“I don’t want this information to be tracked!”: Design and Evaluation of Information-Based Blocking of Web Tracking

1st author, 2nd author, 3rd author
Affilication
email

ABSTRACT
Web tracking is pervasive on the Internet and has raised significant privacy issues. Existing web tracking blockers such as Ghostery operate in a tracker-based model where they usually list the trackers present on a website and allow users to block all or individual trackers. However, prior research has found that users care more about what information is being tracked than who is tracking that information, implying the promise of an information-based blocking model. We designed Privacy Mirror, an information-based tracker blocker, which allows users to block web tracking by the types of information being tracked. To evaluate and compare our information-based model with the traditional tracker-based model, we conducted a user study with fifteen participants testing these two models. Our results show that the majority of our participants (12 out of 15) preferred the information-based model for several reasons, such as easier to understand and more user control.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI)

Author Keywords
Online Tracking; Web Tracker; Online Behavioral Advertising (OBA); Privacy; Blocker

INTRODUCTION
Online tracking can be categorized into first-party tracking and third-party tracking. First-party tracking is initiated by the cookies set by the websites that the users directly interact and visit (i.e., first-party sites) [2]. Third-party tracking refers to the practice that third parties such as ads and social media companies, embed their cookies across the first party sites such as New York Times, and collect users’ behavioral information and preference as they browse the sites [27]. Third-party tracking has become more prevalent and complicated over the years and has spread to the most popular sites [20]. Privacy and security discussion around it has taken place in technical, legal, and policy fields and evolved from one-time studies to some longitudinal and archaeological studies [20]. Online behavioral advertising (OBA) is a typical example of third-party tracking where ads company tracks “an individual’s online activities in order to deliver advertising tailored to the individual’s interests” [10]. In the rest of the paper, we use tracking / trackers to denote third-party tracking / trackers unless specified otherwise.

Given the prevalence of online tracking, Internet users have raised significant privacy concerns. For example, prior research has found that most users do not want to be tracked by OBA because it is an invasion to their privacy [32, 22]. However, people’s attitudes towards OBA could also be complicated and context-dependent based on different types of information being shared with OBA [19] and different online activities they undertake [34]. Some studies futher indicated that people’s understandings of third-party tracking and OBA might be inaccurate since they do not have a good knowledge of the mechanisms behind [23, 36].

Many blocking tools have been proposed and implemented to allow users to block online tracking and OBA. For example, Ghostery is a popular web tracking blocker that can identify the trackers on a website and empower the users to block them entirely or selectively [16]. Disconnect.me is a similar tool that visualizes and blocks the connections of third-party trackers [13]. Another tool, Privacy Badger, blocks third-party trackers’ cookies, prevents ads from loading, and uses different colors to display the blocking status of individual trackers [4]. A commonality among these tools is that they operate in a *tracker-based* blocking model where they usually list the trackers present on a website and allow users to block all or individual trackers.

However, one recent user study of online tracking found that most of the participants were more concerned about what information is tracked rather than who is tracking the information (i.e., trackers) [36]. This result implies the promise of an information-based blocking model where users can block web tracking by the types of information being tracked, for instance, blocking all trackers that collect/track users’ location or address. In this paper, we present Privacy Mirror, an information-based blocking model, implemented as a Chrome browser plugin. To evaluate and compare our information-based model with the traditional tracker-based model, we conducted a user study with fifteen participants testing these two models. Our results show that the majority of our participants (12 out of 15) preferred the information-based model for several reasons, such as easier to understand and more user

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI 2017, May 06 - 11, 2017, Denver, CO, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4655-9/17/05...$15.00
DOI: http://dx.doi.org/

control. Three participants preferred the tracker-based model either because they preferred to know who were tracking them rather than what information was collected or because on some website (e.g., Amazon.com), the tracker-list was shorter and looked simpler to block.

This paper makes three main contributions. First, we designed and implemented an information-based blocking model, a novel model that allows users to block online tracking based on the types of information being tracked. Second, we evaluated and compared the information-based model with the traditional tracker-based model, showing promising results of the information-based model. Third, we provide suggestions for future web tracking blocker. For instance, we suggest a hybrid model which includes the information-based model as a complement rather than replacement of the tracker-based model, and the information-based model can be the default blocking mode since most users prefer to use this model.

RELATED WORK

Our work is inspired by prior research on people's perceptions of online tracking, as well as the limitations of the current blocking mechanisms.

Online Tracking and Countermeasure Mechanisms

Online tracking, or more accurately speaking third-party online tracking in our research context, refers to the third parties' practice of embedding themselves in the first-party sites and record and profile the users as they browse across these sites [20]. Although third-party trackers have their value to provide free surfing experience to the users, and ads, analytic, and social media to the hosting first party sites, they also engender privacy violations and concerns [21]. Also, third-party tracking has become more prevalent which calls for increasing privacy defense for the users [20]. There are numerous tracking mechanisms such as by learning browser properties (i.e., "stateless tracking"), by embedding a unique identifier (i.e., "stateful tracking"), by storing HTTP cookies in users' browser, or by various means of fingerprinting [2].

To combat online tracking, numerous countermeasures have been proposed. For example, Do-Not-Track (DNT) is an HTTP head field that once enabled signifies that the user wants to be opt-out from being tracked [21]. However, since DNT has no enforcement mechanism, its effectiveness still depends on the website's acknowledgment and self-regulation [29]. Proxy server is another countermeasure that functions as an intermediary between the user and the destination server, so the latter does not know directly the user's IP address; also, sensitive information such as password can be encrypted on the proxy server [2]. The other countermeasures include using a more secure search engine such as DuckDuckGo [14], or a more secure browser such as Tor [8]. However, these countermeasures would require the users to have some knowledge or even expertise on online tracking and tool configuration.

People's Attitudes and Understanding of Online Tracking

People showed complicated attitudes and understanding of online tracking. On the one hand, from surveys such as [21], it is found that users do not want to be tracked. In fact, almost all stakeholders including EU (ePrivacy Directive, 2002) and US (FTC) agree that consumers should have certain degree of control over web tracking. Turow et al. [32] also found that 87% of the American respondents did not want to be tracked, especially by the online ads. On the other hand, however, people could be comfortable with online tracking for personalized search results so long as the information is not sensitive [24]. Sometimes people's understanding of online tracking is inaccurate, for example, people might misunderstand web trackers as malware and assume that online tracking directly affects local browsing history [23].

