Do Not Embarrass: Re-Examining User Concerns for Online Tracking and Advertising

Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, Saurabh Panjwani
Bell Labs Research, Bangalore, India
{saurabh.panjwani, sharad.jaiswal, nisheeth.shrivastava}@alcatel-lucent.com, agarwal.lalit91@gmail.com

ABSTRACT

Recent studies have highlighted user concerns with respect to third-party tracking and online behavioral advertising (OBA) and the need for better consumer choice mechanisms to address these phenomena. We re-investigate the question of perceptions of third-party tracking while situating it in the larger context of how online ads, in general, are perceived by users. Via in-depth interviews with 53 Web users in India, we find that although concerns for third-party tracking and OBA remain noticeable amongst this population, other aspects of online advertising—like the possibility of being shown ads with embarrassing and suggestive content—are voiced as greater concerns than the concern of being tracked. Current-day blocking tools are insufficient to redress the situation: users demand selective filtering of ad content (as opposed to blocking out all ads) and are not satisfied with mechanisms that only control tracking and OBA. We conclude with design recommendations for end-user tools to control online ad consumption keeping in mind the concerns brought forth by our study.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Human Factors

Keywords

online advertising, third-party tracking, privacy, embarrassment

1. INTRODUCTION

Advertising on the Internet is a much more complex phenomenon than in traditional broadcast media. Modern Web technologies have made it possible for advertisers to track individual users' online habits and browsing patterns, tailor ad content to match these patterns and consequently increase both the relevance of ads for users and revenues for themselves. Tailored advertising of this form, often referred to as *online behavioral advertising (OBA)*, is cleverly orchestrated by a host of intermediaries (or third parties) who liaise between publishers and advertisers and help accomplish two key goals: the tracking and aggregation of user data for ad tailoring and the eventual delivery of tailored ad content from the advertisers on the publishers' sites. OBA, and personalized ads in general, are being increasingly recognized as the way forward in Web advertising and already constitute a noticeable chunk of the global online ad market [11].

Numerous researchers and privacy advocates have raised concerns with respect to OBA because of its reliance on third-party tracking of users' data. It has been shown, time and again, that users lack awareness of the mechanics of OBA and that current consumer choice mechanisms (for controlling tracking) and education efforts (to raise awareness) have had limited impact in helping users exercise choice with respect to OBA [23]. Studies also indicate that although some users find OBA useful, the general public attitude towards it is negative and that people are deeply concerned about their online activities being tracked by Web third parties [13, 17, 21].

We re-investigate the question of user attitudes towards third-party tracking and online advertising, studying it in the context of 53 Web users in Bangalore, India. Our method is similar to that of Ur *et al.* [23] in that we use one-on-one, in-depth interviews to understand user attitudes towards OBA, but with three key differences. First, we take stronger measures to avoid priming our study participants for privacy when educating them about concepts around OBA. Second, we initiate the study of user sensitivity towards third-party tracking in a *quantitative* manner, measuring sensitivity as a function of individual browsing histories of users. Finally, and most importantly, we study the question of perceptions towards OBA by situating it in the context of how online ads, in general, are perceived by users. This enables us to evaluate concerns for OBA relative to other user concerns in the realm of online advertisements.

Below are our key findings:

- We find that users in our study, like those in previous studies, are concerned about third-party tracking and OBA but their overall attitude is more neutral here than in earlier works. Users' concerns are centered largely around a fear of personally-identifiable information and financial data being lost to third parties and most users want only a fraction (about 25%, on

Symposium on Usable Privacy and Security (SOUPS) 2013, July 24–26, 2013, Newcastle, UK.

average) of their browsing history to be not tracked by third parties. Their concerns towards being *shown* OBA ads are even milder than those towards being tracked, which is not surprising given that OBA is only one discernible outcome of tracking.

- More than the issue of third-party tracking, users are concerned and sensitive about the content they get exposed to in online ads. A majority of the users in our study reported past experiences of being shown ads with embarrassing and suggestive content which had upset them. Others reported being wary of shown embarrassing ads in the future. Users explained, with vivid examples, how the context surrounding their Web browsing behavior (e.g., browsing at home vs. work) can lead to varying levels of embarrassment. Some users, in fact, invoked the ideas of third-party tracking and OBA to explain why consumption of certain ads may lead to greater embarrassment in front of others. Overall, users ranked the possibility of being shown embarrassing ads as a significantly greater concern than that of being shown OBA ads or even that of being tracked by third parties for OBA.

Besides these key findings, our study uncovers new issues in online advertising and user perceptions of it, which seem to have been left unaddressed in previous work on the topic. We find that users have clear and unique topical preferences for and against advertising content and these preferences are *not* necessarily reflected in the browsing behavior of the individual: ads users want to see may not be related to the sites they have recently visited. Furthermore, within the scope of relevant ads based on browsing behavior (i.e., OBA ads), the *perceived* relevance of an ad seems to depend critically on its timing—badly-timed and repetitive OBA ads could easily lead to user dissatisfaction.

Our study culminates in new ideas of tools for addressing user sensitivities towards online advertising. We find that existing tools to control third-party tracking or ad blocking do not address all the concerns raised by the users in our study and there is a need for new technology to fill this gap. We begin the exploration of such technology in this work and outline critical problems of usability and deployability associated with this exercise.

2. RELATED WORK

User perceptions of online ads. Numerous researchers have explored consumer perceptions of Web advertisements and studied how different aspects of online ads like content and interactivity affect these perceptions. A 2007 study by McCoy *et al.* [12] finds that ads can diminish user preference for websites and that certain kinds of intrusive ads (namely, pop-ups and pop-unders), in particular, hamper users' ability to retain the content of websites. Other studies have explored the effects of animation in ads and found that although highly animated ads are generally detrimental to user experience [3], a moderate amount of animation can also produce positive effects, like increased ad recognition rates and brand attitudes [24]. Campbell and Wright studied the interplay between ad repetitiveness, relevance and interactivity, and showed that increasing relevance and interactivity can significantly improve user attitudes towards repetitive ads [4]. Put together, these studies suggest that users generally perceive ads as an annoyance, but careful design choices and increasing relevance of ad content can sometimes change this perception.

While ad annoyance has been the subject of much research in the past, we find very little work on the issue of embarrassment induced by online ads on sensitive topics. One recent study explores ad-induced embarrassment in the context of TV ads and demonstrates that the social context surrounding ad viewing can determine both feelings of embarrassment and advertising effectiveness [16]. Embarrassment, as investigated in our paper (and in [16]), is not strictly a privacy issue but we find it to be intricately tied with the privacy-affecting phenomenon of personalized advertising. In particular, we find that embarrassing ads tend to cause greater concern to users than third-party browser-history tracking, and that knowledge of the latter and its effects can heighten user concerns for embarrassment.

Perceptions of OBA. Before we describe related work on OBA, we give a brief primer on this topic. Behavioral advertising or OBA is a modern technique in online advertising which is used by advertisers to tailor ad content to users based on their past browsing habits. It is not the most popular form of tailoring today (e.g., *contextual* advertising which involves tailoring ads based on the ad publisher's content has a much higher incidence rate [10]) but its popularity is growing and it is increasingly being discussed in both industrial and academic circles because of its privacy implications. OBA is normally implemented by Web third parties who partner with different websites, track information about individual visits to these sites and use this information to create browsing profiles of the visitors. The more websites a third party partners with, the more information it gathers about the sites' visitors and the better profiles it can create of individual users. This information is subsequently used by third parties to channel advertisements to users based on the profiles it created for them. The most popular third party which does this kind of "cross-site" tracking today is DoubleClick [18], a fully-owned subsidiary of Google.

Normally, third-party tracking used in behavioral advertising is meant to be anonymous from the user's perspective. Third parties usually rely on cookies to identify individual users' machines and these cookies are intended only to collect information about the different websites a user visits, without linking these visits with personally-identifiable data. In response to concerns around potentially violational data collection practices, industry self-regulatory bodies like the Network Advertising Initiative (NAI) and Digital Advertising Alliance (DAA) have recently taken shape. Part of the objective of these bodies is to enforce ethical data collection in OBA (e.g., ensuring anonymity of user data) and to enable appropriate choice mechanisms for users (e.g., the ability to opt out from being shown OBA ads) [6, 14].

Even with the assumption of anonymity, OBA, and tracking in general, raises concerns around individual Web behavior being observed by third parties, and the extent to which this is desired by users. Numerous surveys and studies have addressed this question since the inception of OBA [8, 13, 21, 22, 23]. Initial work by Turrow *et al.* and McDonald *et al.* [13, 22] and the 2011 TRUSTe and Harris Interactive online survey [21] paint a negative picture, with statistics as high as 85% for the number of users who are opposed to third-party tracking and OBA. All of this work was conducted in the context of American users. The more recent

and more international survey by KPMG [8], on the other hand, suggests that nearly two-thirds of Web users are willing to be tracked by third parties, provided this happens "under the right circumstances" and "in exchange for cheaper content". Finally, the study by Ur *et al.* [23] conducts a more in-depth investigation of user attitudes towards OBA (in the US) using in-person interviews and demonstrates that user preferences for tracking are highly complex and while more people are worried about it than not, the extent of the worry depends upon the situation in which tracking happens.

