%), compared with scenario S6 where the information would be used on other websites and for "unspecified purposes" (35%).

ZIP code. As in the case of gender, fewer participants (21%) in scenario S6 were willing to share this data type than participants in all other scenarios (25%-34%). This difference was significant (β=-0.64, p-value=0.005) when compared with the baseline scenario (S1, 29%). For both the previous and this data type, the "unspecified" purposes seemed to negatively impact willingness to share, likely because participants assumed more "negative" than "positive" purposes for these data types.

Sexual orientation. Participants' comfort sharing their sexual orientation was low and similar in scenarios where this data type was used for targeted ads either on the first-party website (S1, 13%) or other visited websites (S3, 16%), going to be linked with offline data to receive coupons (S4, 18%), and used on other websites for other unspecified purposes (S6, 13%). However, participants were statistically more comfortable sharing this data type in scenarios where the information was going to be shared with Facebook (S7,

23%, β=0.8, p-value<0.001) or used on other websites for other unspecified purposes but without sharing it with third-parties (S5, 21%, β=0.56, p-value=0.04).

Email address. While only few participants (10%) were comfortable sharing their email address, a larger fraction of them were comfortable sharing it to receive coupons (S4, 13%, β=0.56, p-value=0.07) and with Facebook (S7, 16%, β=0.88, p-value=0.002) when compared to scenario S1 (8%).

Computer and sensitive information. None of the scenarios affected participants' willingness to share either computer's information such as IP Address and operating system or more sensitive information such as income bracket of credit score.

Effect of unspecified purposes. For the scenarios that included "other purposes" uses, we coded participants' interpretations of those purposes as *positive* (e.g., suggesting content, measuring success of ad targeting, observing consumer trends), *negative* (e.g., selling information to other companies, creating mailing lists, sharing with the government), or *ambiguous* (e.g., participant uncertain or response unclear). The majority of participants perceived other purposes as negative (52%), some as positive (35%), and a smaller number of responses were ambiguous (13%).

We included this variable in our regression models to evaluate whether opinions about "other purposes" had an impact on sharing comfort. A positive perception of "other purposes" had a positive impact on the level of comfort for sharing online activities (β=0.32, p=.07), gender (β=0.32, p=.07), and ZIP code (β=0.46, p=.002). It is important to mention that to measure the effect of "other purposes" we had to split our sample into three subsets (positive, negative, ambiguous), decreasing statistical power.

Taken together, these quantitative results suggest that participants paid attention to the tradeoffs presented in their given scenario. We now turn to a qualitative discussion that provides further insights into the participants' reasoning.

4.1.2 Why people would share information

We asked participants open-ended questions to understand why they were comfortable sharing some data types with advertisers. We coded and qualitatively analyzed a representative 10% sample consisting of 206 participant responses, resulting in 255 coded statements leading to the reasons shown in Table 2. The two main reasons why participants were willing to share data with advertisers were receiving relevant advertisements (25.8%), and feeling that the data was public rather than personal, or private (18.8%). The top data types that participants considered public were operating system information, gender and online activities. Note that when participants considered a data type as public they were more willing to share it for receiving relevant advertisements.

Participants (12.5%) felt that sharing data types such as gender or operating system did not matter to them. For example, they stated, "...it is not big deal" or "I really don't care," but did not provide further explanations. Some participants (9%) mentioned that they were comfortable sharing general information, like ZIP code and operating system, as it could not be used to personally identify them. Other participants (4.7%) saw no harm in sharing their data, because at most advertisers would send them ads. Some argued that there was some harm, but not much. One participant said regarding sharing operating system data, "I suppose it could maybe cause some security issues but other than that it doesn't bother me."

Participants also expressed data type-specific reasons for sharing. For example, participants wanted to share ZIP code to receive location-specific benefits, like local deals and news. Participants felt comfortable

Reason	Count	Percent
Receive relevant ads	66	25.8%
Not personal/secret/private	48	18.8%
Does not matter	32	12.5%
Not personally identifiable	23	9.0%
Required to provide relevant service	18	7.0%
No harm in sharing	12	4.7%
Easy to infer	8	3.1%
Technical aid	7	2.7%
Not embarrassing	7	2.7%
Receive better deals	7	2.7%
Location targeting	7	2.7%
Not privacy invasive	6	2.3%
Other	14	5.5%

Table 2: Reasons why participants are willing to share data with advertisers. Participants could provide multiple reasons (255 codes=100%, n=206).

sharing information about the articles, videos and pages they visit to receive better service, for example, recommendations for interesting news articles. Participants wanted to share information about the products they were interested in to receive discounts. A few participants felt that operating system information was required for the website to display properly on their computer. Some participants indicated that they were proud of their sexual orientation and were not embarrassed to share it.