As a typical instance of online tracking, OBA stirs up much concern and controversy. People regarded it "creepy and scary" because of its online tracking nature and capability, but also perceived it "smart and useful" because of its tailored recommendation [33]. People from different cultural backgrounds also showed varied attitudes towards it, for example, comparing to the American users, Indian users were more concerned about the embarrassing content in OBA [1], and Chinese users were more willing to share their information to OBA [34]. People's attitudes could also vary according to different types of information and their retention [19], different online scenarios or activities [34], and different first-party sites they visited [9]. Finally, people's overall understanding of how OBA works, as represented by their "folk models" could also be incomplete or inaccurate [36].

Popular Tracking Blockers and Their Limitations

There are several popular tracking blockers which to the best of our knowledge, all target at blocking certain trackers or entities. For example, Ghostery [16] could detect and block various types of trackers such as OBA, social media, and site analyzers. Similarly, Disconnect.me [13] also detects and blocks trackers and could visualize the potential connections of a site to various trackers. Both Ghostery and Disconnect.me are classifying and blacklisting trackers by examining their service domains [31]. Privacy Badger [4] leverages a slightly different mechanism that requires less "custom configuration to block non-consensual trackers," and it judges and prevents a third-party by detecting its behavior across different sites hence it could be regarded as "behavior-based" instead of "blacklist based" [31]. There are also more severe blockers which inhibit all third-parties, such as Request Policy [26] or more ads oriented blocker such as Adblock Plus [25].

However, these blockers have their limitations. In a most recent study, Traverso et al. compared seven most popular blockers and found that none of them could deliver their promise for protection, even though Ghostery outperformed the other blockers [31]. Roesner et al. also found that third-party cookie blocking is ineffective and plugins would remove the buttons (e.g., the social media buttons) when blocking the tracking, which could be inconvenient to users [27]. Wills and Uzunoglu also found that the effectiveness of the current blockers varied, for example, Disconnect.me could only provide moderate protection and Ghostery, Adblock Plus and Adguard have to be manually configured to enable better protection because their default protection would be minimal [35]. Furthermore, Leon et al. tested nine online ads blocking tools and identified a

number of serious usability flaws, such as the interfaces were hard to understand, the opt-out options were hard to setup, and the settings were complex to configure [18].

In addition, in a previous study, Yao et al. found that most of their participants perceived the information being collected by web trackers are more important than who are the web trackers [36]. Inspired by these findings, we designed and implemented an information-based blocking model, an alternative way to allow users to block third party web trackers by the types of information they track. We conducted a user study to compare the new information-based model with the traditional tracker-based model, and to collect people's feedback for future tracker blocker development.

METHODOLOGY

Our research methodology consists two main parts: a prototype system design and a user study. In the prototype system design, we developed a Chrome browser plugin called "Privacy Mirror" in two different versions. One version mimics the tracker-based blocking mechanism that spreads in the current web tracker blockers; the other version mimics the information-based blocking mechanism which we propose based on the existing literature. In the user study, we asked our participants to go through these two different plugin versions in three online scenarios, evaluate the pros and cons, and explain their preferences. We present the details below.

Privacy Mirror Prototype Design

We developed the two prototypes in an iterative manner. Such a research strategy is also inspired by several prior HCI studies [5, 30]. In the early exploration phase, we designed a prototype of the information-based blocker, and conducted a preliminary comparative study using this version of Privacy Mirror and Ghostery. We found that our preliminary study participants overwhelmingly (three out of three) preferred Ghostery mainly because of its commercial reputation and its elegant and sophisticated User Interfaces (UI). As the key idea of the study is to evaluate two different blocking models while controlling for factors such as the quality of graphical UI design or the reputation of the tool developer, based on the feedback we collected from the preliminary study, we made two major improvements to our system design and study design.

First, we redesigned the user interface of the Privacy Mirror to the current version with many details. Second, to control for these other factors that can potentially affect people's preferences, we developed another blocker using the tracker-based model, which had the same tracker-based blocking features as Ghostery. The information-based version would enumerate in real-time what types of information was tracked on a particular website, and the tracker-based version, like other existing blockers' practices, would show in real-time the names of the trackers on a particular website. Both versions allowed the users to block tracking or trackers selectively or by all. We used these two versions of blockers in our user study.

In terms of data source, we used Ghostery as our benchmark tool, and used the data provided by Ghostery as our primary data source, since Ghostery has been proven to be able to block over 90% of trackers, and thus provide best protection from trackers among all the commercially available blockers [31]. In the next section, we explain in detail how we get data about what information is collected, and who are the trackers on each website. We will take reddit.com as an example.

Trackers. We identified trackers on reddit.com with the following steps. First, we installed Ghostery plugin in our Chrome browser. We then visited reddit.com, and used Ghostery to examine all the trackers on the site. Second, we identified three tracker categories based on the number of trackers in each category provided by Ghostery, i.e., Advertising, Site Analytic, and Social Media. All other categories (e.g., Essentials, Comments) were put in "Other" due to the space limit of our design. Third, we manually clicked on each tracker to access its profile and URL. We copied the profile to our system, and used the URL as identifier to search for trackers in our system. It is worth noting that, every time we refreshed the page, we received a slightly different tracker list. Thus, to ensure we captured all the trackers on the site, we repeated step one to three a few times till there was no new tracker appeared. Figure 1b showed the final tracker list on reddit.com, and Figure 1c showed the tracker list provided by Ghostery on the same site.

Information Collected. After we captured all trackers on the site, we then manually checked the full profile page of each tracker on the Ghostery website. In the tracker full profile page, there was a collapsed section named "Privacy Information." In this section, there was a sub-section called "Data Collected" where it showed all the information that can be collected by this tracker. We manually scripted all the information in this sub-section, and added these information to our system. It is not clear how Ghostery retrieved or inferred this information, but since the main goal of our study is to compare the two blocking models, we believed that hardcoding this information in our information-based blocking model is acceptable because Ghostery can easily implement the information-based model. In a way, this hardcoding aspect of our study is similar to the Wizard of Oz method [11]. We repeated this process for all trackers on the site, and captured what information was collected by each tracker. Figure 1a showed the final information list on reddit.com.