Our work adds to this literature in three different ways. Like Ur *et al.*, we use interviews to understand user attitudes towards third-party tracking and OBA but study these attitudes in the context of Indian users. (We also take greater care to avoid privacy-priming our study participants, as discussed later.) Second, we take the first steps in *measuring* user sensitivity towards third-party tracking via a unique clustering apparatus that is applied to real browsing histories of users. Finally, we study user attitudes towards tracking and OBA within the larger framework of user perceptions of ads, and find that although tracking does concern users, other issues in online advertising (like the possibility of being shown embarrassing ads) concern them even more.

Consumer Control Mechanisms. There is a growing body of work to support end-users in controlling how their data is tracked for OBA and in customizing ad consumption in general. Amongst the industry-led efforts, the most popular ones are blocking tools like AdBlock Plus and Ghostery[1], which assist users in blocking out individual third-party trackers and, in the case of the former, also blocking out ads. AdBlock Plus is currently the most popular of all browser plugins but its primary purpose is ad blocking not tracking control. Much recognition is being given to Do-Not-Track (DNT), a W3C-led mechanism for enabling users to seamlessly signal tracking preferences to third parties via a new HTTP header, but its implementation seems challenging, given consistent signs of non-compliance from advertisers [1]. The academic literature has also looked at OBA control mechanisms (e.g., [20]), but largely using more privacy-preserving (client-side) approaches to profile users while still maintaining ad relevance. These tools significantly transform the workflow of third-party based ad delivery, either by introducing new intermediaries into the picture or by making ad targeting more cumbersome for third parties, which limits their overall deployability. We believe there is a need to re-think the problem of designing client-side mechanisms for ad and tracking control, while balancing both user concerns and deployability, and discuss a potential solution for this towards the end of the paper.

3. METHODOLOGY

Our study consisted of two parts—a qualitative part which involved in-depth interviews with users on the topic of online advertising and OBA, and a quantitative part aimed at measuring user sensitivity towards third-party tracking.

3.1 The Interview

Most of our interaction with users was via a one-on-one, semi-structured interview centered around online advertising and OBA. We started with the expectation that users are generally unaware of or carry misconceptions about third-party tracking (based on evidence from [23]), so we incorporated an educational component into our interviews. As we report later, our findings confirmed this expectation.

User Education. To educate users about OBA and third-party tracking, we initially considered the possibility of using a canned video on this subject, on the lines of [23]. Our search for finding a suitable online video for our task resulted in disappointment. Most videos we found online were either privacy-priming in nature (projected a bias that third-party tracking violates privacy), too long (incorporated redundant concepts), or narrowly scoped (e.g., a marketing video from a third party explaining how only that party, and not others, collects user data). We found several of these limitations reflected in the video used by [23]: the video is more than 7 minutes long, carries redundant information (like how and by whom cookies were invented), mentions the word "privacy" in several places and makes misleading suggestions about how tracking can be suppressed (e.g., it suggests that deleting cookies is a control for cookie-based tracking).

Therefore, we prepared our own educational materials on third-party tracking and OBA. Our materials consisted of a PowerPoint slide deck (which explains third-party tracking and OBA through an example and gives data about the prevalence of these practices on the Web) and a script to be spoken alongwith. We contextualized our example to suit the Indian audience (e.g., we used "Cleartrip.com", a popular Indian travel website, as the advertiser and "Google" as the third party) and took care to emphasize the difference between tracking and behavioral advertising (the latter being just one consequence of the former). We also emphasized the fact that cookies are anonymous (contain only a random identifier to identify a user's browser and no personally-identifiable information), something which we found inadequately expressed in online videos on the subject. Our exposition was entirely around cookie-based, cross-site, browser-history tracking, which is the most prominent tracking form used for OBA [11, 18]. We did not mention sophisticated techniques like flash cookies [19] in order to avoid overwhelming our users. We also included a brief tutorial on how to control behavioral advertising, which came later during the discussions around user perceptions of OBA.

We kept our presentation of materials interactive, asked questions as we spoke and encouraged users to ask questions in return. This not only enabled us to gauge users' comprehension of the materials but also elicited data about their instinctive reactions to OBA. Our training session lasted between 3 to 7 minutes depending upon the amount of interaction injected by the participants. While training participants, we did not make any statements to the effect that third-party tracking impacts individual privacy.

Questionnaire. Our interview script consisted of six parts. In Part 1, we collected user demographic data and gauged users' prior understanding of third-party tracking and OBA. This was followed by user education, and then by Part 2, which assessed user perceptions of third-party tracking. Part 3 examined users' overall perception of ads and their topical preferences. This part used an ad categorization inventory shown to participants on a computer screen. The inventory was derived using Google Adsense's top-level "general" ad

[1] http://www.adblockplus.org, http://www.ghostery.com/

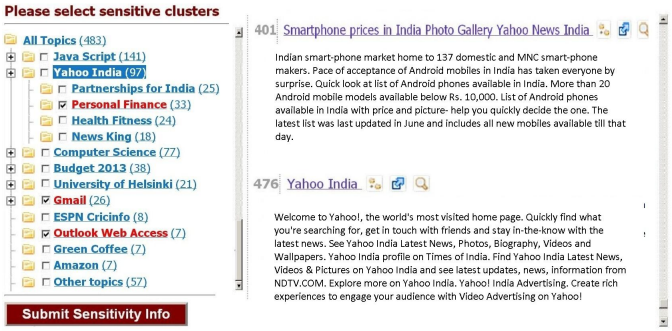

Figure 1: User interface for the clustering apparatus used during the study. Labels of clusters appear on the left in an expandable-list format. Clusters can potentially contain sub-clusters. By clicking on a cluster, the user can view its contents—URL titles and page snippets—on the right. There are checkboxes next to each cluster label which the user can use to identify sensitive clusters (colored red). Some fonts have been enlarged for legibility.

categories[2] merged, randomly, with some sensitive ad categories (e.g., sex-related, get-rich-quick ads) also defined by Adsense. Parts 4 and 5 investigated user sensitivity towards advertising content, which included questions around ad-induced embarrassment and perceptions of OBA. This was followed by the tutorial on tools for OBA control. Part 6 studied user expectations for ad and tracking control.

3.2 Measuring Sensitivity

A unique feature of our study, compared to prior usable privacy work on OBA, was that we collected quantitative data on user sensitivity towards third-party tracking. Our goal was to compute, across users, the fraction of the users' browsing history that a user regards as "sensitive" i.e., not being appropriate for being tracked by third parties. In order to simplify the process of identifying sensitive URLs for the user, we used a clustering technique to pre-process the browsing data of the user. Our data collection setup had two components: a client-side browser extension (written for Chrome and Firefox) that would be installed on a user's machine and a server-side clustering engine run on a remote server controlled by us. The extension extracts the last 1,000 URLs in the user's history and transmits them to the server along with a unique user ID. The server crawls these URLs, extracts text from them, and clusters them based on word similarities using a commercial clustering tool called Lingo3G[3]. The server then transmits the cluster information to the client, which displays the clusters in an expandable-list format, as shown in figure 1. The user can inspect the contents of all clusters (URL titles and snippets are displayed on the right side of the screen) and mark a subset as sensitive using a checkbox interface. Clusters can potentially contain sub-clusters, which can be individually marked sensitive. Users provide two sets of inputs—one with respect to arbitrary third parties, and another with respect to a "trusted" third party, picked by them from a list of seven major third parties shown by us on a computer screen.

Because this part of the study relied on individual browsing histories, which are generally viewed as sensitive by users, we gave participants the choice to *not* participate in it or to participate with limited engagement i.e., submit history data to the server temporarily till the point we compute sensitivity ratios. For the participants who opted for the latter, data was removed from the server after the study.

3.3 Participant Sample

We recruited 53 participants to take part in our study. Most (44) of the participants were recruited from the office premises of Alcatel-Lucent (ALU) India—a leading telecommunications company in India—via a combination of email fliers, announcements and in-person solicitations. Printed versions of the fliers were also placed in numerous locations of ALU's premises in Bangalore which is used by more than 2000 people everyday. We chose to advertise our study in ALU for convenience: our research lab is employed by ALU, and we were interested in conducting face-to-face interviews only. To add variety, we also recruited 9 participants from outside ALU using personal contacts, while ensuring not to enroll any close kin as a way to minimize response biases. These participants added some demographic richness to our sample and included people from another IT company, employees of an NGO, housewives, college students (two participants from each of these categories) and one employee of a travel agency. The non-ALU participants were interviewed closer to the end of the study, and we limited ourselves to nine because we seemed to be reaching a point of diminishing returns in terms of lessons learnt per participant.

Our sample was gender-balanced (26 F, 27 M) and we took care to sample a mix of people from both technical (bachelor's degree in engineering) and non-technical backgrounds (25 technical, 28 non-technical). The age range was 22 to 42, with a mean age of 30.7. Participants were fairly well-educated, all of them holding a bachelor's degree and 20 with at least a master's degree, and came from middle and upper-middle class Indian families (daily per capita income of over 20 USD). They were all active Web users, reporting to be spending between 1 to 8 hours (mean 3.4 hours) browsing on a personal PC everyday. At least 30% of the participants reported that they possessed a smartphone, but PC-based browsing was more common. Google Chrome was reported to be the favored browser by more than half of the participants, followed by Firefox and Internet Explorer.

