Who's Most Targeted and Does My New Adblocker Really Help: A Profile-based Evaluation of Personalized Advertising (extended)

Sofia Bertmar Johanna Gerhardsen Alice Ekblad Anna Höglund Julia Mineur
Isabell Öknegård Enavall Minh-Ha Le Niklas Carlsson
Linköping University
Sweden

ABSTRACT

The success stories of targeted and personalized advertisements can be intimidating and offend some. One popular way to reduce the exposure to such targeting is to use adblockers and other privacy enhancing browser extensions. While there are a lot of works studying the effectiveness of adblockers, there is very limited prior work studying how the personalization experienced by different users is impacted by the use of these technologies, geographic location, the user's persona, or what browser they use. To address this void, this paper presents a novel profile-based evaluation of the personalization experienced by carefully crafted user profiles. Our evaluation framework impersonates different users and captures how the personalization changes over time, how it changes when adding or removing an extension, and perhaps most importantly how the results differ depending on the profile's persona (e.g., interest, occupation, age, gender, etc.), geographic location (US East, US West, UK), what browser extension they use (none, AdBlock, AdBlock Plus, Ghostery and CatBlock), what browser they use (Chrome, Firefox), and whether they are logged in to their Google account or not. While the extensions reduce the number of ads that a user is exposed to, we found that three out of the four extensions let through a significantly bigger portion of personalized ads. The geographic region had a big impact on the number of ads and the ad campaigns that users were exposed to, whereas the choice of browser or whether a user was logged in did not. The level of personalization ramped up quickly at the start of our 21-day measurement campaign and we observe significant differences between the level of personalization achieved for personas with different interests. By comparing and contrasting these differences we provide insights that help explain why some user groups may feel more targeted than others and why some people may feel even more targeted after having turned on their adblocker.

A preliminary version of this paper appears as a 4-page workshop paper at [4].

Submitted, Date, year, Address

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

KEYWORDS

Profile-based analysis, Personalized ads, Targeted advertisement, Privacy, Browser extension

ACM Reference Format:
Sofia Bertmar Johanna Gerhardsen Alice Ekblad Anna Höglund Julia Mineur, Isabell Öknegård Enavall Minh-Ha Le Niklas Carlsson. 2021. Who's Most Targeted and Does My New Adblocker Really Help: A Profile-based Evaluation of Personalized Advertising (extended). In *Submitted, Date, year, Address.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The online advertising market is expected to top 400 billion US dollars in 2021 [39]. Given that companies want the most out of their advertising spendings, it is perhaps not surprising that today's users are highly tracked and that significant efforts are being made to use this information to build online profiles of these users and to present personalized and targeted ads to potential consumers [29].

With some third-party tracking services having very good coverage of the web [5, 37] and state-of-the-art machine learning techniques continually improving, it is also not surprising that we start to hear an increasing amount of eerie stories of when targeted ads have been so successful in their targeting that they have creeped out and scared away the intended customers. Although marketers can follow certain guidelines to avoid too much backlash [27], it is still clear that personalized advertisements easily can overstep people's personal boundaries.

While some people have argued that the exposure of such ads is a price that users must pay to receive free content/service (e.g., [36]), others have buckled down and developed adblockers and other privacy enhancing browser extensions [3, 17, 25, 26]. Such extensions typically attempt to block third-party trackers, advertisements, or even replace the advertisements with something else (e.g., an image of a cat [1]). Although the use of these services has their own privacy and security risks [12, 15], adblockers and other privacy enhancing browser extensions have become a popular way to reduce the number of ads that a user is exposed to.

This has prompted several papers to study the effectiveness of adblockers [21] and their performance tradeoffs [12]. However, to the best of our knowledge, no prior work has studied the personalization experienced by realistic users selecting to use these technologies and how the level of personalization is affected by various other factors. In this work, we present a novel profile-based evaluation of the personalization experienced by carefully crafted user profiles that aim to surf the web in a similar fashion as many modern web users. We next summarize our key contributions.

First, we develop a Selenium-based data-collection tool that emulates user sessions of specific personas. The tool implements a basic user model and is driven by search terms associated with a set of fictive personas that we also developed for the project. While making search queries and browsing the linked results, the tool simultaneously gathers screenshots, scrapes HTML code, and stores away other information about each website visit.