User Study

We conducted a user study during the summer of 2017 to evaluate and compare the two different blocker versions. Combining different recruitment approaches such as Craiglist, word-of-mouth, and on-campus flyers, we recruited 15 participants with diverse backgrounds and demographics. We also selected five websites representing five common online contexts such as online shopping (Amazon.com), online banking (Chase.com), Health consulting (Healthcare.gov), Blog & social media (reddit.com), and online news (CNN.com). In each user study, we randomly selected three from these sites, and tested the two blocker versions on the selected websites. All interviews were audio-recorded and screen-recorded upon participants' permission. Our study has been approved by the IRB.

The study procedure is as follows: first, we invited our participants to our lab and asked a few general questions about their demographic information and the Internet usage. Without

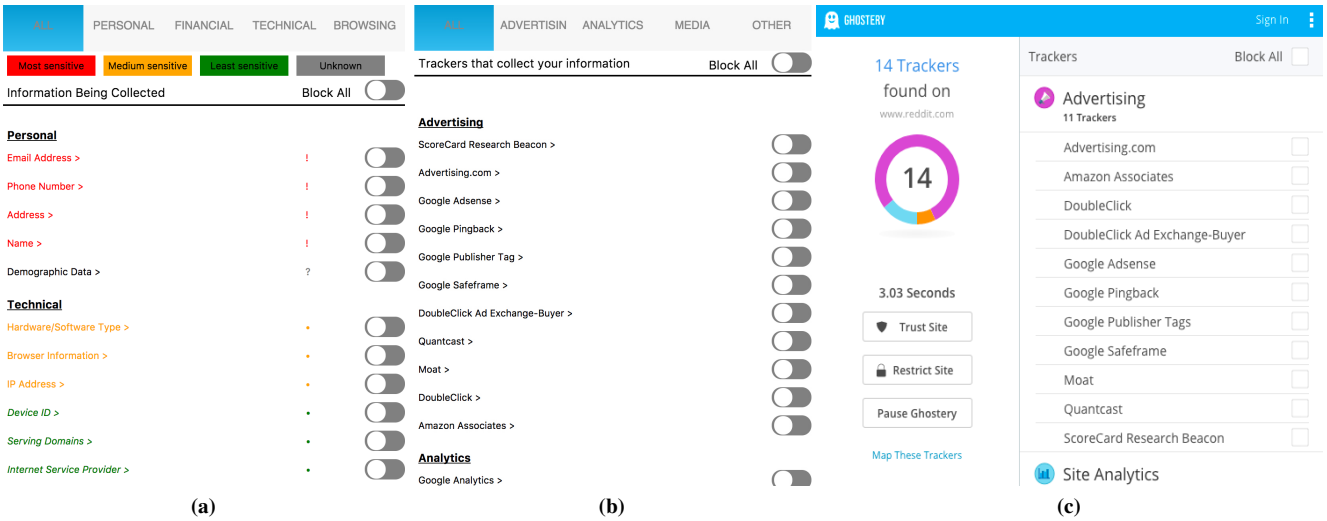

Figure 1: Prototype of Privacy Mirror on reddit.com. (a) Information-based Blocker. (b) Tracker-based Blocker. (c) A screen shot of Ghostery.

priming them to privacy or security concern, we also inquired about their experience and attitude of online advertising, and we probed whether the ads they mentioned were targeted ads or random ads. Regardless of their prior knowledge of targeted ads, we provided a definition, "Targeted ads refer to the practice of tracking an individual's online activities in order to deliver advertising tailored to the individual's interests." Then we asked our participants about their experience and perception of web trackers. Regardless of their prior knowledge of web tracker, we introduced a definition, "Web trackers are entities that collect user information and monetize it either by selling it or directly using it (e.g., providing targeted ads)." Both definitions were inspired by [36].

Second, we led our participants to the scenario-based lab study and let them surf the Internet via the Chrome Browser for three minutes. Then, we asked our participants to browse a website we randomly chose from the five scenarios, for example, we asked them to browse Amazon.com for a few minutes. It should be noted that to protect our participants' privacy, we have cleaned the cookie and browsing history before every scenario, and we assured our participants that they need not and should not provide any personal information when browsing any website on our lab computer. In the scenario, we would further let our participants do a task, e.g., shopping a pair of shoes, and encourage them to think aloud and explain what they were doing and what information might be potentially collected during this process.

After a period of browsing and thinking aloud, we let our participants to randomly click one of the two plugins on the Chrome Browser. One plugin was the information-based blocker prototype, and the other was the tracker-based blocker prototype. We carefully avoided introducing the plugins as developed by us to reduce the potential social desirability bias. Then, we let our participants play around with the plugin and express whether anything concerned or confused them. We suggested them block necessary items (specific type(s) of information in the information-based blocker or tracker(s) in the tracker-based blocker) to protect their information from being tracked. We then asked them to explain their blocking choices and rationale. After the test with one blocker, we repeated the same test procedures to the other.

Finally, we conducted an exit interview to probe our participants' understandings and preference of these two blockers. We started by asking "What do you see as the major differences among these tools?" Then, we asked which one they would prefer and why. After receiving their feedback and preference, we revealed to our participants that these two blockers represent two design ideas: the *information-based* blocking model and the *tracker-based* blocking model, and we asked them again about their understandings and attitudes between these two design ideas. We ended our interview by soliciting our participants' suggestions on how to improve each blocker design and how to better protect Internet users' online privacy in general.

Data Analysis

The interview audio recordings were transcribed, and analyzed through a thematic analysis [7]. Two coders (co-authors) independently coded the same subset of our data at sentence level to develop codes that captured the underlying meanings of our data. Then two coders discussed, and merge their codes into a code book. The two coders then used this code book to code the rest of our data. Upon finishing, the coders discussed the codes, and groups all codes in the code book into seven themes, including: prefer information-based blocker, prefer tracker-based blocker, decision about blocking information, decision about blocking trackers, understandings of blockers, system limitations, and suggestions.