Our study shares some limitations with other interview-based studies like limited geography, limited sample size and consequent limits on the generalizability of results. However, the depth with which we were able to collect data from individual users—particularly on the issue of embarrassment—would have been difficult, or entirely impossible, using surveys or empirical studies. Besides, situating the study in India gave us the opportunity to understand user perceptions on a widely-discussed privacy issue which has been explored only in the context of Western cultures till now.

3.4 Study Protocol

Most interviews were conducted in a closed-room laboratory environment, except in a few cases (of the 9 participants outside of Alcatel-Lucent) where interviews took place at

[2]https://support.google.com/adsense/bin/answer.py?hl=en&answer=186376&topic=23398&ctx=topic

[3]Lingo3G is available from http://carrotsearch.com/lingo3g-overview.html. In prior work, we have experimented with different approaches to cluster search queries and collected user feedback on these experiments as well. We have also applied the tool for measuring sensitivity of search queries [15].

home. The first and last author jointly moderated 13 of the interviews (conducted earlier in the study), and the remaining interviews were conducted by each of them individually following the same script. (The jointly-conducted interviews served as practice to reach a point of moderator consistency.) Interviews typically lasted between half an hour and 50 minutes, although a few participants volunteered more than an hour of their time. All interviews were audio-recorded. Each participant provided written, informed consent and received a verbal note of thanks at the end of the interview.

Participants were requested to bring their personal laptops, where available, and at the end of Part 2 of the interview (tracking-related questions), they were told about the optional quantitative component of our study. If the participant opted in, we installed our browser extension on his/her machine at this point. The clustering engine ran in the background (taking up to 15 minutes), while we continued with the interview. At the end of the interview, we instructed participants about the process of marking sensitive clusters on our clustering UI (using a mock-up) and let them provide their inputs as we stepped away and observed them distantly. (We kept distance in order to avoid making them feel uncomfortable or embarrassed.) We emphasized the definition of "sensitive" multiple times to each participant: a cluster was meant to be marked "sensitive" if the participant did not desire third parties to track him/her on URLs that might fall under the cluster[4]. Participants were given discretion in deciding the extent to which individual clusters were examined. At the end of the exercise, we uninstalled our extension from the participant's machine.

3.5 Analysis

Data collected through the interviews was recorded and analysed by the two researchers who moderated the interviews. One of them transcribed the interview audios as they were generated; the other read and verified the transcribed text in parallel. Since the same researchers were involved in collecting, recording and verifying the data, it was possible to identify prominent themes in the data even before formal analysis began; e.g., the link between ad-induced embarrassment and OBA was discovered by us at this stage.

Our analysis used a bottom-up inductive coding technique, commonly used in qualitative research. We built an initial codebook based on pre-determined themes, the structured questions in our questionnaire and some emergent themes noted during the transcription process. Data was stored in a set of 5 matrices—one "main" matrix M_1 (capturing most of the participant responses, hierarchically categorized); 3 focused matrices M_2-M_4 (dedicated to 3 topics which generated particularly long response vectors: tracking topical preferences, ad topical preferences and general attitude towards ads, respectively); and one "quant" matrix M_5 (which stored data from the quantitative part of the study). M_5 included some data on tracking preferences (from M_1, M_2) which helped study correlations. Most of our data codes were binary variables corresponding to structured questions (e.g., *"did the participant report to have experienced OBA ads?"*) and some unstructured ones (e.g., *"did the participant report to want to view personalized ads?"*); some were numeric variables with a broader range (e.g., *"participant's concern level for third-party tracking (1-5)"*, *"in which context did the participant experience ad embarrassment? (1 = alone, 2 = with children, 3 = with elders, ...")*) and a few were plain text (*"what reason did the participant provide for undesirable ads being undesirable?"*).

Starting with the initial codebook, we iteratively expanded the codes, based on emergent themes (participant responses not captured by existing codes) and by compiling plain-text codes into numeric variables. After introducing or updating a code, previous entries in the matrices were revised to fit the updated coding scheme. We also eliminated some codes which we felt were not exposing interesting themes.

The task of coding and data entry was split between the two researchers as follows. Researcher A, the *lead* researcher, performed most of the coding and data entry and researcher B assisted in parts of the process. The initial codebook for M_1 was built collaboratively by both of them; in particular, researcher A made a codebook definition, collected feedback from B and updated the definition based on this feedback. The quantitative data in this matrix was entered by both A and B, with B making about half of the entries; the remaining data was entered by A only. Researcher A identified emergent themes, updated code definitions based on these themes, collected feedback from B on the new definitions and entered the corresponding data into M_1. Researcher B later verified that data entry for the emergent themes (done by A) complied with definitions developed earlier. Matrices M_2, M_3 and M_5 (all of which contained quantitative responses only) were filled by B. M_4, which stored data on general attitude towards ads, was filled by A.

Although most of the coding and analysis activity was undertaken by the lead researcher, we took measures to minimize the chances of researcher bias influencing our results. First, researcher B took part in developing the initial codebook (as described above) and also gave feedback on the emergent themes identified and coded by A. (For example, B gave feedback on whether definitions of individual codes were exhaustive or not, based on his own perusal of the transcripts.) Second, researcher B verified entries in M_1 which corresponded to emergent themes; in particular, B identified discrepancies between a matrix entry, a code definition and a transcript, as applicable. Third, the two researchers who did not participate in formal analysis, did read the majority of the transcripts and verified that the results reported here are not evidently deviant from the transcripts' content. This verification was not formal but it still helped us increase our confidence in the validity of our results and test whether the results are free from individual biases or not.

4. FINDINGS

Even though nearly half the participants in our sample were from a technical background and were employed in technical jobs, overall awareness about cookies and third-party tracking was abysmally low. All but two participants had heard about cookies prior to the study, but only three could correctly explain their functionality (even discounting the possibility of third-party cookies). We noted different misconceptions about cookies like "cookies save my browser history," "they store my recently used passwords," and even "a cookie is a small software embedded by a vendor which spies on my data". (Similar misconceptions have been reported by McDonald *et al.* [13], though awareness levels

[4]Note that the definition of the word "sensitive" was provided after part 2 of the study (discussion on third-party tracking) was over. As such, it is unlikely to have had priming effects on participant responses to third-party tracking.

seem lower in our study.) Participant preparedness for third-party tracking and behavioral advertising was even poorer, most not even having heard of these terms. Some participants confused tracking for "hacking", with one participant referring to third-party tracking as a way for "unwanted applications to steal our banking details."

During our presentation on third-party tracking and OBA, we tried to eliminate these misconceptions and noted some indicators to this effect. Participants expressed surprise at the idea of a cookie being a text file and that of third parties serving content on other websites. Some were curious about observing real cookies, in which case we demonstrated the process of viewing cookies on the browser of their preference, on their own machines where available. Participants asked advanced questions about third-party tracking and OBA (e.g., "Is there a way to know that third parties are tracking me on a website?" or "Do they track just the fact that I visited [a website] or anything else?" or "[Do] you mean, against one cookie, there are several websites that Google knows [I am visiting]?"), which further indicated their comprehension of our content. (We responded with a fixed set of answers to such questions, clarifying that third-party tracking for OBA is largely anonymous, unless unethically implemented.) Finally, when asked questions about the benefits of OBA, participants were almost universally able to articulate the idea that OBA enables them to see relevant ads and indirectly helps advertisers as well.

More than forty out of the fifty-three participants in our sample reported to have experienced OBA in the past. Some interrupted us during the presentation and pointed out that our example (of Cleartrip ads shown on a news website) reminded them of their past experience of seeing OBA ads on Facebook and other sites. A few were visibly pleased to have learnt the internals of a phenomenon previously observed by them: "Oh! This is how it works?"

4.1 Concerns about Tracking

Unlike past studies, we isolated participants' perceptions for third-party tracking and OBA, studying their reactions to both these concepts separately. As in prior research [13, 23], participants exhibited a range of reactions to third-party tracking but their general outlook seems to be less negative here. We believe that our approach to user education, which limited privacy priming, had an effect on this outcome although this is difficult to prove using our data alone.

At a high level, our participants found third-party tracking to be a useful idea but held concerns regarding (a) the lack of transparency with which it is implemented and (b) the extent to which their browsing history is tracked. Roughly as many participants expressed a positive attitude towards third-party tracking as a negative one (25% positive, 28% negative, the rest neutral)[5]. At the positive extreme, there were five participants who believed there was no online activity of theirs which they did not want to be tracked (provided it was done anonymously). Participant F-11 (F being short for female) was particularly supportive of the idea:

> Why wouldn't I want them to track me? Can't think of a reason. Even if I'm doing anything hanky panky, big deal, if I'm doing it.. What's the harm if they [know]? That this particular ID.. XYZ.. is fond of music, or is fond of movies, is fond of reading news, sports. If someone is able to capitalize on that information and pass me information that I might be interested in, why not?