We analyzed why participants were more comfortable sharing certain data types for specific scenarios. Recall that participants were significantly more willing to share email in the Facebook (S7) and offline (S4) scenarios compared to other scenarios. Participants in the offline scenario were willing to share email to get better deals and services. In the Facebook scope they felt that they had already voluntarily provided their email to Facebook. Participants were more willing to share gender in the Facebook scenario as some did not care if Facebook knew their gender, and others felt that their gender was not a secret. Participants were also more comfortable sharing their sexual orientation in the Facebook scenario (S7). They were also more comfortable when sexual orientation could be used for other purposes, but would not be shared with other parties (S5). Participants' felt that it was no big deal, no harm would come out of it, no shame, or that it did not point to their identity. They also felt that services could be tailored to their interests. One participant said, "they can't do anything knowing that," and another mentioned, "the ads and services could be tailored to feature products that are in line with my lifestyle."

4.1.3 Why people would not share information

We further investigated reasons for feeling uncomfortable with sharing data with advertisers. In this case, the representative 10% sample analyzed consisted of 575 participant responses, resulting in 786 coded reasons. The larger number of coded statements reflects the higher percentage of participants not willing to share information with advertisers. We extracted the reasons shown in Table 3. Overall, participants were not comfortable sharing data they considered personal information (23.8%), there was no need to know (14.5%), or was unnecessary for advertising (12.2%). Note that the reasons for unwillingness to share are almost opposites of those for willingness to share, discussed previously. For example, top reasons for willingness to share included considering information to be non-personal and also relevant for advertising.

Reason	Count	Percent
Personal information	187	23.8%
None of their business	114	14.5%
Unnecessary for advertising	96	12.2%
Invasion of privacy	81	10.3%
Location tracking	50	6.4%
Ad spam	43	5.5%
Lack of consent	42	5.3%
Inference of information	24	3.1%
Personally identifiable	22	2.8%
General tracking	20	2.5%
Computer harm	20	2.5%
Unreliable information	18	2.3%
Other	68	8.6%

Table 3: Reasons why participants would not share data with advertisers. (786 codes=100%, n=575).

As with willingness to share, participants' unwillingness to share varied by data types. Participants were mainly concerned about their location being tracked if they shared ZIP code or IP Address. And the main reason why participants would not like sharing their email address was because they would not want to receive unsolicited emails. Participants were unwilling to share IP address with advertising companies, because they considered it personally identifiable information. Furthermore, certain data types such as gender, income bracket or online activities were not perceived as reliable indicators of their interests for ads. One participant stated, "I don't see how what I read about accurately reflects any products or services that I would or could actually purchase or even be interested in."

The majority of participants were not comfortable sharing credit score, income bracket, and sexual orientation because they considered these types of information personal, unnecessary for advertising, or thought it was nobody's business. Some participants viewed the collection of these types of information as an invasion of their privacy. Some participants expressed concerns about discrimination based on price, gender or sexual orientation. Some did not trust Facebook and other advertisers. They felt that advertising companies might sell or share their data with third-parties. A few believed that their data may be stored insecurely, or that they may become victims of identity theft. Lastly, some participants thought that sharing data such as gender may increase the risk of assault or physical harm.

Participants considered what they did on Facebook personal and believed that they could not use Facebook freely if Facebook tracked their habits and activity history. For the offline scenario, participants were uncomfortable combining offline and online information. Participants in the "other purpose scenarios" considered collecting online activities as too much collection of information and compared it to spying on them. One of the participants said, "That is personal information. I feel like they are spying on me if they know that. That makes me feel uncomfortable."

Participants were uncomfortable sharing purchasing interests for Facebook (S7) and other purpose scenarios (S2, S5 & S6). Participants did not want Facebook to know too much of their browsing and shopping habits. They were concerned about Facebook hounding them with ads, and that Facebook could announce or share purchases that were deemed personal. Participants in the "other purpose" scenario were uncomfortable sharing purchase interests for several reasons including invasion of privacy, not relevant to their interests, unwilling to purchase from ads, and because they did not specifically give permission for collecting such information. Participants expressed that location could be inferred from ZIP code, and their location, or

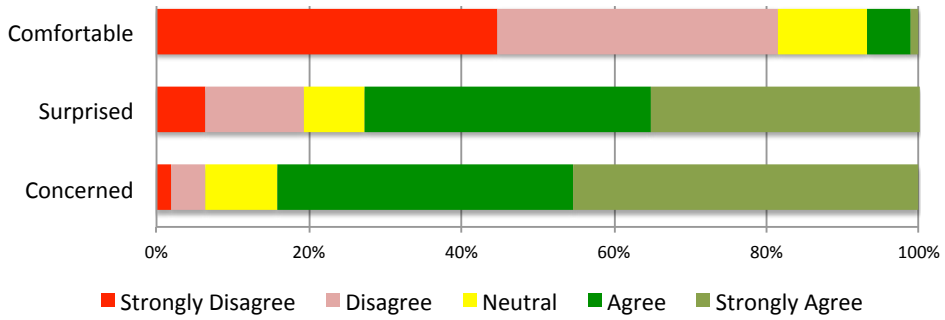

Figure 2: Participants' responses to the statement, "I am *comfortable / surprised / concerned* with the information that such profiles may contain." Participants were not comfortable, but surprised and concerned.

where they live, was personal information. They felt that sharing such personal information was invasion of privacy. They also did not want to be targeted by local businesses.