When participants were asked to try out the blockers, we logged their behaviors, including which blocker they were using, which items they blocked, which website they were on, and time-stamp. We also screen recorded this process. The

log data was used to calculate participants' average time spent on every task, and the screen recordings were used to provide context for us to better understand the log data.

RESULTS

Next, we report our results, focusing on people's perceptions and experiences of the two blockers.

Participants

A total of 15 participants completed our user study. Their ages range from 20 to 40 (Mean=28, Median=29, STD=6.2). Nine were male and six were female. Twelve of our participants were from local communities (e.g., Craigslist), and three were recruited using snowball sampling [17]. Our participants also represent a wide range of occupations, including student, librarian, social worker, university employer, software developer, nurse, and accountant. In terms of the Internet usage, all of our participants use the Internet on daily basis, and eight of them spend at least 50% of their time on desktop computers.

Prior Knowledge about Web Trackers

Twelve participants reported that they have never heard of web tracker. Three participants have heard of web tracker, but not exactly sure what that means. We asked them to define web tracker before we provided an explanation. For example, P4 thought web tracker as a program that can collect user information such as search history and visit history.

"Maybe web tracker is like a program tracking what you are looking at on the Internet, like your browsing history, and the cites you frequently go to such as sports or campaigning." (P4)

P11 considered web tracker as a person who collect your search history in order to provide target ads.

"It's kind of a step before targeted ads, like the person collecting the data for what you are searching." (P11)

In general, although most participants did not have any prior knowledge of web tracker, through our ice-breaking questions, most of them formed reasonable understandings of it and perceived web tracker as having the ability to track their online information, such as search history.

Current Practice to Prevent Online Tracking

When asked what did they do to protect their information online, most participants mentioned that they were not aware of any measures that they could take to protect their privacy online. The common way was to be cautious online, and to avoid providing too much information online. For example, P1 commented,

"The only way I can control is to watch what I say, watch what I post, watch what I go on and put in privacy settings based on which site you are on. Nothing is private." (P1)

A few participants mentioned about the actions they took, such as installing tools like AdBlocker (four out of fifteen), using Incognito mode in Chrome (one out of fifteen), and paying attention to the *https* icon in front of the URL (one out of fifteen). None of our participants mentioned any commercially available tracker blocker, such as Ghostery [16] or Disconnect.me [13].

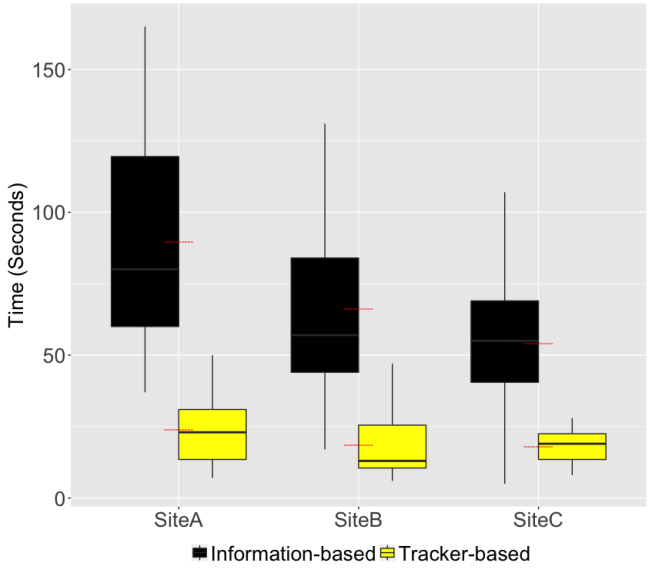

Figure 2: Time of successfully complete task with information-based blocker and tracker-based blocker. The red line in the middle of the box represents the mean of each catetory.

Task Completion Time

Figure 2 shows the average time our participants used to complete the tasks in the three websites. Overall, participants on average spent significant more time (Wilcoxon Test, $P<0.05$) finishing the tasks on the information-based blocker (70.3 seconds) than the tracker-based blocker (20.6 seconds). Based on our observations during the user study, one major reason why participants spent more time on the information-based blocker was that they took their time in exploring the blocker and thinking about whether to block specific type(s) of information. For example, before blocking any item, many participants thought about whether the information was sensitive to them, and whether the website needed such information to provide them services. On the contrary, when using the tracker-based blocker, many participants indicated that they were not familiar with the trackers, thus they made their decisions purely based on the tracker's name. For this reason, even though participants spent more time with the information-based blocker, the majority of them still preferred it because they could make well-thought out decisions. However, one can argue that Internet users often do not have the time or patience to spend much time to make these blocking/privacy decisions, therefore ways to reduce the amount of time needed for users to make informed decisions in the information-based model is an important area for future research.

Why Prefer Information-based Blocker?

When asked about which blocker they preferred to use and why, twelve out of fifteen preferred the information-based blocker for a number of reasons. We present the results below.

Features and Interfaces

Easy to Understand. Several participants preferred the information-based blocker because it was easy for them to understand what was displayed in the blocker. For example, P1 said,

"The first one [information-based] was easier to understand, name, demographic, email, I understood what each tab meant. The other one [tracker-based] I didn't understand...So if you use words that are familiar with people then I guess that's okay. It was more friendly. I know exactly what I was blocking." (P1)

She suggested the information-based blocker was very easy to understand since it used many common words, while in the tracker-based model, she did not understand the meaning of trackers, thus she was more certain about what she was blocking with. This highlights that people could easily understand the different types of information being collected, and it was harder for them to understand trackers, especially for those who were not technically savvy. Thus, the information-based blocker leverages the fact that it is generally easier for people to understand information than trackers.

For example, P3 noted that the information-based blocker could help him to block the most sensitive information based on his own privacy mental models.

"Specifically, it is easier for me to block the most sensitive information." (P3)

He thought about which information was sensitive to him, and block all that he considered sensitive. P11 explained more specifically.

"Like demographics and contact information, there is the financial information, it makes me nervous, it's more personal...This looks more sensitive to me." (P11)

She listed several information that were more sensitive to her. In summary, due to the information-based blocker's ability to show users' information sensitivity based on their preferences, it is more favored by most participants.