However, most participants portrayed a more neutral view on third-party tracking. The most common reaction we received from participants to the idea was: "It is ok as long as my personal information is not tracked." Several participants were divided about their opinion on third-party tracking because of questions around transparency and choice: "If they can track what I am browsing, maybe they can keep tab of the information that I give out on those sites?" (F-16), "It doesn't concern me so much [but] they should not track me unless I have given [them] permission" (F-23). Finally, two participants were completely opposed to the idea of third-party tracking, not wanting to be tracked on any website they visited.

One concern that emerged from our interviews (and is unreported in past OBA studies) was that third-party tracking could lead to marketing calls or emails from advertisers, which could become a source of annoyance. Six participants explicitly mentioned that they fear advertisers could gain information about their *phone numbers* through this method: "In these [shopping] sites, they take my mobile number. I am concerned that then I [might] get those marketing calls.. [My] only concern is that they take and publish this information [to marketers]" (M-11).

4.1.1 Context Dependence

In general, participants reported that their preferences for third-party tracking depends on the context in which it happens. The issue of context dependence was studied in [23] using six hypothetical scenarios given to users and gathering their preferences for permitting tracking in these scenarios. We took a deeper look at this issue in our work.

First, we posed two open-ended questions to participants: *"Can you give examples of websites on which you would not want third parties to track you? Examples of websites on which you would want them to track you?"* For the former, the pre-dominant answers we received were: online banking, email, social networking sites (in that order)[6]. For the latter, the pre-dominant answer was e-commerce websites. Besides email, banking and SNSs, there were a few types of websites for which participants reported not being comfortable tracked e.g., job search ($n = 4$), insurance ($n = 4$), politics ($n = 4$) and gossip ($n = 3$). One participant (M-15) stated that his preference depends on the *age* of the website: "The website should be older.. like 5 years old. The websites which are new, I don't want to be tracked [there]."

Second, along the lines of [23], we gave participants a list of 7 website topics and asked: *"For which of these topics would you not want any third party to track you, when you browse a website on the topic?"* The topics were *financial investments, job search, critical illnesses, travel, adult content, gossip* and *news*[7]. Our hope was that even though participants may not have been able to recall their sensitiv-

[5]Our finding of dominant privacy-neutrality amongst users is reminiscent of Ackerman *et al.*'s partitioning of users into three groups—privacy fundamentalists (17%), privacy pragmatists (56%) and the marginally concerned (27%) [2].

[6]Note that this expectation of participants does not conflict with current tracking practices on popular email and SNS sites; the latter largely rely on *first*-party tracking.

[7]These topics capture the scenarios used in [23], with minor modifications e.g., we included *adult content* and *gossip* and we excluded *online food shopping* (not common in India).

ity towards certain topics from the earlier question, showing them candidate topics would help uncover some sensitivities. A majority of the participants reported not being comfortable to be tracked on financial investments (59%) and adult content (51%) websites. A smaller fraction of the sample was concerned with respect to critical illnesses (32%) and job search (25%), too[8].

Finally, we explored participant sensitivity towards third-party tracking quantitatively by applying the clustering apparatus described in Sect. 3.2 on their browsing histories. Twenty-seven participants (51%) took part in this part of the study; of the remaining, ten were unwilling to participate and the rest either did not use Chrome/Firefox or had recently deleted their history. The key findings from this part of the study were:

High sensitivity variance. Participants varied widely in their sensitivity preferences, ranging from finding nothing in their history sensitive ($n = 2$) to 100% sensitivity ($n = 1$). Participants viewed an average of 65.3 clusters (including sub-clusters) on their browsing history, out of which an average of 17.6 were marked sensitive by them. The average sensitity ratio—ratio of the number of sensitive clusters to the total number of clusters—we computed was 0.257 and the variance was extremely high (SD = 0.233).[9]

Quantitative responses more inclusive. Sensitivity preferences largely reflected the qualitative responses given earlier by participants, but tended to be slightly more inclusive. Fourteen participants (52%) found some topic in their history sensitive which was not voiced as sensitive in response to the qualitative questions. The most commonly reported topics of this variety carried the labels "videos" or "movies" ($n = 6$). Besides these two labels, there were up to 15 topics other than emails, banking and SNS'es which were reported as sensitive but were distributed sparsely across participants e.g., jobs/careers ($n = 2$), blogs ($n = 2$), photos ($n = 2$) and one case of Bollywood gossip being marked sensitive.

Third-party-indifference. Even though almost all participants expressed a bias towards a few third parties during the interview (Google being a common favorite), we observed hardly any difference between participant sensitivities towards "all" third parties versus a "trusted" third party. This suggests that even though participants may instinctively express a preference towards some third parties, differences in trust levels across parties may actually be small.

In sum, we learnt that there are a few topics (namely, email, banking/financials, social networking, adult content) which are commonly regarded as sensitive by most users from the perspective of third-party tracking but besides these topics, users are divergent about the topics on which they want to control or allow tracking. Also, it seems that even though participants may project a neutral attitude towards third-party tracking "on the whole", they carry (highly varying) situational biases with respect to it.

Feedback on History Clustering. While investigating participant sensitivity towards third-party tracking, we also probed them about their perceptions of our clustering apparatus. Although there were some reports of badly-formed clusters and of inappropriate or non-descriptive cluster labels, most participants provided positive feedback for our clustering tool and expressed that they would like to see such a functionality integrated with their regular browsing experience (independently of tracking controls). Some were visibly excited to see how clustering provided a quick gist of their browsing history: "Wow! I have been visiting only shopping websites" (F-19). Others liked the idea of clustering because it could potentially simplify searching over their browsing history. In the words of F-13:

> Nowadays I am visiting Bharat Matrimonial, so in that ways, daily 10-20 boys will be there while I am browsing their sites. If I want to go through somebody's profile, I can go to Bharat Matrimony [cluster] and find out the person directly.

One participant went further to suggest that a clustered view of browsing history could help users keep track of how others could be using their machines: "I give my laptop to my daughter also; it will be interesting [for me] to know what she is browsing" (F-22). In ongoing work, we are exploring alternate techniques (e.g., use of fixed topic directories) to categorize browsing histories so as to improve accuracy and usability over our current system. While there is much work on classification of Web documents, we are unaware of studies on classification of individual browsing histories and corresponding usability evaluations. This topic would be interesting to explore in future work.

4.1.2 Third-Party Tracking Vs. OBA

Participants' attitudes towards being *shown* OBA ads were, in general, more positive than those towards third-party tracking: when asked if they were concerned about being shown OBA ads, fewer than 20% of the participants answered positively and more than a third said they would like to be shown such ads. One concern that was voiced several times (by more than 70% of all participants) was with respect to the repetitiveness of OBA ads. Participants reported being annoyed by OBA ads shown to them repeatedly, sometimes even long after they had made a purchase through the ad. Some went a step further and articulated suggestions to suppress repetitive OBA ads; e.g., M-8:

> There should be a time after which I can flush it.. [that is,] I can say, 'Now, change my [profile].' Because if I have [already] purchased a hard disk and then everyday it is showing me that buy-a-hard-disk ad.. [it] is irritating.

Interestingly, 15 participants in our sample (28%) perceived OBA ads to be *more* repetitive than non-OBA ads. It would be interesting to empirically evaluate whether this is just a perception or whether ad networks enforce greater repetitiveness in OBA ads than in other ads.

[8]As a comparison, user attitudes towards being tracked on job search were more negative in the study of [23], participants of that study being "evenly divided" for this topic.

[9]Separately, we collected 5-point Likert-scale ratings from all participants on their concern for being tracked. For the participants who used our clustering tool, we observed a mild positive correlation ($r_s = 0.24$) between these ratings and sensitivity ratios. In particular, the participants who found nothing sensitive in their histories provided a concern rating of 2.5 each, and the one who found everything sensitive provided a rating of 5.

Ten participants (19%) expressed concern that OBA ads could potentially leak information about their past browsing behavior to other individuals in their vicinity, but these concerns were largely speculative in nature; e.g., M-7:

> You never know. There might be something like a search or a job hunt, and then suddenly [an ad] shows up related to that. You don't want your manager to see that.

One participant expressed being more concerned about being shown OBA ads than about third-party tracking itself on the grounds of *unfairness*: "They're sort of taking advantage of your previous browsing [history], and trying to later on kind-of tempt you to do things that you otherwise may not have wanted to do. I don't see that as fair."

4.2 Perception of Ads

One of the main goals of our study was to understand user attitudes towards OBA, relative to other concerns they may have with respect to online ads. For this, we first queried participants on their general perception of ads. Participants exhibited mixed reactions towards ads, ranging from expressions of serious contempt ("I have never found any ad useful") to profuse liking ("I love ads, whether on TV or online"). In response to the open-ended question—*"What comes to your mind when you hear the term online advertising?"*—53% of the participants had something negative to say including words like "annoying", "distracting", "irrelevant" and even "fraud", while 32% of them (not all disjoint) gave positive responses like "informative", "entertaining", "offers" and "smart". A majority of our participants (89% of all females, 55% of all males) reported to be paying visual attention to ads, a sizeable number reported having clicked on them intentionally (67% F, 37% M) and a noticeable fraction reported having converted a click into a purchase (17% F, 11% M)[10]. Participants were largely (89%) unwilling to pay to get rid of ads, indicating that even with some of the negative influences of ads, their presence was not hindering enough to induce a switch to paid browsing.