4.1.4 Other factors affecting disclosure

Based on our regression models, we further explored the impact of participants' personal characteristics such as gender, age, and attitudes towards privacy, targeted ads and the AllNews website. Attitudinal variables showed a strong effect on sharing comfort for all explored data types. In particular, positive opinion about targeted ads and about the AllNews website significantly increased sharing comfort (βs ranged from 0.58 to 1.87, p-values $<$.001). On the other hand, the more privacy concerned participants were, the less comfortable they were sharing any data type (βs ranged from -0.79 to -0.42, p-values $<$.002). Demographics had a milder effect. Participants' gender was only significant for the ZIP code and gender data types, where male participants were more comfortable sharing those types of information than female participants (p-values $<$.002).

4.2 Reaction to User Profile

We further showed participants a realistic example of a behavioral profile and elicited their reactions. Qualitative analysis showed that contextualizing the profile for each participant had the desired effect as many participants believed that the profile pertained to them. Participants rated their comfort, surprise, and concerns about the profile's content on a 5-point Likert scale, as shown in Figure 2. The majority was surprised (73% agree or strongly agree), concerned (84%), and not comfortable (82%) with the contained information. We first discuss what concerned and surprised participants, followed by perceived benefits of accessing their profiles.

4.2.1 Concerns and surprise

Participants who indicated surprise or concern could express their reasons in an open-ended question. For concern, the analyzed sample consisted of 157 responses, from which we identified 20 unique concerns. For surprise, the analyzed sample consisted of 138 responses, from which we identified 17 categories. Concern categories are shown in Table 4; surprise categories are shown in Table 5. Due to the large overlap between surprise and concern categories, we discuss them together.

Major reasons for surprise and concern were the amount of information collected by online advertisers (17.9% concern, 35.5% surprise) and the level of detail (7.5% concern, 15.2% surprise). A major concern

Reason	Count	Percent
amount of information	55	17.9%
personal information	45	14.7%
general concern	31	10.1%
level of detail	23	7.5%
lack of consent	21	6.8%
general harm	20	6.5%
inaccurate information	13	4.2%
household (info)	11	3.6%
credit score (info)	11	3.6%
accurate information	11	3.6%
income (info)	9	2.9%
unnecessary for ads	9	2.9%
inferred information	8	2.6%
location (info)	7	2.3%
other	33	10.7%

Table 4: Reasons for concern about the sample profile (307 codes=100%, n=157).

was that the collected information is considered personal (14.7%), many participants also voiced general concern about the profile (10.1%).

Participants (6.8%) mentioned lack of consent to the collection of data as a concern and were concerned about selling and sharing of information with third parties. One participant stated, "Collection of that much information on a person without their knowledge seems dishonest. Also the possibility of that information being used against a person or ending up in the wrong hands is disturbing." Lack of transparency regarding data collection was also a concern. Further concerns included both the inaccuracy (4.2%) and accuracy (3.6%) of information. Particularly inaccurate information was associated with potential harm: "... incorrect information could lead to problems to the person being profiled through no fault of their own."

There were concerns about a number of specific data types, such as household (3.6%) or credit score (3.6%), specially as the amount and detail of collected data was seen as unnecessary or excessive for advertising (2.9%): "The amount of information collected, most of which is very personal and unnecessary for their purposes." For some participants, awareness of the scale of data collection increased their distrust in advertisers, e.g., "... NO corporation has any business collecting this much information ... this is really incredibly upsetting ... my mistrust of internet advertisers has increased tenfold by seeing that. I will make it my goal to block their access to ALL of my information ..."

Surprise about the information in the profile largely aligned with stated concerns. One participant stated, "I was very surprised by just how much info there was; it was like a resume or background check." Participants were most surprised that the profile contained detailed information about an individual's household (5.6%) and financial information, such as credit score (5.6%) and income bracket (3.9%). One participant stated: "How much they made, where they lived, the credit bracket... stuff that I have been told all of my life to keep to yourself, but they have it on a spread sheet."

The level of accuracy was another reason for surprise (5.6%). Many participants felt uncomfortable with accurate information being collected about them, for example, "the profile almost matches me EXACTLY. It was a little terrifying." Participants were also surprised about the potential of creating inferences (2.6%) and data aggregation (2.2%). One participant summarized it as, "it has a lot of information that seems meaningless, but put together creates an alarmingly specific picture." Others were more surprised with the

Reason	Count	Percent
amount of information	82	35.5%
level of detail	35	15.2%
inaccurate information	14	6.1%
household (info)	13	5.6%
credit score (info)	13	5.6%
personal information	13	5.6%
accurate information	13	5.6%
income (info)	9	3.9%
interests (info)	7	3.0%
inferred information	6	2.6%
location (info)	5	2.2%
aggregation	5	2.2%
other	16	6.9%

Table 5: Reasons for surprise about the sample profile (231 codes=100%, n=138).

level of inaccuracy in parts of the profile (6.1%), e.g., "Many items on the profile did not fit my situation or personality. I am surprised at the assumptions made by the ad software based on very few points a reference."