Perceived More Control of Information

Another reason why many participants preferred the information-based blocker is that these participants perceived that by showing them what information was being collected, they would have more control to their own information online. For example, P4 noted,

"I feel like it [information-based blocker] gave me more detailed like gave me more control, like I could say I want to block my name and address and my e-mail address but I don't care if they have my location or whatever. I feel it was more detailed." (P4)

He highlighted that in the information-based blocker, he has the ability to decide which types of information he would want to share with web trackers, and which types he would not. He felt such ability gave him more control over his own information. P3 made a similar argument,

"I feel like I would have a little bit more freedom with the information-based." (P3)

By saying *more freedom*, she referred to the ability to block any information she wanted to block based on her preferences, and she thought such ability was limited in the tracker-based blocker since she did not know what information each tracker could collect.

Contextual Adaptability. Some participants preferred the information-based blocker because it could adapt to different contexts and offer varied blocking choices accordingly. By contrast, they regarded the tracker-based blocker as lacking such a contextual adaptability because it would show similar trackers across the online scenarios. P2 explained,

"[On] Amazon I didn't block my phone number and my e-mail because I know Amazon already has that. Then reddit I did block a lot of the more personal things because it seemed like things that reddit might be collecting that I don't think are necessary like operating system, then Washington Post I talked about that a little bit earlier but I think my personal information for Washington Post [that] I'm not interested in sharing." (P2)

In her view, she could use the information-based blocker to selectively block different types of information according to the genres and her preferences of the websites.

Some participants also anticipated that certain information should be prevented from tracking in all online contexts because they are sensitive regardless of the sites. P6 noted,

"I'm not sure how the other person is going to use that data in the website, but I'm sure what kind of data I don't want to share with others...Maybe I'm not putting the e-mail address in some of the websites, but who knows if they can track it with that web page they may track it from another web page but I may be putting the e-mail address, I don't want to do that, so that is the main reason." (P6)

She preferred to blocking similar kind of information across all websites because what should and should not be collected does not depend on websites.

Our participants did not mention about their contextual choice or preference regarding the tracker-based blocker, indicating the potential for the information-based blocker to provide context-based blocking suggestions.

Perceived Effectiveness

For those who preferred information-based blocker, they perceived that the information-based blocker would be more effective than the tracker-based blocker for a number of reasons. The first reason is that it would be effective in the long run, especially when a new tracker comes out but it is not added to the database of the tracker-based blocker. For example, P10 highlighted this point,

"If some new web tracker emerges, and the developer of the blocker tool didn't know, and then the new web tracker can collect my information and since he didn't update the tool, so then I think it's [tracker-based] useless." (P10)

In his opinion, if there is a new tracker, the blocker developer may not know about it, and thus fails to update the blocker in time, then his information would still be tracked. In this case, he preferred the information-based blocker as he perceived that once he blocks a certain type of information, this type of information would not be tracked by anyone, even if it is a new tracker.

Another reason why some participants believed the information-based blocker to be more effective is that compared to the tracker-based blocker, the information-based blocker generally provides a more detailed list of types of information. Our participants perceived such detailed list as more robust.

Why Not Information-based Blocker?

Three of our participants preferred the tracker-based blocker mainly for two reasons. The first reason is that some participants wanted to know who was tracking them, and that was more important than their information. P7 explained,

"Well they're already collecting information so probably who's doing it is more important ... Like if it's just a marketing company trying to make, trying to do business, alright they're trying to sell me stuff, that isn't going to hurt me, it might be annoying but if it's like the Russians, the anonymous, whoever is out there, if it's somebody who may want to do something harmful with the information, [that's a problem]." (P7)

He has this stereotype that his information would be collected anyway. He explained that if his information was used by unknown people or entities for malicious purposes, then that would hurt him. Thus he cared more about who could collect his information rather than what information would be collected.

Another main reason was that the tracker-blocker normally contains a shorter list compared to the information-based blocker, thus it looks simpler and cleaner. This is mainly because, in the current information-based blocker, all information that could be collected by trackers are displayed. Even though the system only shows the types of information that are being collected on the current site, it always has a relatively long list compared to the tracker-based blocker. Such phenomenon is particularly true on websites such as Amazon, as there is one web tracker (i.e., Amazon Associate) on Amazon.com. However, this tracker tracks over twenty types of information, and thus makes the list in the information-based blocker noticeably longer than the list in the tracker-based blocker. P9 first complained that he clicked too many items, indicating that he did not like the long list in the information-based blocker. He further commented about the tracker-based blocker,

"If I click Google Analytics, then all the information that is collected from Google Analytics will be blocked, so it's more convenient for the user." (P9)

In his explanation, he argued that in the tracker-based blocker, as long as he blocked a tracker, all types of information would be blocked, thus it was more convenient.

Mental Models

Perceived Purposes of the Blockers In the user study, when we showed our participants the two blockers, we asked them to describe their perceived purposes of both blockers. We found that most participants formed a reasonable understanding of the information-based blocker regarding its purpose, while fewer participants had a comparable level of understanding of the tracker-based blocker. For example, on Amazon.com, when asked what was the purpose of the information based blocker, P2 said,

"It gives you more privacy and more control over what information is made available to I guess marketing companies as well as any sort of like anything that you may be tracking. So like if I click block e-mail address, if I try to type in, like if it asks me to create an account or something and it will not allow me to put an e-mail address." (P2)

She suggested that once she blocked the email, her email address would not be collected, which reflects a good understanding of the tool's functionality.

In comparison, when asked what was the purposes of the tracker-based blocker, P2 said,

"This one I guess for me would be more difficult to use I guess because I don't know what any of these are, and maybe for like the general user." (P2)

Her response showed that because she did not know what did the tracker names mean since she had never heard of those, she had a hard time figuring out the purpose of the tracker-based blocker. In fact, this was the most popular reason why many participants did not reasonably understand the purpose of the tracker-based blocker.