We found high variability in user preferences for ad categories. Three participants reported to be wanting to view online ads on all categories and each of the rest expressed a unique combination of positive, negative and neutral preference for different topics. The most wanted category across all participants was "travel and tourism" with 66% of the participants expressing a desire to see ads of this category and only one asking to be not shown such ads. The topics reported as being least desired were "Sex-related ads" (60%), "Get-rich-quick ads" (55%) and "Religion" (43%), well-aligned with Google Adsense's definition of sensitive ad categories. We observed strong gender differences on some categories e.g., a 7:1 female-to-male ratio for not wanting business and industrial ads.

A key lesson we learnt through our discussions around online ads was that participants did not necessarily view ad relevance to be tied to their recent browsing history. When asked whether or not they want to see ads on their preferred categories only when they browse for information on that topic, twenty-four participants (45%) answered negatively. Participants expressed an inherent bias towards some ad categories: "I like apparel ads; I may not buy it online but I generally like to see ads of this type" (F-1); "I like cars. I want to be shown ads on any vehicles." (M-12). Clearly, recent browsing patterns are not always a reliable indicator of user preferences for or against ad topics, which suggests that non-behavioral, and even non-contextual, forms of advertising still provide value to users.

4.3 Concerns about Embarrassing Ads

From the general discussion on online ads, our conversation steered towards understanding participant sensitivities for advertising content. We were surprised to find that a majority of our participants (39/53, or 74%) reported to have experienced situations in which they were shown online ads they perceived as containing embarrassing content. Nearly all of these participants also admitted to having been embarrassed by the ad in such situations. In fact, 9 participants (17%) of our participants uttered the word "embarrassing" or a synonym *pro-actively* (to describe some categories of undesirable ads), even before our prompting them about it.

What embarrasses users about ads? Those in our study were fairly consistent in defining embarrassing ads as graphic ads that either contain sexually explicit content, information on online dating or else a display of swimwear or lingerie. There were a few outlying cases in which religious, beauty and personal care, maternity-related and matrimonial ads were also referred to as embarrassing. Four participants reported to have experienced embarrassing ads in the past but declined to provide details of their experiences.

Participants were very vocal about their vexation with embarrassing ads. While some stated having been embarrassed by ads when browsing in private, the majority reported instances in which the embarrassment was caused by being in the vicinity of other people: "At home, if I am alone, I will just be annoyed [by seeing such an ad]. It is more embarrassing when you are in a public setup" (F-12); "I feel embarrassed when something like this comes up when I am doing [a] presentation. It is only in the group or when you have people around" (M-15). Participants varied widely in their descriptions of the context in which ads had embarrassed them in the past. Although there were more reports of embarrassment in the workplace, nearly half of the participants narrated incidents around home browsing. F-3 provided a representative example:

> Sometimes, [even] when you log into good sites, you [see] certain things which really embarrasses. They show Sunny Leone, you know. It doesn't matter when adults are watching, but it does matter [otherwise]. Once, me and my niece were trying to browse something for her. She was on holidays so I wanted to give her some math questions and stuff like that... Something [like that] came up and you know, she didn't know how to react, I didn't know how to react. I just closed it but that was, like, pretty embarrassing. As far as the TV goes, we have locked "those" channels. On the Internet we have to be extra careful.

Equally palpable were participant concerns with respect to being shown embarrassing ads in the presence of parents or other elders, nine participants narrating incidents of this type; e.g., M-12:

[10]The gender differences observed by us seem consistent with prior work on gender effects on ad click-through-rates e.g., http://www.reuters.com/article/2011/08/30/idUS125944+30-Aug-2011+BW20110830

> When I watch online movies, the movies open in a pop-up player of a third party. They don't care about what kind of people are watching. They show ads of scantily-clad women. I was watching it with my mom [once] and it became pretty embarrassing.

A plausible cause for such experiences being so pronounced at home was the high incidence of collaborative browsing in the home environment. In the words of M-3:

> I don't care about it on my personal computer at work but on the home computer which I share with my kids, I don't want ads which kids are not supposed to see.

We probed participants about the nature of websites on which they observe sensitive ads, and the associated frequency. Seven of the participants (18%) reported to be seeing such ads only on websites with pirated content or otherwise interpreted by them as "suspicious" sites. In contrast, more than 60% claimed to have seen embarrassing ads even on "normal" websites which they categorized under email, news or video-streaming sites. Participants who mentioned the first of these categories (email) also reported to have had multiple experiences of embarrassment, often intermittently. e.g., M-14:

> The best [example] is Yahoo [Mail]. It constantly shows available girls less than 25 in your area. I don't have a clue how it can do that, it is extremely embarrassing at work.

As for the 14 participants who did not report experiences of embarrassment in the past, nearly all expressed concern towards sex-related ads, when asked about ad categories that might embarrass them. For example, F-26 spoke: "I would not want to expose [my daughter] to any adult ad. That's the only ad I would not want."

4.3.1 Amplification by OBA

Knowledge about personalized advertising seemed to be amplifying participant concerns with respect to embarrassing ads. Seven of the participants who had experienced embarrassing ads in our sample (18%) considered them embarrassing because they could influence others' perceptions of their browsing tendencies, even if, wrongly so. Two out of these seven made explicit mention of OBA in how such perceptions could be created; e.g., F-23:

> I assume that other people may also know what this [OBA] thing is, and may assume that the reason I'm being shown these ads is because of my previous history. There's nothing wrong with it being my previous history but I certainly don't want other people to [see].

The remaining five pointed to people's potential awareness about personalized advertising in general and how this could influence perceptions. M-17 explained this idea cogently:

> I think what is embarrassing is that the ad reveals what kind of person you might be. The other person who would be with me when this ad came up would interpret more into that. Everybody has knowledge that your ads are being customized to your taste. I would obviously not like [this interpretation].

Even though behavioral advertising constitutes a small fraction of all forms of online advertising [11], it seems plausible that awareness about its existence is increasing (e.g., in our own study, most participants seemed to have experienced its effects a priori), which would make users' concerns for such interpretations quite realistic. Also, note that interpretations like the above are unique to the context of online ads (as opposed to other forms of advertising); neither is personalization possible in other forms nor are ordinary people likely to be able to conceive such a possibility on their own.

Besides the issue of OBA influencing external perceptions, participants expressed concerns about being *wrongly* targeted with embarrassing ads. Three participants hypothesized that the reason they had been shown embarrassing ads in the past had something to do with their browsing behavior being misinterpreted by advertisers. One participant (F-15) explicitly mentioned "health-related searches" as having caused this misinterpretation. Similar concerns were voiced by one participant in the study by Ur *et al.* [23].

4.3.2 Third-person References

Not only did participants talk about their own experiences with embarrassing ads, they volunteered other people's experiences into their responses. Participant F-24 narrated an interesting anecdote from her workplace.

> It happened with a colleague of mine. The other day he opened IE and something had happened with his settings, it went straight to MSN.com and there was one bikini-clad lady standing on the website, the main page. Everyone [around him] was like—what were you browsing last night? And he was like—nothing. Poor guy; it was really embarrassing.

The casual reference to "browsing last night" in the referred conversation suggests how the idea of ads being determined by prior online activities has been internalized by some users. Another participant (not one who had experienced embarrassment from ads) described an episode in which a friend and his father encountered an ad from "certain embarrassing sites" when browsing at home and later told her about it. The participant expressed concern about facing similar situations herself in the future.

Such third-person references further accentuate participants' concern for being shown embarrassing ads. The concern seems significant enough that users talk about it (at least in some circles) and are concerned by others having experienced them in their lives.

4.3.3 Perceptions relative to OBA

Amongst the participants who had experienced embarrassing ads, their concern towards being shown such ads was markedly greater than that towards being shown OBA ads, or even towards third-party tracking. Nearly all participants in this category stated to be more worried about being shown embarrassing ads than about tracking or OBA. We asked participants to provide a qualitative concern rating on a 5-point Likert scale to three eventualities—being tracked by third parties, being shown OBA ads and being shown embarrassing ads. The results are shown in figure 2.

Lack of prior awareness of OBA and third-party tracking could have biased participant response on these questions. However, it is difficult to determine the direction of bias. It is plausible that being newly-educated about a concept

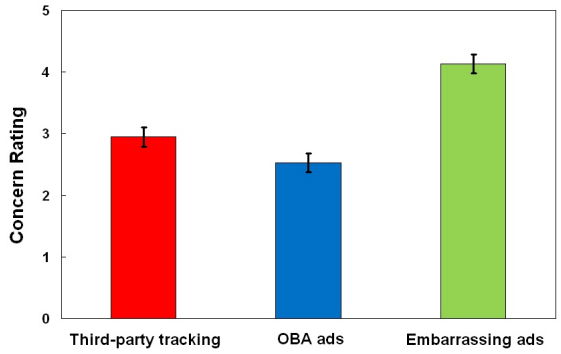

Figure 2: Mean participant concern ratings for being tracked by third parties ($\mu = 2.95, n = 53$), being shown OBA ads ($\mu = 2.53, n = 53$) and being shown embarrassing ads ($\mu = 4.14, n = 39$). Error bars indicate standard error. Concern ratings for embarrassing ads are significantly greater than those for tracking, by the Wilcoxon signed-rank test ($z = 4.11, p < 0.001$). The latter are significantly greater than ratings for OBA ads ($z = 2.39, p < 0.05$).