4.2.2 Benefits of profile access

We asked participants to "[t]hink about the ability to view and edit the information that advertising companies know about you. How much do you agree or disagree with the following," showing them six statements. 90% of participants believed (agreed, strongly agreed) that they should be given the opportunity to view and edit their profiles. A large percentage wanted to be able to decide what advertising companies can collect about them (85%) and saw benefits in being able to view (79%) and edit profiles (81%). The majority thought that the ability to edit their profiles would provide companies with more accurate data (70%) and allow them to better serve the participants (64%). These results indicate that the motivation for accessing their own behavioral data is not only to remove information but actually enhance the utility of targeted ads.

After viewing the sample profile, participants rated their level of comfort with sharing different categories of information, see Figure 3. The majority was not comfortable sharing contact information (83% disagreement), which matches the results obtained before they viewed the profile (79% unwilling to share email). After viewing the profile, participants were either comfortable sharing interests (46%) or neutral about it (21%), which is similar to the willingness to share purchasing interests reported before (43% agreement, 19% neutral). However, the comfort level for sharing online activities, such as the articles read, was much lower (18% agreement) than the willingness to share this information reported before viewing the profile (45%). Thus, participants are more willing to share interests than online activities with advertisers, which conflicts with the current practice of primarily collecting activity data.

Participants described the perceived benefits (if any) of having access to their profiles. We analyzed 199 responses, resulting in 259 coded statements categorized in Table 6. While 13.5% saw no benefit in gaining access to their profile or only benefits for the advertiser (3.9%), the main reported benefit was being able to know what information advertising companies were collecting (28.6%). Participants perceived it as "fair" to have access to the information collected about them, e.g., "I think transparency is best. At least I know what information they have gathered about me." Yet, only a small fraction (1.2%) explicitly stated that access to the profiles would increase their trust in the advertising company.

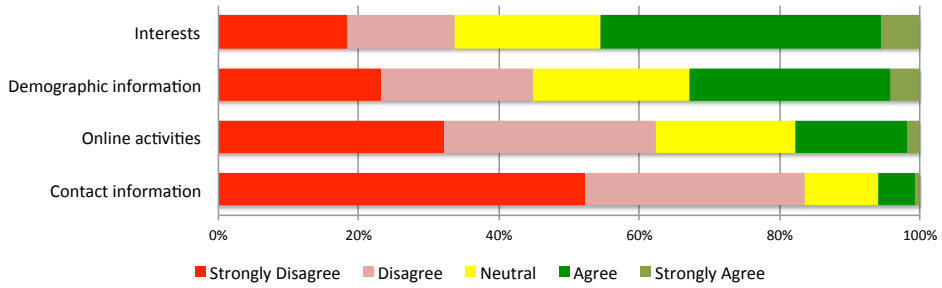

Figure 3: Participants' responses to the statement, "In general, I am comfortable sharing with advertising companies," after they saw the sample profile.

Benefit	**Count**	**Percent**
know what is collected	74	28.6%
make ads more relevant	49	18.9%
no benefit for consumer	35	13.5%
consequences of online behavior	24	9.3%
make profile more accurate	20	7.7%
want to edit or delete info	15	5.8%
benefit for advertiser	10	3.9%
make better privacy decisions	9	3.5%
understand reasons for ads	8	3.1%
make profile less accurate	6	2.3%
other	9	3.5%

Table 6: Perceived benefits of access to own profiles (259 codes=100%, n=199).

Other stated benefits fall into two categories: a way for improving the utility of targeted ads or for enhancing privacy protection. Many participants would like to increase the relevance of ads shown to them (18.9%) and believed that by accessing their profiles they could correct inaccuracies (7.7%), and remove or edit some information (5.8%). A related reason was the interest in understanding why they see specific ads (3.1%), for example, *"to correct any misinformation if a user was interested in having targeted ads and to understand the information that is driving the ads you receive."*

Participants also thought that accessing the profile would help them better understand the consequences of their online behavior (9.3%). Some participants saw that as a motivation for making better privacy decisions or changing their behavior in the future (3.5%), e.g.,"It might persuade more of them to be more careful about what information they give out and avoid websites that abuse their privacy." Only a small fraction of participants (2.3%) wanted to make profiles less accurate by removing personal information.

Participants who agreed that "they should be given the opportunity to view and edit their profile," explained their reasons. The analyzed sample of 173 responses led to 285 annotations across the categories shown in Table 7. The main reasons for wanting to edit the profile were the ability to control what is being collected (22.5%), to protect personal information (16.5%), to make the profile more accurate (16.5%), and to know what is being collected (15.1%). This reflects the perceived benefits of access as well as the major concerns and surprises about the amount of collected information. Other participants were generally interested in editing or deleting information from the profiles (14.0%), some with the goal of making ads more relevant (10.9%).

Reason	Count	Percent
control what is collected	64	22.5%
make profile more accurate	47	16.5%
protect personal information	47	16.5%
know what is collected	43	15.1%
want to edit or delete info	40	14.0%
make ads more relevant	31	10.9%
make profile less accurate	9	3.2%
other reasons	4	1.4%

Table 7: Reasons for wanting edit profiles created by advertisers (285 codes=100%, n=173).