Surprised by the Amount of Information Being Tracked

In the user study, right after we asked our participants to browse the web page, we asked them whether their information could be collected during their browsing. Many participants thought that some minimum information could be collected. We then asked our participants to use the two blockers to explore what information was being tracked or what trackers were tracking them. Many participants were startled to see the amount of information that were tracked. For example, P8 noted,

"My financial information for one thing, I mean this is really I didn't think would be on there, and demographics and PII [personal identifiable information] should be posted on here...Yeah, that's basically what surprised me." (P8)

She was on reddit.com, and she assumed very minimum data could be collected. However, the information-based blocker showed her a list of information that could be tracked on the site which included financial information, demographics, and PII, thus she was very surprised to learn this.

On the contrary, in the tracker-based blocker, our participants were generally not surprised to see who are the trackers that were tracking them. The main reason is that our participants did not recognized most trackers. The ones they did recognize

were from Google and Amazon, which they already expected in their mind.

Why Block/Not Block Information?

When our participants used the information-based blocker, we asked them to explain why they decide to block or not block certain information types. The most popular reason to block a type of information was that the information was sensitive to them, and they felt uncomfortable to share, such as financial information, name, and address. For example, P15 noted,

"It's just that financial information I would always block, phone number always because if I don't know the sites and if I were to visit them, I don't want them to have my name, my phone number, my address and my financial information. The other things is not a big deal." (P15)

In P15's case, she would always protect her financial information, name, phone number and address since she did not feel comfortable sharing that information with anyone. Another reason for our participants to block a certain type of information is that they felt the sites do not need this information to provide services. For example, P2 explained,

"I don't really care that reddit know what I'm looking at and how long I'm spending on it, I mean that might be helpful for their developer, but I don't think they need to know like my demographics." (P2)

She explained that on reddit, she would be willing to share some technical information such as click-through or page views since she believed that information can help the website developers provide better services. However, she refused to share some other types of information, such as demographic information, since she believed that reddit, as a social media sites, did not need that information to provide her services. The last reason our participants mentioned about why they decided to block certain types of information was that they believed that by blocking certain information on certain sites, they could receive better services. P11 explained,

"Because I feel like CNN, it has news as relevant information and more political. So I guess it's [technical information] more valuable in a sense. Valuable because it has my searched and stuff. And I feel like this is more of a personal, saying like I am on reddit for fun, trying to see what interests me. Whereas the news is, it's important and I need that information. I don't want it to be affected by anything." (P11)

When she was on the news site, she mentioned that technical information, such as page views and click through, were very important to the site, since as a news site which contains political information, the site could use her technical information to infer her news preferences or even political preferences, then display tailored content on her homepage to bias her opinions. She was hoping to get more neutral and justice news, thus she was very cautious and blocked all technical information on the news site. She further mentioned that that information was not important on social media sites such as reddit since she would only use that site for fun.

Interestingly, when our participants talked about the reasons why they did not block certain types of information, some of them mentioned that they did not block certain types of information because they wanted to receive better services from the sites, such as targets ads and product information. P15 explicitly mentioned that he would not block certain information from being tracked since he wanted to receive better targeted ads.

"I don't mind targeted ads for specific reasons...Maybe if I want my experience with the ads better, I don't mind if it is collected for that reason." (P15)

Another reason is that for some participants, they understood why there was a need to track their information. They considered their information as the price they pay to receive free Internet services. For example, P7 said,

"Nothing is for free right, so somehow they're warehousing this information, they're monitoring it and then trying to monetize it right by these targeted ads...If they would pay me for my information, like if they're already getting this anyway and now I have the ability to block it, if they would somehow pay me somehow, I might say okay track me, like give me a free whopper once a month or something." (P7)

He understood that his information was used to pay for the free Internet he got. Once he is given the ability to block his information using the information-based blocker, he was willing to share some information with the third party trackers for some payment.

In summary, our participants decided whether to block certain types of information or not depending on a number of factors, such as whether the information is sensitive, and whether the information is useful and needed for the website.

Why Block/Not Block Tracker?

We made similar requests to our participants, asking them to explain why they chose to block or not block certain trackers in the tracker-based blocker. Our findings suggested that most participants chose to block or not block a certain tracker depending on whether they had knowledge of this tracker. Specifically, some participants blocked a few trackers because they did not know these trackers' name or behavior, which made our participants feel insecure. Other participants did not block the unrecognized trackers because they wanted to know more about these trackers before blocking them. Previous literature suggests that most users blocked trackers from all companies since they did not recognize the trackers [18]. Therefore, our results revealed a major flaw of the tracker-based blocker, that users tend to make their blocking decisions purely based on their familiarity with the names of trackers. Such flaw in the tracker-based block creates opportunities for the information-based blocker, which provides users more transparency and control of their online information in an accessible way.

Suggestions for Improvement

At the end of the user study, we asked our participants to provide suggestions for future improvement for both blockers, or online privacy protection in general. Our participants sug-

gested that for the information-based blocker, there should be a brief explanation of each individual term such as the meaning and purpose of collecting that type of information. Such explanation would be particularly useful for the more technical terms, such as "IP address", "Personal Identifiable Information", and "Click Through". For the tracker-based blocker, our participants suggested that there should be more information about each tracker, such as the type of information it would collect. Some participants also suggested that the information-based blocker could categorize similar types of information together so that the users would not be overwhelmed by a long list of information. These suggestions shed light on the improvements for future blocker design and implementation.

DISCUSSION

Our results suggest that the information-based tracker blocker is a promising approach to help users manage their online privacy. Even though on average, our participants spent more time completing tasks using the information-based blocker than the tracker-based blocker, most participants still preferred the information-based blocker because it is easier to understand and use. Participants also perceived that they would have more control of their information in the information-based blocker, so it appears to be more effective than the tracker-based blocker. Because of these reasons, most participants preferred the information-based blocker. Three participants favored the tracker-based blocker because they wanted to know who were tracking them. Besides, due to the limited number of trackers and the ample amount of information trackers could collect, the tracker-based blocker provided a relatively cleaner interface, which was another reason why a few participants chose the tracker-based model. Such findings indicated that both blockers have their own pros and cons. We compare them below.

Comparison Between Two Blockers

The results of our study show a few advantages of the information-based blocker compared to the tracker-based blocker.