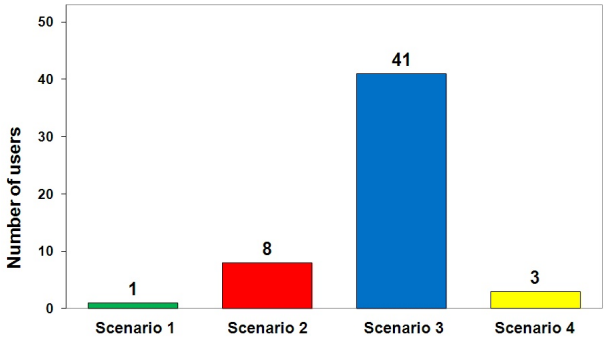

Figure 3: Participant preferences for ad blocking. Scenario 3, the most preferred scenario, is the one in which ads of some user-defined topics are shown, the rest being blocked. The next preferred scenario (Scenario 2) is one in which all non-annoying ads are shown. Even less preferred is Scenario 4 (all ads blocked).

like third-party tracking, which involves indiscernible data collection from users, could have biased them against it, but this would only imply an even smaller relative concern for tracking and OBA. Contrarily, not having experienced any harmful effects of OBA in the past may have softened their stance on OBA (and, consequently, on tracking). Nevertheless, given the descriptiveness and depth in participant responses on the topic of embarrassing ads, it appears that ad-induced embarrassment is a concern of general importance amongst Web users and one that is at least as worthy (if not more) of further investigation as OBA is.

4.4 Perception of Ad Blocking Tools

In our discussion on controls for behavioral advertising, we informed participants about the use of blocking tools as a mechanism to limit exposure to advertising as well as third-party tracking. (With suitable tuning, blocking tools can be an effective defense against tracking as well [11].) We centered our discussion around AdBlock Plus, the most widely used ad blocking tool, explaining its high-level functionality to participants, including its ability to restrict third-party tracking, and gauging their subsequent reactions.

By and large, participants responded favorably to the idea of AdBlock Plus (ABP) but a few were also displeased by the sheerness with which it restricted ad consumption. Only three participants in our sample reported to have used the tool prior to the study. Amongst the remaining, the majority (62%) expressed an interest in using the tool, some even making follow-up requests to help install the tool on their machines. Two participants were explicit in stating that their main motivation to use the tool was to eliminate embarrassing ads. Four others wanted to use it only in some situations (e.g., only at work). Participants questioned us on the tool's ability to entertain individual preferences and of the participants who did not want to use the tool, more than half explained this choice with a desire to see ads filtered based on personal choices. Our conversation with F-24 is illustrative of this line of thinking:

> **Interviewer (I):** Would you like to use ABP?
> **Participant (P):** Can I choose the websites where I want it to work?
> **I:** When you're on a website you can specify whether you want ads on that site or not.
> **P:** But I cannot choose by category of ads?
> **I:** No.
> **P:** When I'm saying category, can I choose ads from certain service providers? Like for example, I want travel ads, is that possible to say? If I can specify what are the kind of ads I want to see or which are the service providers I'm ok to see ads from, then it would help. If it completely blocks out everything, then not.

To understand participants' desire for such category-based control in ad filtering, we presented four scenarios to them:

1. All ads are shown to you as is
2. All ads except certain annoying ads (pop-ups, pop-unders and distracting ads) are shown to you
3. Ads on some topics of your choice are shown to you, the rest are blocked
4. All ads are blocked

An overwhelming majority of participants, including those who earlier sided with AdBlock Plus, chose to be in the third scenario (figure 3). It is worth pointing out that current-day blocking tools are largely focused on achieving either scenario 4 or 2, which seem to be less preferred by users.

5. DISCUSSION

While much emphasis has been given to the issue of privacy in behavioral advertising in prior work, our study suggests that this may not be the issue that Web users are most worried about today (within the realm of online advertising). A large number of users in our study reported being more concerned about seeing embarrassing advertisements online than about their browsing history being tracked by third parties, which means that at least in some geographies, embarrassment from online ads is a matter of significant user

concern. Still, we find that this matter has not received adequate attention from the research community in the broader discussion around online advertising.

This is not to suggest that the issue of third-party tracking be ignored. Even in our study, like in others', participants were emphatic about issues of transparency and choice in behavioral advertising and varied significantly in terms of their willingness to being tracked in different contexts and on different topics. A few also raised concerns around OBA ads leaking private information to proximal users. The problem of improving user controls for third-party tracking and OBA is important, but studying it *in isolation*, ignoring other user concerns towards online ads, is what we call into question.

5.1 The Need for Selective Ad Blocking

One might suggest that selecting appropriate ad content (e.g., eliminating embarrassing ads) should be the responsibility of publishers and ad networks: with time, users would naturally gravitate towards the more responsible publishers (with better ad selection policies), and the problem of ad inappropriateness would simply disappear. This has not happened thus far. For example, users in our study reported to have encountered embarrassing ads even on certain "good" websites (e.g., email sites) and claimed that they continue to visit such sites for their information needs. Others admitted seeing bad ads only on "bad" websites (e.g., sites which promote piracy), but carried no intentions of abandoning such sites. Besides, user preferences for advertising content seem complex and highly individualistic: some users want OBA ads; others don't. Some are embarrassed by ads; others are not. User preferences for and against ads also seems context-dependent e.g., one participant in our study reported a desire for child-friendly ads in the home environment but not at work; another asked for blocking OBA ads on shared computers but not on others.

One might also suggest that current-day ad blocking tools are meeting users' needs but this, as we discussed in Sect. 4.4, is not the case. More than half of the participants in our study who were disinterested in AdBlock Plus said that they lacked interest not because they did not want to see ads blocked, but because they wanted to stop irrelevant and embarrassing ads only, something that ABP still does not have good controls for. Besides, prior work has reported usability flaws with ABP and other blocking tools which make these tools difficult to use by novice Internet users [9].

Finally, tools like ABP have a singular focus on ad blocking and control third-party tracking as a secondary objective. We believe that controlling third-party tracking should be a key design criteria for an ad blocking tool, given the findings from recent studies, including ours. But again, a singular focus on controlling tracking (as done by a variety of other tools [9]) also seems insufficient.

5.2 Towards an End-User Tool

Our proposition is simple. We envision an end-user browser-assisting tool which enables users to achieve two objectives—*selective tracking control* i.e., disabling tracking on websites which fall under certain user-defined sensitive topics and *selective ad blocking* i.e., blocking out advertisements which contain information on certain, potentially different, user-defined topics. Selective (topic-based) tracking control is an idea that has been discussed in prior work [5]; selective ad blocking, on the other hand, is something that seems unexplored as yet. As such, we don't know of any system which gives both these functionalities together to end users.

Figuring out the right implementation of such a tool will take some iterations and is left open for future work. Below, we list down some of the design guidelines that we hope to follow in the process.

- **Categorization approach:** A critical design decision for a tool like this is the approach to take for categorizing websites and ads. Our suggestion is to start with standard topic directories (e.g., Google ad categories), develop a classification system around them and test the system for real browsing and ad histories. We expect that building a comprehensive and usable category listing for browsing histories will be challenging and that standard directories will need a lot of tuning. Existing approaches [5] have relied on very coarse directories and rudimentary classification mechanisms, whose usability has not yet been evaluated.

- **User input:** User input will consist either of a set of configuration parameters (e.g., categories marked as sensitive by the user) or dynamically-specified attributes (e.g., ads defined as embarrassing during browsing) or a combination of the two. Our discussions with users during the study indicate that dynamic input might be the more preferred modality, although users also seem to be comfortable viewing and navigating hierarchical category lists with more than 50 items (as required by our clustering apparatus). Pre-marking some categories as sensitive based on common preferences (as found in Sec. 4.1.1) may simplify user input.

- **Filtering:** The actual filtering can happen using a combination of approaches like setting DNT headers, blocking cookies, javascripts or ad loading, publisher opt-out settings, etc. Existing tools should be suitably leveraged, based on performance-accuracy trade-offs. It may be worthwhile to introduce limits on ad blocking in order to protect publisher interests.

- **Performance constraints:** A key constraint to keep in mind is that users are sensitive to seeing their browsing experience (e.g., browser performance) degrade with add-ons and extensions. Out of the three participants in our study who reported to have used ABP in the past, two reported to have stopped using it recently. For one, the reason was a perceived browser slowdown; for the other, the reason was restricted functionality on some websites (because of ABP's control on pop-ups and javascripts). While we may be able to circumvent the latter problem to some extent (our hope is to be more inclusive with advertisers and third parties than ABP), we reckon that improving over ABP in terms of speed will prove challenging. Plausibly, some forms of efficiency loss (e.g., a slowdown in ad delivery only, as opposed to that in loading entire pages) will be tolerable for most users.

5.3 Deployment Challenges

Even though AdBlock Plus is the most widely-used browser plugin today, it does not seem to have been installed by more than 5% of all Web users[11]. Its actual usage is probably

[11] http://www.getadzap.com/blog/how-many-people-use-ad-blockers/

much less than this. Tools that are focused only on tracking control (not ad-blocking) seem to be used by a significantly tinier fraction of the Web[12]. Before we develop a new system for tracking control and ad blocking, it is worthwhile to reflect a bit on these statistics and try to reason them out.