A number of participants thought that their online behavior did not reflect their interests and would prefer to provide or adjust their interests and ad preferences instead. One participant expressed,"it would help both myself and the advertising companies if they knew my preferences from me and not my searches. My searches don't necessarily indicate my interests." Another said, "I think it's important because the articles I read might contribute to a different profile about me than what I actually like. For example, I might read a lot about Republican politics but be a registered Democrat." This may explain why participants indicated to be more comfortable sharing their interests than their online activities, see Figure 3. Some participants further stated that being able to control what information advertisers collect and use about them would increase their willingness to share certain information with them, "there are certain details of my life that I do not want known. I would be more willing to share some information, to allow properly targeted advertising, if I were also allowed to guard other information that I feel is personal and private. This would also help the advertisers because they would have accurate information and I would be more likely to respond to their ads." While a few participants stated that they would purposefully make the information in a profile inaccurate (3.2%), a much larger fraction (16.5%) was willing to correct inaccuracies and they felt that correcting inaccuracies could be beneficial to them as well as advertising companies.

In summary, our participants were concerned and surprised by the amount and level of detail of information that advertisers collect. Yet, the reported reasons for and benefits of having access to their profiles show that this could potentially be addressed by being transparent about data collection and offering control about what data types can be collected and used for advertising.

5 Discussion and Conclusions

Public policy discussions related to OBA have focused on the need for privacy notices and opt-out opportunities [8]. The advertising industry has developed icons that indicate the use of targeted advertising on websites, and has provided tools and websites to enable consumers to opt-out of targeted advertising from individual or a list of advertising companies. However, these are "all or nothing" solutions that require users to make decisions without a full understanding of advertising companies' practices. We investigated how practices routinely used by advertisers influence users' willingness to disclose different types of information. We discuss how our results could assist advertising companies in refining their practices to better meet people's expectations, and how to improve regulation in this area.

5.1 Context Matters

More than half (55%) the participants were not comfortable sharing any type of information. However, those who were comfortable sharing some data showed nuanced preferences. Participants were comfortable sharing information that they deemed necessary for advertising or delivery of a better service. For example, 49–53% of participants were comfortable sharing their online interactions when data was used for targeted ads on the visited website as well as other websites, but not if the data was used for other purposes outside the scope of the visited website, or combined with PII (e.g., Facebook and offline scenarios). They were also more comfortable sharing their email addresses to receive coupons (13%) and with Facebook (16%) than with other websites or for other purposes (7–9%). Participant's were comfortable sharing their gender, ZIP code and other data types under some circumstances, but not always. These results suggest that binary approaches like Do-Not-Track or opt-outs, which do not consider the context of collection, as well as ad companies' current disclosure practices are less than optimal.

Disclosure preferences were generally influenced by perceived benefits and potential harms, as well as the perceived trustworthiness of the parties involved in collecting or processing the information. Generally, participants were more accepting of their browsing activities being used for advertising on the primary website, but were less comfortable when other parties had access, especially if those parties or their purposes were unspecified. Thus, more restrictive use of behavioral data and clear specification of purposes would likely result in higher acceptance of targeted advertising.

In contrast to related work, we did not find that retention had a significant effect on participants' willingness to share. Whereas Leon et al. [13] compared one-day or indefinite retention and found significant differences. Our bounded and less extreme retention periods (one week, three months, one year) had no effect. This difference in results suggests that people dislike indefinite retention but that implications of limited retention periods are likely too abstract to impact disclosure preferences. Further investigation is required to determine how the privacy implications of retention periods of different lengths can be communicated effectively to users.

5.2 Meeting Users' Expectations

Participants were more comfortable sharing some data types when they assumed that the "other purposes" for which companies would use that data were positive. However, many participants tended to assume the worst. This suggests that collection and use purposes should be clearly specified to ensure that the users' expectations and consent align with the actual data practices rather than having (potentially erroneous) assumptions drive their willingness to share or their perceptions of advertising practices.

Consistent with findings from previous studies [24], participants indicated that they were uncomfortable sharing data with companies that had not obtained proper consent to collect and use their data. Further, even users who recognized benefits in sharing remained uncomfortable sharing sensitive data types such as income bracket, credit score bracket, sexual orientation, email address and IP address. They considered them not necessary for advertising, and potentially harmful. Participants were surprised and concerned about the types and amount of information shown in the sample profile.

Many participants did not see the need to collect so much data for advertising purposes. Participants also voiced concerns about sharing information with third parties. Furthermore, many participants felt that their online activities did not really reflect their purchasing preferences. Yet, participants were willing to share information that they thought would result in improved and more relevant ads, such as location when they were interested in learning about local deals, or gender or operating system, which were considered not or less personal. Many participants (46%) were also comfortable sharing their purchasing interests directly.

These insights suggest that there is opportunity for interest-based rather than behaviorally targeted advertising. Advertisers could—and likely should—provide their potential customers with options to refine and specify there preferences and respect their expectations. Making it more transparent what information a displayed ad is based on and enabling users to influence and adjust these aspects would likely result in fewer violations of privacy expectations and an increase in perceived utility of ads as well as generally more positive attitudes towards advertising practices.