The most noticeable advantage is that the information-based blocker provides a higher level of perceived ease of use and perceived usefulness in protecting users' information from being collected online. Based on our study results, most participants were able to figure out the purpose of the information-based blocker, but had a relatively hard time to figure the purpose of the tracker-based blocker. The most important reason is that information-based blocker provides more intuitive and familiar items to the users, while the tracker-based blocker provides a list of tracker names, which most users seldom interact with directly [18]. In addition, since the information-based blocker provides more detail information in an user friendly way, a number of our participants felt that they had more control over their own information, and the blocker was also more robust and effective, thus more useful. The Technology Adoption Model (TAM) has suggested that people's perceived ease of use and perceived usefulness can largely affect people's decisions about future technology adoption [12], thus we believed that the information-based blocker has the potential to be widely adopted in reality.

We also believe that the information-based blocker initiates people's thinking process regarding which ones to block or not block. In terms of the task completion time, we noticed that our participants spent more time completing tasks on the information-based blocker, and less on the tracker-based blocker. From our observation during the user studies, participants spent more time exploring the information-based blocker and thinking it through before blocking any item in the list. Additionally, in terms of reasons to block or not block, in the information-based blocker, participants chose to block a certain type of information for a number of reasons, such as they did not feel comfortable sharing that information, they wanted better services, and whether the sites they were visiting need such information. These rationales came from participants' thorough considerations. As a comparison, in the tracker-based blocker, participants made their blocking decision based on the name of the tracker, such as whether they have heard of the company, whether the company was trustworthy, etc. Thus, we believe that the information-based blocker can really help users think though, and eventually make decisions that better suit their privacy preferences.

Meanwhile, we also notice a few disadvantages of the information-base blocker. One major drawback is that, often times, the information list is longer than the tracker list since many web trackers track more than one type of information. For example, Amazon Associate collects more than twenty types of users information, including ad vies, demographic data, serving domains, IP address, etc [3]. On Amazon.com where there was only one tracker ("Amazon Associate") in the tracker-based blocker, there was a relatively long list of information in the information-based blocker. Thus some participants perceived the tracker-based blocker to be much simpler.

The Importance of Information-based Blocker

After discussing the pros and cons of information-based blocker, the next natural question to ask is why it is important to have an information-based blocker, especially when the current commercially available blockers all adopt a tracker-based mechanism.

We believe the information-based blocker is a critical step in protecting people's online privacy for three reasons. First, due to its high level of perceived ease of use and perceive usefulness, an information-based blocker would be more accessible to users, especially to those who do not have strong technology background.

Additionally, in our study, many participants expressed their surprises when they realized the amount of information that were collected online. This is an important observation which suggests that many users are not aware of the privacy risks they are facing online. By implementing an information-based blocking model, the blocker can provide users with a holistic view of information collection online, which eventually will contribute to a more transparent online environment.

Lastly, the information-based blocker allows users to block the information that they are uncomfortable to share and unblock certain insensitive information to the trackers. Since the

Internet economy is based on web tracking and users' information [28], such practice will not only protect users' privacy online, but also protect the entire Internet economy.

System Limitations and Future Improvement

Based on the study results, we now discuss the limitations of current information-based blocker, as well as how to improve it in future development.

First, the information-based blocker is designed to show the users what information can be collected on each site that a user visits. However, at this stage of development, we have not automated this process. In the current version, we hard coded all the trackers on our testing sites and the information they collect into the system. As discussed in previous section, we used Ghostery as our data source to extract tracker list as well as what information each track collects.

As far as improvement, since the information-based model needs to know what information is being tracked, we advocate the tracker sites to disclose such information explicitly in their privacy policies so that in the future, this information can be more accessible for tracking blockers. We also believe disclosing what information is collected is a mutual beneficial way for both users and tracker companies, as users will have more transparency to protect their sensitive information, and at the meantime, share other non-sensitive information which tracker companies will desire. Moreover, in the next version, rather than hard coding, we plan to automate this process by applying machine learning algorithms in detecting trackers and capturing what information is being collected by each trackers. For example, with OpenWPM web privacy measurement framework [15], one possible approach is to infer what information is being collected using the URLs parameters through which tracks use to communicate and send data to server.

Second, as shown in Figure 2, participants spent significant more time on the information-based blocker across all three test sites. In the current version, participants needed to manually select which types of information they wanted to block on each site, as we cleared users' choices from other test sites for the purpose of the study. This process was time consuming. Even though participants preferred the information-based blocker, literature has suggested that people are not willing to spend too much time opting out from data collection [6].

In the future version, one critical aspect we need to consider is how to reduce the configuration time for users. We plan to equip more advanced features in the blocker using artificially intelligence and machine learning. For example, when the users start to use the blocker, they need to manually configure it. Then the blocking preferences will be applied across all sites, and the users can make adjustments based on which sites they visit. Overtime, the blocker can learn users' contextual preferences of their information, and configure the blocker automatically when users visit new sites. Additionally, based on users' preferences, the blocker can selectively display information that users most care, and hide the other information in the main interface. Those hiding information will still be accessible by users with extra clicks, but in this way, the amount of information displayed in the main interface will be largely reduced, which will make it more favorable by users based on our results.

Additionaly, the current system did not show who were tracking each type of information. Our work suggests that no single blocking mechanism is a silver bullet that fits the needs of all users. The information-based blocker is more favorable by most participants, while the tracker-based blocker has some implications to draw. For future development of blocker, rather than complete replacement, we suggest a hybrid model can be applied where the information-based model and tracker-based model can co-exist and complement each other. For example, a hybrid model can use information-based model as the default mode since the majority of participants preferred it. For each type of information collected, the blocker should also show who are the trackers that collect such information. This hybrid model will suit the needs of those who think the information is more important, and at the same time, provide enough details for others who prefer to know more about the trackers.

Finally, the current system did not include the purpose why each type of information is collected. Based on our findings, such information was desired by many participants, and would be useful in helping users make blocking decisions. We plan to add this feature in our future version.

Study Limitations

We admit that our study has a few limitations.

First, our study was conduct in our lab, and our participants were observed by researchers while performing tasks. Thus the testing environment is not participants' natural environment. In addition, each user study took about one hour to finish, which did not allow our participants to observe the long term effect of using the blockers. However, the observation and the think loud session during the study allowed us to understand participants' perceptions of two blockers more closely. These insights are invaluable in deeply understanding participants as well as informing future designs.