Why are tracking prevention tools not popular yet? We believe that the reason for this is deeper than just the relative novelty of these tools or the associated teething problems (e.g., flawed usability [9]). While there is active debate around privacy issues in third-party tracking in the US and beyond, and even recent cases of third-party violations and retributions [7], to the best of our knowledge, there is no real, demonstrated evidence of actual "harm" caused to an individual by a third party's tracking activities, anonymous or otherwise. Neither are there widely known examples of users experiencing differentiated services due to irresponsible tracking practices, or a published report, as yet, on tracking data compromised from third-party storage. To the extent that such evidence remains scarce, the fear about the potential damages caused by third-party tracking will be limited in the mind of the average user which will, in turn, reduce the incentive to seek tracking control or to even learn about it. This may also explain why users in our study felt less concerned about third-party tracking and OBA than the phenomenon of embarrassing ads, even though the latter portends no threat to personal data.

An important area of future work is to build mechanisms that can attract users to use tools that control third-party tracking. Portraying real privacy leaks on user data, or questionable examples of tailored ads in the user's browsing history could be one incentive. Until that is shown to be possible, other alternatives are worth exploring. For one, combining tracking control functionality with ad blocking functionality (as proposed in the preceding section) is likely to help: ad blocking is clearly more desired by users and a tool that functionally improves ABP to match some of the non-ABP-users' requirements (e.g., selective ad filtering) without worsening its usability or destroying advertisers' interests, can hope to be at least as popular as ABP. Another approach could be to add "bonus" features which are not directly related to tracking or ad blocking but satisfy other information needs of users. An example of this would be visualization and search features, implemented on top of history categorization, which could be used to improve users' history navigation experience (as suggested in Sec. 4.1). Finally, one important aspect of deploying end-user tools is user education. Tracking control mechanisms can bring value only to the users who are aware about tracking and lack of awareness can be a dampener for deployment. Awareness is likely to spread as real privacy breaches surface, but until then, other approaches will be needed.

5.4 Limitations of the Study

Our study evaluated user perceptions with respect to third-party tracking for a specific form of third-party tracking which is commonly used in online advertising. Other forms of tracking also exist (e.g., non-anonymous third-party tracking using social widgets [18]) but we did not consider them for several reasons: one, to maintain a focus on user concerns towards advertising; two, for practicality—most participants in our study were unaware about any form of tracking and we wanted to minimize training effort.

As we mention above, user concerns around privacy are influenced by exposure to privacy breaches, or the after-effects of tracking-based profiling. We did not present any such effects, real or potential, that could be caused by third-party tracking, to participants. Going forward, it is important to understand the extent to which privacy breaches can arise in real-world third-party tracking and to study user responses to such breaches when effected on their profiles.

Our quantitative evaluation of user sensitivity, though first of its kind, is quite preliminary in nature. First, we analyzed only a portion of the browsing history of the participants (for efficiency reasons). And second, the method suffers from a risk of selection bias: users who participate in the evaluation are conceivably less privacy-conscious than the rest. Future work is needed to address these limitations.

On the topic of embarrassing ads, we note that although many users in our study recounted personal experiences with such ads, not all spoke about them until they were prompted by us. This limits the validity of our findings around ad-induced embarrassment and the intensity of user concerns for it. However, discussing this topic is not easy for all users (e.g., four of them in our study did not share details of their experiences, even when asked); so, *some* form of prompting seems necessary to understand individual perceptions of it.

It is plausible that the cultural context of our participants influenced some of our findings but this is difficult to verify using the results of the current study alone. We do not have any reason to believe or disbelieve that our findings on relative concerns for OBA and ad-induced embarrassment are specific to Indian users; more work is needed to resolve this issue. Independent of the answer, this remains the first study which explores perceptions of third-party tracking and OBA in depth with a non-Western population, examines questions from prior work on OBA in this context, and introduces new hypotheses on the topic of ad-induced embarrassment, which has remained unexplored in the literature on online ads.

6. CONCLUSION

While there has been much public debate and discussion around privacy issues in third-party tracking and OBA and growing consensus around improving consumer control mechanisms to address these issues, it is important to understand these issues relative to general user concerns around online advertising. Our study reinforces some of the concerns that have surfaced during this debate but it also finds that there is another prominent concern—that of embarrassment caused by online ads—which is also widespread amongst users and has gone largely unaddressed in previous research on ads. This concern, as we found, is intricately related to that of third-party tracking and OBA and is, in fact, amplified by the growing awareness around the phenomenon of OBA. Furthermore, current-day choice mechanisms do not seem to enable users to tackle this concern satisfactorily nor do they provide a satisfactory solution to all user concerns for tracking or OBA, as brought forth by our study. We hope that the ideas presented in this paper will take us closer to building end-user tools that satisfy users' requirements with respect to online advertising at large, give them greater choice and flexibility in their interactions with advertisements and simultaneously, satisfy the interests of the advertisers who support the Internet.

[12]For example, Ghostery, a popular tracking control tool, has less than 5% of the number of downloads of ABP on Firefox: https://addons.mozilla.org/en-US/firefox/addon/ghostery/.

7. ACKNOWLEDGEMENT

Thanks to Dhruv Patel and Saurabh Shukla, who contributed to the clustering apparatus used in our study. Thanks also to the anonymous reviewers of SOUPS 2013 and our shepherd, Sameer Patil, whose comments significantly improved the paper.

8. REFERENCES

[1] Do Not Track: an uncertain future for the web's most ambitious privacy initiative. http://www.theverge.com/2012/10/12/3485590/do-not-track-explained, 2012.

[2] M. Ackerman, L. F. Cranor, and J. Reagle. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. of EC*, 1999.

[3] M. Burke, A. Hornof, E. Nilsen, and N. Gorman. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction*, 12(4):423–445, 2005.

[4] D. Campbell and R. Wright. Shut-up I don't care: Understanding the role of relevance and interactivity on customer attitudes toward repetitive online advertising. *Journal of Electronic Commerce Research*, 9(1):62–76, 2008.

[5] Context Aware Do-Not-Track. http://www.context-aware-dnt.com/.

[6] Digital Advertising Alliance. Self-regulatory principles for online behavioral advertising. http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf, 2009.

[7] Federal Trade Commission. FTC puts an end to tactics of online advertising company that deceived consumers who wanted to "opt out" from targeted ads. http://ftc.gov/opa/2011/03/chitika.shtm, 2010.

[8] KPMG International. The converged lifestlye. http://www.kpmg.com/convergence, 2011.

[9] P. G. Leon, B. Ur, R. Balebako, L. F. Cranor, R. Shay, and Y. Wang. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proc. of CHI*, 2012.

[10] J. Mayer. Do Not Track is no threat to ad-supported businesses. http://cyberlaw.stanford.edu/node/6592, 2011.

[11] J. Mayer and J. Mitchell. Third-party web tracking: Policy and technology. In *Proc. of IEEE Symposium on Security and Privacy*, 2012.

[12] S. McCoy, A. Everard, P. Polak, and D. Galletta. The effects of online advertising. *Communications of the ACM*, 50(3):84–88, 2007.

[13] A. M. McDonald and L. F. Cranor. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *TPRC*, 2010.

[14] NAI. FAQs. http://www.networkadvertising.org/managing/faqs.asp.

[15] S. Panjwani, N. Shrivastava, S. Shukla, and S. Jaiswal. Understanding the privacy-personalization dilemma for web search: A user perspective. In *Proc. CHI*, 2013.

[16] S. Puntoni, I. E. de Hooge, and W. J. M. I. Verbeke. Embarrassment without Agency: A Theory of Ad-induced Embarrassment. Unpublished manuscript, Rotterdam School of Management, 2013.

[17] K. Purcell, J. Brenner, and L. Rainie. Search engine use 2012. http://pewinternet.org/Reports/2012/Search-Engine-Use-2012.aspx.

[18] F. Roesner, T. Kohno, and D. Wetherall. Detecting and Defending Against Third-Party Tracking on the Web. In *Proc. of NSDI*, 2012.

[19] A. Soltani, S. Canty, Q. Mayo, L. Thoma, and C. J. Hoofnagle. Flash cookies and privacy. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862, 2009.

[20] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas. Adnostic: Privacy preserving targeted advertising. In *Proc. of NSDI*, 2010.

[21] TRUSTe. Privacy and online behavioral advertising. http://www.truste.com/adprivacy/TRUSTe-2011-Consumer-Behavioral-Advertising-Survey-Results.pdf, 2011.

[22] J. Turow, J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy. Americans reject tailored advertising and three activities that enable it. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214, 2009.

[23] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proc. SOUPS*, 2012.

[24] C. Y. Yoo and K. Kim. Processing of animation in online banner advertising: The roles of cognitive and emotional responses. *Journal of Interactive Marketing*, 19(4):18–34, 2005.

APPENDIX

A. THE INTERVIEW SCRIPT

A.1 Part 1 - Introduction, demographic data collection and training

We are conducting a user study on online advertising. It is going to be an interview cum discussion where we tell you a few things about how ads work and collect feedback from you on some tools that we are building. Some parts of the study may require your laptop.