5.3 Access is Needed

A large fraction (90%) of participants wanted to view and edit their profiles for various reasons, including awareness and control over the collection of their data, to make their profile more accurate, to remove unwanted information, and to receive more relevant ads. Indeed, 70% indicated that they would improve the accuracy of their profiles to improve the ads shown to them, while only a small fraction indicated that they would purposefully provide false information.

These results suggest that providing access to users could increase both the effectiveness of targeted advertising and users' comfort and trust. This contrasts with today's practice of making inferences about users' preferences with the risk of making mistakes and infringing on their privacy. Ad companies could limit inferences to non-sensitive information such as demographic or non-sensitive interests and allow users to opt in to receive ads tailored to more sensitive categories (e.g., according to income range, sexual orientation, or health interests).

A handful of advertising companies are already giving users access to anonymous (i.e., cookie based) profiles, such as BlueKai Registry, but it is unclear whether users are aware of those access options, and the vast majority of advertising companies don't offer any access. Furthermore, the current access tools are often difficult to find and difficult to use and understand [12, 20]. Advertisers should strive to create usable tools that make it transparent what and how information about an individual is collected and used, and enable individuals to easily refine these profiles so that they reflect their interests and the online self-image they are comfortable with.

5.4 Improving Privacy Notices

Understanding the factors that are relevant to users is important for designing notices that communicate OBA practices in an effective and actionable manner. Our results show that users are interested in knowing not just what data is being collected, but also how it will be used and shared, and how those uses may benefit or harm them. However, it is not reasonable to expect users to locate and read the privacy policy of every advertising and tracking company to determine whether their practices are acceptable. Therefore we need tools that can automatically identify the trackers that match users' preferences. Current efforts are attempting to extract important privacy policy elements through crowd sourcing and natural language processing [21, 25]. Success of such efforts, however, relies on companies being transparent about their data practices, and adhering to their stated collection, use and sharing practices.

References

[1] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass: re-examining user concerns for online tracking and advertising. In *Proc. SOUPS '13*. ACM, 2013.

[2] Howard Beales. The value of behavioral targeting. *Network Advertising Initiative*, 2010.

[3] Michael Benisch, Patrick Gage Kelley, Norman Sadeh, and Lorrie Faith Cranor. Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing*, 15(7):679–694, 2011.

[4] BlueKai. Little blue book: A buyers guide. White paper, BlueKai, 2013.

[5] BlueKai. The BlueKai Registry. `http://bluekai.com/registry/`, 2014.

[6] Federal Trade Commission. Data brokers: A call for transparency and accountability, May 2014.

[7] Ayman Farahat and Michael C. Bailey. How Effective is Targeted Advertising? In *Proc. WWW '12*. ACM, 2012.

[8] Federal Trade Commission. Protecting consumer privacy in an era of rapid change, March 2012.

[9] Joshua Gomez, Travis Pinnick, and Ashkan Soltani. KnowPrivacy: Final Report. Technical report, KnowPrivacy.org, June 2009. `http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf`.

[10] Adam N Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B Paine Schofield. Privacy, trust, and self-disclosure online. *Human–Computer Interaction*, 25(1):1–24, 2010.

[11] Anja Lambrecht and Catherine Tucker. When Does Retargeting Work? Information Specificity in Online Advertising. *Journal of Marketing Research*, 5:561–576, 2013.

[12] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proc. CHI '12*, 2012.

[13] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What matters to users?: factors that affect users' willingness to share information with online advertisers. In *Proc. SOUPS '13*. ACM, 2013.

[14] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I Hong. Modeling usersÕ mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Porc. SOUPS '14*, 2014.

[15] Naresh K Malhotra, Sung S Kim, and James Agarwal. Internet users' information privacy concerns (iuipc): the construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[16] Aleecia McDonald and Lorrie Cranor. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *Proc. TPRC '10*, 2010.

[17] Jonathan Mugan, Tarun Sharma, and Norman Sadeh. Understandable Learning of Privacy Preferences Through Default Personas and Suggestions. Tech report CMU-ISR-11-112, School of Computer Science, Carnegie Mellon University, 2011.

[18] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 2010.

[19] Lee Rainie, Sarah Kiesler, Ruogu Kang, and Mary Madden. Anonymity, privacy, and security online. Technical report, Pew Research Center, 2013. `http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online/`.

[20] Ashwini Rao, Florian Schaub, and Norman Sadeh. What do they know about me? Contents and concerns of online behavioral profiles. In *Proc. ASE Conference on Privacy, Security, Risk and Trust (PASSAT)*, 2014.

[21] Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Aleecia M. McDonald, Joel R. Reidenberg, Noah A. Smith, Fei Liu, N. Cameron Russell, Florian Schaub, and Shomir Wilson. The usable privacy policy project: Combining crowdsourcing, machine learning and natural language processing to semi-automatically answer those privacy questions users care about. Tech. report CMU-ISR-13-119, Carnegie Mellon University, December 2013.

[22] Horst Treiblmaier and Irene Pollach. Users' perceptions of benefits and costs of personalization. In *Proc. International Conference on Information Systems (ICIS)*, 2007.

[23] Catherine Tucker. Social Advertising. SSRN, Massachusetts Institute of Technology, 2012. http://ssrn.com/abstract=1975897.