Second, we used our lab computer in the study rather than participants' own computers. Since participants' own computer contains more personal information and lab computer does not, participants may have different behaviors on lab computers. However, we asked our participants to make their decisions based on the assumption that this is their personal computer, so we believed that our results are reasonably validate.

CONCLUSION

Online tracking is prevalent and many tracking blockers have been developed. However, one major drawback of existing systems is that they adopt a *tracker-based* model, and provide the users with a list of trackers which users generally have difficulties understanding. In this research, we proposed and tested Privacy Mirror, a prototype of an *information-based* blocker that allows the users to selectively block different types of information from being tracked. Our user study indicated a positive evaluation of our prototype. We propose that the design of such an *information-based* blocker could complement to the current *tracker-based* blocker to give the users more choices and empowerment.

REFERENCES

1. Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. 2013. Do not embarrass: re-examining user concerns for online tracking and advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*. ACM, 8.
2. Muhammad Aleem, Asra Ishtiaq, Salma Hussain Abbasi, and Muhammad Arshad Islam. 2017. User tracking mechanisms and counter measures. *International Journal of Applied Mathematics, Electronics and Computers* 5, 2 (2017), 33–40.
3. Amazon Associates. 2017. (2017). https://apps.ghostery.com/en/apps/amazon_associates
4. Privacy Badger. 2017. Privacy Badger blocks spying ads and invisible trackers. (2017). https://www.eff.org/privacybadger
5. Natã M Barbosa, Jordan Hayes, and Yang Wang. 2016. UniPass: design and evaluation of a smart device-based password manager for visually impaired users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 49–60.
6. Cara Bloom, Joshua Tan, Javed Ramjohn, and Lujo Bauer. 2017. Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. In *Symposium on Usable Privacy and Security (SOUPS)*.
7. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
8. Tor Browser. 2017. The Tor software protects you by bouncing your communications around a distributed network. (2017). https://www.torproject.org/projects/torbrowser.html.en
9. Farah Chanchary and Sonia Chiasson. 2015. User Perceptions of Sharing, Advertising, and Tracking.. In *SOUPS*. 53–67.
10. Federal Trade Commission and others. 2009. Self-regulatory principles for online behavioral advertising. (2009).
11. Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studiesâĂŤwhy and how. *Knowledge-based systems* 6, 4 (1993), 258–266.
12. Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.
13. Disconnect. 2017. Used by over a million people âĂŞ Disconnect lets you visualize block the invisible websites that track you. (2017). https://disconnect.me/disconnect
14. DuckDuckGo. 2017. The search engine that doesn't track you. (2017). https://duckduckgo.com
15. Steven Englehardt and Arvind Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of ACM CCS 2016*.
16. Ghostery. 2017. Ghostery detects and blocks tracking technologies to speed up page loads, eliminate clutter, and protect your data. (2017). https://www.ghostery.com
17. Leo A Goodman. 1961. Snowball sampling. *The annals of mathematical statistics* (1961), 148–170.
18. Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. 2012. Why Johnny can't opt out: a usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 589–598.
19. Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. 2013. What matters to users?: factors that affect users' willingness to share information with online advertisers. In *Proceedings of the ninth symposium on usable privacy and security*. ACM, 7.
20. Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016.. In *USENIX Security Symposium*.
21. Jonathan R Mayer and John C Mitchell. 2012. Third-party web tracking: Policy and technology. In *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 413–427.
22. Aleecia McDonald and Lorrie Faith Cranor. 2010. Beliefs and behaviors: Internet users' understanding of behavioral advertising. (2010).
23. William Melicher, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu, and Pedro Giovanni Leon. 2016. (Do Not) Track Me Sometimes: UsersâĂŹ Contextual Preferences for Web Tracking. *Proceedings on Privacy Enhancing Technologies* 2016, 2 (2016), 135–154.
24. Saurabh Panjwani, Nisheeth Shrivastava, Saurabh Shukla, and Sharad Jaiswal. 2013. Understanding the privacy-personalization dilemma for web search: a user perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3427–3430.
25. Adblock Plus. 2017. Surf the web without annoying ads! (2017). https://adblockplus.org
26. Request Policy. 2017. RequestPolicy is an extension for Mozilla browsers that increases your browsing privacy, security, and speed by giving you control over cross-site requests on pages you visit. (2017). https://requestpolicycontinued.github.io
27. Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 12–12.

28. Roland T Rust, PK Kannan, and Na Peng. 2002. The customer economics of Internet privacy. *Journal of the Academy of Marketing Science* 30, 4 (2002), 455–464.

29. Iskander Sanchez-Rola, Xabier Ugarte-Pedrero, Igor Santos, and Pablo G Bringas. 2017. The web is watching you: A comprehensive review of web-tracking techniques and countermeasures. *Logic Journal of the IGPL* 25, 1 (2017), 18–29.

30. Elliot Tan, Huichuan Xia, Cheng Ji, Ritu Virendra Joshi, and Yun Huang. 2015. Designing a Mobile Crowdsourcing System for Campus Safety. *iConference 2015 Proceedings* (2015).

31. Stefano Traverso, Martino Trevisan, Leonardo Giannantoni, Marco Mellia, and Hassan Metwalley. 2017. Benchmark and comparison of tracker-blockers: Should you trust them?. In *Network Traffic Measurement and Analysis Conference (TMA), 2017*. IEEE, 1–9.

32. Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy. 2009. Americans reject tailored advertising and three activities that enable it. (2009).

33. Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *proceedings of the eighth symposium on usable privacy and security*. ACM, 4.

34. Yang Wang, Huichuan Xia, and Yun Huang. 2016. Examining American and Chinese Internet Users' Contextual Privacy Preferences of Behavioral Advertising. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 539–552.

35. Craig E Wills and Doruk C Uzunoglu. 2016. What Ad Blockers Are (and Are Not) Doing. In *Hot Topics in Web Systems and Technologies (HotWeb), 2016 Fourth IEEE Workshop on*. IEEE, 72–77.

36. Yaxing Yao, Davide Lo Re, and Yang Wang. 2017. Folk Models of Online Behavioral Advertising.. In *CSCW*. 1957–1969.