Please be honest while answering the questions. You can always say 'pass' if you don't want to answer any question.

1. Tell us your name, age, educational qualification, occupation and employment history.
2. How many hours do you spend online every day? 0-2? 2-4? 4-6? etc.
3. How many hours do you spend on a PC every day?
4. How many machines do you use? Do you own a smartphone?
5. Which browser on each machine do you use?
6. Which browser do you use the most? Any specific reason for using this browser?

7. Do you understand the term 'cookie'? If so, please explain what it does, in your own words.
8. Have you heard the term 'third-party tracking'? If so, please explain it in your own words.
9. Have you heard any of these terms: behavioral targeting, online behavioral advertising, targeted ads. If so, please explain in your own words.

At this point, the participant views the presentation on third-party tracking and OBA (referred to as 'behavioral targeting' during the rest of the interview). The interviewer paces the presentation based on the answers to the last three questions. During the presentation, some questions are asked spontaneously:

- Did you understand how third-party tracking works? Can you explain in your own words?
- Have you ever seen behaviorally targeted (or BT) ads before? If so, can you give us an example?
- How do you think BT ads help you? How do you think BT ads help advertisers?

A.2 Part 2 - Perceptions of tracking

1. What is your overall feeling about third-party tracking? Do you like it, dislike it or are you neutral? Are you concerned about third parties tracking your visits to different websites? What is your main concern?
2. On a scale of 1 to 5 (1 being least concerned and 5 being very concerned), how concerned are you about third-party tracking?
3. Can you give examples of websites on which you would *not want* third parties to track you?
4. For which of these topics would you not want any third party to track you, when you browse a website on the topic?
   - Financial investments
   - Job Search
   - Critical illnesses
   - Travel
   - Adult content
   - Gossip
   - News
5. Can you give examples of websites on which you would *want* them to track you? For what?
6. Which of these third parties do you trust the most (or do you trust all of them equally)? **(The participant is shown the slide from the presentation which portrays seven companies which practice third-party tracking: Google, Facebook, Yahoo, Microsoft, AOL, BlueKai, Quantcast.)**

At this point, we explain the quantitative part of the study to the participant and based upon consent, install our plugin on his/her laptop. The plugin runs in the background and the participant uses it towards the end of the interview. Just before the participant uses the plugin, we define two terms for him/her: (a) a cluster is called "sensitive" if the participant does not desire third parties to track him/her on URLs that might be present in the cluster; (b) a cluster is "bad" if it contains URLs that are completely unrelated to its label. The following questions are asked when the participant is using the plugin or has just finished using it.

1. **(While cluster-browsing)** How many bad clusters do you see? Can you give us an example of a bad cluster?
2. **(Later)** Would you like a feature in your browser which enables you to see your history in a clustered manner like this? If yes, why?

A.3 Part 3 - General attitude towards ads

Let us talk about ads now, first about ads in general and then about behaviorally targeted (BT) ads.

1. Mention a few words which come to your mind when you hear the term "online advertising". Does any adjective (good or bad) come to your mind?
2. Which ads do you prefer to see—TV ads or online ads? Which of the two do you find more useful?
3. Think of all online ads that you have seen in the past. These ads are on a variety of topics like travel, finance, electronics, etc. Are there topics on which you do *not want* to be shown ads? You can look at this list of topics as a reference. **(The participant is shown the ad inventory created using Google AdSense.)**
4. For each topic that you picked, what is the reason you don't want to be shown ads on that topic?
5. Are there topics on which you *want* to be shown ads? If yes, name a few.
6. Do you want to see these ads only when you are browsing for information on corresponding topics or even otherwise?

A.4 Part 4 - Embarrassing Ads

1. Have there been situations in which you were shown embarrassing content in an ad? If yes, could you give us an example of a situation in which this happened?
   - What kind of an ad was it? Did it embarrass you?
   - What website did you see it in? Did you see it on a Torrent site or some other website?

   (If the answer is no, we show the participant the ad inventory again and ask: Would there be topics on which if you were shown an ad, you might feel embarrassed? **We skip the rest of the questions.)**
2. On a scale of 1 to 5 (1 being least concerned and 5 being very concerned), how concerned are you about the possibility of being shown embarrassing ads?
3. Are these ads more embarrassing at work or at home?
4. Are these ads always embarrassing or are they embarrassing only when you see them in certain pages?
5. How frequently do you see such ads?
6. Besides the issue of embarrassment, is there any other reason why you may not want to see some ads when you are browsing in a place where others can watch your screen? If so, give an example of such an ad and the reason you may not want others to see such an ad.

A.5 Part 5 - Perceptions of OBA

1. Are you concerned about being shown BT ads? Why?
2. On a scale of 1 to 5, how concerned are you about being shown BT ads?
3. Suppose you see a certain ad while browsing and your browser told you that that ad is a BT ad. Would this extra information make the ad less or more acceptable to you? Or would it not change anything?
4. Have you noticed a BT ad being shown to you repeatedly? Do you think BT ads are more repetitive than non-BT ads?

At this point, the participant is told about three techniques to control behavioral targeting: browser controls, opt-out mechanisms and the use of blocking tools. In the last part, the participant is introduced to AdBlock Plus.

A.6 Part 6 - Questions around ad blocking, Miscellaneous questions

1. Have you heard of AdBlock Plus (ABP) or any other ad blocking tool?
2. **(For ABP-aware users)** Have you used it? Do you see any disadvantages to using it?
3. **(For ABP-unaware users)** Would you like to use ABP? Why or why not? Do you see any disadvantages to using it?
4. What do you think might happen if everyone started using ad blocking software?
5. Which of the following scenario would you prefer to be in?
   (a) All ads are shown to you as is
   (b) All ads except certain annoying ads (pop-ups, pop-unders and distracting ads) are shown to you
   (c) Ads on some topics of your choice are shown to you, the rest are blocked
   (d) All ads are blocked
6. Suppose you start using a tool like AdBlock Plus on your machine and your favorite news website starts charging you for browsing the site. How much would you be willing to pay?
7. Suppose I gave you two tools that you could install on your browser:
   - Tool 1 is a tracking prevention tool. It protects you from being tracked by third parties on particular topics. It will control third-party tracking on certain topics you select and will enable you to not see BT ads on those topics. But it will not block ads on other topics.
   - Tool 2 is an ad blocking tool. It blocks ads which you find embarrassing or irrelevant. But it does not stop third-party tracking on any website.

   Rate each of these tools on a scale of 1 to 5, 5 being very useful and 1 being least useful.
8. Have you ever used a browser plugin? If so, which one and for what?
9. Do you ever change your browser settings to (a) view or change your browsing history, (b) change cookie settings?

B. PRESENTATION SCRIPT

We used the following five-part script when making our presentation on third-party tracking and OBA to the participants. The *'s in the script indicate points where slide transitions occur.

Part 1 (Intro, Example). Advertising companies on the web use a technique called "behavioral targeting" to present ads to you that match your interests and browsing habits. For example, if you search for flight tickets from Mumbai to Aizawl on Cleartrip today*, it is possible that when you go to a different website* later on, you will be shown ads from Cleartrip on that site and these ads will be for the same* sector that you were trying to book tickets for earlier (i.e. from Mumbai to Aizawl).

Part 2 (OBA mechanics). How does behavioral targeting work?* When you visit a website like Cleartrip*, your browser interacts not only with Cleartrip's server*, but also with some third parties* that Cleartrip may have a relationship with*. These third parties could be entities like Google* which could either be serving content on Cleartrip (e.g., they may be showing ads on that site) or else they could be tracking different visits to that site. As Google interacts with your browser, it may place bits of information called "cookies" * on your machine which helps it recognize you in the future. A cookie is nothing but a small text file* containing a unique random identifier which is stored on your machine against the name of the party who sent it to you, which in this case is Google. It does not contain any personally identifiable information like your name or email address, only a random identifier which is uniquely assigned to your machine.

Part 3 (OBA mechanics, contd.). Later on, when you visit another website,* which also has a third-party relationship with Google*, Google can "link" these 2 visits as coming from the same computer by reading the cookie* it earlier placed there. Google* may now decide to show you a Cleartrip ad* on the new site because it knows that you were previously trying to book tickets on Cleartrip earlier. As you visit more and more websites (like Times of India, Flipkart, Clker, etc)* which also have third-party relationship with Google, Google will learn more and more about your browsing patterns and behaviors and can make smarter decisions about the ads to show you. This is what Behavioral targeting is.

Part 4 (Re-iteration, Questioning). So there are two concepts here. One is that of third-party tracking, using which third parties like Google track your visits to different websites and use cookies to collect information on the sites you are visiting. The other is that of behavioral targeting: based on third-party tracking, Google can get a better idea of the kind of sites you are visiting and serve you ads accordingly. And this process of serving ads is called behavioral targeting. Did you understand? Can you explain that to us in your own words?

Part 5 (Generalization). Google is not the only company which does this—there are several companies* other than Google which practice both third-party tracking and behavioral targeting (or BT). Here are some of them. Google

alone* has relationships with over 2 million websites on the Web* on which it either tracks you as a third party or serves you BT ads. These include Times Of India, Cleartrip, Myntra, etc. Not all the ads you see online are BT ads but a good number of them are.