[24] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proc. SOUPS '12*. ACM, 2012.

[25] Sebastian Zimmeck and Steven M. Bellovin. Privee: An architecture for automatically analyzing web privacy policies. In *Proc. USENIX Security '14*, 2014.

A AllNews Website

Figure 4: The All News homepage. We asked participants to visit this webpage before providing them notice about [Best Ads / Facebook]'s OBA practices.

B Survey Scenario

Note: We only show scenario 6 as example, other scenarios have different practices.

Please read this information carefully. Then answer the questions below.

Many websites, including AllNews, are able to offer free services to their visitors by contracting with online advertising companies. The advertising companies pay websites for every ad they show, allowing the websites to provide free services to users.

Imagine that you provided some information about yourself (e.g., email address, gender, etc.) when you signed up for an account with the AllNews website. Further imagine that AllNews has contracted with Best Ads, an advertising company that is interested in learning about you to show you ads that are most likely to be of interest to you. These ads are known as targeted ads.

For example, if you watch a video about the 2014 winter Olympic games on the AllNews website and then visit a traveling website and look up hotels near the Olympic venue, next time you visit the AllNews or any other news, entertainment, travel, or retail website, Best Ads could show you a targeted ad for a discounted hotel near the Olympic venue.

The following table summarizes Best Ads' data collection and use practices.

Best Ads may collect information from	• The AllNews website • Other news, entertainment, travel, and retail websites you visit
Best Ads may use information it collects to show you targeted ads on	• The AllNews website • Other news, entertainment, travel, and retail websites you visit
Best Ads may use information it collects for	• Targeted ads • Other purposes
Best Ads may retain information for	• **[One week / 3 months / One Year]**

C Sample Profile

Advertising companies create individual profiles based on the information they collect or infer from users' online activities. Some of these companies provide Internet users access to their profiles. The table below shows an example of what information such a user profile may include. The information has been taken from actual user profiles created by an advertising company. Please review this sample profile carefully and then answer the questions below.

Data Type	Value
Location	Region:[Participant's Region]
	City:[Participant's City]
	IP Address:[Participant's Computer IP Address]
Demographics-Individual	Gender:[Participant's Gender]
	Single
	[Participant's Age] years old
	Education: [Participant's Education]
	Type of Job:[Participant's Occupation]
Demographics-Household	Income: $50K - $75K
	Household size:1
	Number of Adults:1
	Children in Residence: No
	Home Type: Multifamily Dwelling
	Home Value: Less than $100K
	Length of Residence: Fewer than 3 years
	Discretionary spending: $30K-$40K
	Voter Indicator: Republican
	Automobile: Less than $20K
Interest	General Health>Bones, Joints, Muscles>Pain
	Religion code: Tiers 1 - 3
	Video Games: Computer, PlayStation 3
	Travel Destinations>North America>US>New York>NYC
	Miscellaneous>News>Business and Finance
	Automobile:Coupes
	Online Activities: Research
Activities	Past Purchase>Products>Clothing>Jeans
	Offline Purchases>P&G>Charmin Ultra Soft
	Student Loan Consolidation
	Volunteering: Tier 1 - 3
Attitudes	Buy American: Not likely
	Look at Me Now: Most likely
	Never Show Up Empty Handed: Most likely
	It's all in the Name: Most likely
Behavior	Green Living
	Eco Friendly Vehicle Owner
	Mass Market and Discount Shopper
	Gift buyer
	Prepaid wireless plan subscriber
	Premium channel viewer
Predictive	Credit card interest score: 16-17%
	Credit card appl. intent score: 10 -11%
	Auto insurance online buyer: High propensity
	Online Higher Education Enrollee: High propensity
	In-market: Cell phones and plans

D Regression Model

Table 8: The logistic regression models of participants' willingness to disclose information. In addition to the scenario treatment, we included the following co-variates: age, gender, whether or not a participant used Facebook, Internet literacy, privacy concerns, whether participants like targeted ads, opinion of the AllNews website, and whether or not a participant answered at least one of the scenario understanding questions correctly. Only variables significant at α<0.05 are shown. If one or more levels of a categorical variable was significant, we show all the levels of that categorical variable.

Independent Variable	Control Category	Coefficient	Std. Error	Z	P>|Z|
Dependent Variable: Articles Read, Videos Watched, and Pages Visited (Online Interactions)					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	-0.16	0.2	-0.8	0.42
Scope 3: AllNews + Others	Scope 1: Only AllNews	-0.18	0.18	-0.97	0.33
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	-0.58	0.19	-3.06	0.002
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.62	0.21	-2.87	0.004
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.55	0.20	-2.80	0.005
Scope 7: AllNews + FB	Scope 1: Only AllNews	-0.85	0.19	-4.5	<0.001
Purpose: Positive	Purpose: Negative	0.32	0.17	1.84	0.07
Purpose: Ambiguous	Purpose: Negative	0.17	0.24	0.71	0.48
Privacy Concerned: Yes	No	-0.62	0.14	-4.43	<0.001
Targeted Ads Opinion: Positive	Negative	1.54	0.15	10.12	<0.001
Opinion on AllNews	Negative	0.91	0.10	8.85	<0.001
Has FB Account: Yes	No	0.33	0.13	2.61	0.009
Dependent Variable: Purchasing Interests					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.07	0.20	0.36	0.72
Scope 3: AllNews + Others	Scope 1: Only AllNews	-0.014	0.19	-0.08	0.94
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	-0.06	0.19	-0.29	0.77
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.33	0.22	-1.51	0.13
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.44	0.20	-2.21	0.03
Scope 7: AllNews + FB	Scope 1: Only AllNews	-0.35	0.19	-1.83	0.07
Purpose: Positive	Purpose: Negative	0.22	0.18	1.22	0.22
Purpose: Ambiguous	Purpose: Negative	0.37	0.24	1.55	0.13
Privacy Concerned: Yes	No	-0.43	0.14	-3.06	0.002
Targeted Ads Opinion: Positive	Negative	1.87	0.17	11.28	<0.001
AllNews Opinion: Positive	Negative	0.71	0.10	6.79	<0.001
Dependent Variable: Gender					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.21	0.20	1.05	0.29
Scope 3: AllNews + Others	Scope 1: Only AllNews	-0.009	0.19	-0.05	0.96
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	0.24	0.19	1.28	0.20
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.05	0.21	-0.25	0.80
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.34	0.20	-1.70	0.09
Scope 7: AllNews + FB	Scope 1: Only AllNews	0.33	0.18	1.78	0.08
Purpose: Positive	Purpose: Negative	0.32	0.17	1.82	0.07
Purpose: Ambiguous	Purpose: Negative	0.32	0.24	1.32	0.19
Privacy Concerned: Yes	No	-0.42	0.14	-3.11	0.002

Continued on next page . . .

Continued from previous page

Independent Variable	Control Category	Coefficient	Std. Error	Z	P>|Z|
Targeted Ads Opinion: Positive	Negative	1.1	0.14	7.93	<0.001
AllNews Opinion: Positive	Negative	0.72	0.10	6.99	<0.001
Has FB Account: Yes	No	0.57	0.13	4.33	<0.001
Age	NA	-0.01	0.004	-2.82	0.004
Gender: Male	Female	0.29	0.10	2.79	0.005
Dependent Variable: ZIP code					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.06	0.22	0.29	0.77
Scope 3: AllNews + Others	Scope 1: Only AllNews	-0.05	0.21	-0.24	0.81
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	-0.25	0.21	-1.20	0.23
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.40	0.24	-1.64	0.10
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.64	0.23	-2.82	0.005
Scope 7: AllNews + FB	Scope 1: Only AllNews	-0.13	0.21	-0.63	0.53
Purpose: Positive	Purpose: Negative	0.46	0.20	2.35	0.02
Purpose: Ambiguous	Purpose: Negative	0.39	0.27	1.47	0.14
Privacy Concerned: Yes	No	-0.74	0.14	-5.25	<0.001
Targeted Ads Opinion: Positive	Negative	1.17	0.14	8.18	<0.001
AllNews Opinion: Positive	Negative	0.60	0.12	5.1	<0.001
Age	NA	0.03	0.004	5.52	<0.001
Gender: Male	Female	0.26	0.11	2.28	0.02
Dependent Variable: Sexual Orientation					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.30	0.27	1.08	0.28
Scope 3: AllNews + Others	Scope 1: Only AllNews	0.25	0.25	1.01	0.31
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	0.37	0.25	1.46	0.14
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	0.56	0.28	2.01	0.04
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.07	0.28	-0.25	0.80
Scope 7: AllNews + FB	Scope 1: Only AllNews	0.80	0.24	3.35	<0.001
Purpose: Positive	Purpose: Negative	0.21	0.22	0.96	0.34
Purpose: Ambiguous	Purpose: Negative	0.39	0.30	1.30	0.19
Privacy Concerned: Yes	No	-0.79	0.15	-5.29	<0.001
Targeted Ads Opinion: Positive	Negative	1.0	0.15	6.82	<0.001
Has FB Account: Yes	No	0.29	0.17	1.71	0.09
AllNews Opinion: Positive	Negative	0.58	0.13	4.36	<0.001
Dependent Variable: Email					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.37	0.34	1.11	0.27
Scope 3: AllNews + Others	Scope 1: Only AllNews	-0.10	0.34	-0.30	0.77
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	0.56	0.30	1.84	0.07
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.19	0.40	-0.47	0.64
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	0.22	0.34	0.65	0.51
Scope 7: AllNews + FB	Scope 1: Only AllNews	0.88	0.29	3.04	0.002
Purpose: Positive	Purpose: Negative	0.10	0.28	0.34	0.74
Purpose: Ambiguous	Purpose: Negative	-0.49	0.50	-0.98	0.33
Privacy Concerned: Yes	No	-0.58	0.19	-3.05	0.002
Targeted Ads Opinion: Positive	Negative	0.98	0.18	5.45	<0.001
Has FB Account: Yes	No	1.02	0.28	3.64	<0.001
AllNews Opinion: Positive	Negative	0.62	0.18	3.53	<0.001