Second, using the tool, we develop and implement an experimental design that allows us to evaluate the impact that many different factors have on the personalization perceived by six (or in some cases three) carefully handcrafted user profiles, while controlling for the other factors of consideration. Here, we study how the personalization changes over time for users with different persona (e.g., interest, occupation, age, gender, etc.), how the results differ depending on the profile's geographic location (US East, US West, UK), what browser extension they use (none, AdBlock [6], AdBlock Plus [35], Ghostery [2], CatBlock [1]), what browser they use (Chrome, Firefox), and whether they are logged in to their Google account or not. To ensure clean results, we run each profile inside a separate virtual machine (VM) given a unique public IP address and maintain state information throughout a measurement campaign.

Third, we performed a longitudinal measurement campaign for 21 days. Given the above experimental design, we ran 51 VMs in parallel for 21 days, where we used the first 14 days as our main experiments (described above) and used the last seven days to study the effects of adding or removing an extension. Here, we uninstalled the extension of interest after 14 days from the VMs that originally were running an extension, and at the same time installed an extension on the VMs that originally started without one. Throughout the campaign, we also extract the Google profiles of the users that were logged into their Google accounts during the tests. This gives us a natural reference point of how much personalization can be expected at each point in time. In total, the users visited 178,650 websites, for which we collected 230,175 screenshots, including an estimated 115,000 ads. (For this study, we manually identified and labeled the ads for seven out of the 21 days.) Tools and datasets will be made available with the paper.

Finally, we present analysis, report findings, and share key insights. Here, we summarize some of these findings:

- We observe significant differences between the level of personalization achieved for different profiles, with those most targeted often having interests associated with topics where we expect higher advertisement spending (e.g., fashion, travel, and electronics). The female personas experience almost twice as much targeted ad content as the male personas, and the most targeted personas are in the early-to-middle part of their careers (age groups 25-55).
- All personas also experience targeting based on their private life features (e.g., occupation, relationship status, or whether they have kids). However, this targeting represents a smaller fraction than the interest-based personalization.
- The level of personalization quickly ramped up, showing the effectiveness of the underlying tracking that the advertisement companies achieve. Furthermore, despite the logged in persona's Google profiles having identified most characteristics within two days, there are no significant differences in the level of targeting depending on whether the users are logged in to their Google accounts or not.
- While the use of extensions helped reduce the number of ads that users were exposed to, for three out of the four extensions, the ads that slipped through the filters provided significantly higher average ad personalization than what was experienced by users not using an extension. These results show that the filters are better at reducing more generic ads than personalized ads.
- We observe that companies appear to successfully build user profiles also during the time a privacy enhancing extension is used. For example, the targeting on day 15 when having an extension active during primary collection (days 1-14) and then inactivating it after day 14 results in more personalized ads than was seen by the corresponding user (without any extension running) on day 1. This shows that the tested extensions are not capable of fully stopping third-party tracking, despite some claiming so.
- In terms of volume, the UK users only see 32% of the number of ads seen by the corresponding US users. However, in terms of the level of personalization there are only small geographic differences, with UK-based users seeing somewhat more personalization. We do not observe any significant differences between users in the eastern or western US.
- During the study we observed several ad campaigns that were shown to a majority of the profiles, regardless of location and persona. Given the year that has been, it is perhaps not surprising that most of the identified campaigns centered around societal issues such as the coronavirus. In general, most of the campaigns were both US-focused and shown to US-based users to a higher extent. In fact, even the campaign most frequently seen by the UK users was for a US-based organization ("Feeding America").
- Finally, we note that our conclusions appear consistent regardless of whether the users used Chrome or Firefox. Instead, the users observed similar levels of targeting with the two competing browsers. While there are variations, the personas that observe the highest/lowest personalization with one of the browsers also observe the highest/lowest if using the other browser. Perhaps the biggest difference was that we observed some differences in the number of ads observed and the effectiveness of the adblockers when using the two browsers. However, in terms of the relative personalization seen with/without the different ad blockers, the results are still similar, regardless of which browser is used.

Outline: Section 2 describes the design of the personas and the framework used to emulate their browsing behavior. Section 3 presents our experimental design, the data collection, as well as the manual identification and labeling of ads. The analysis is presented in Section 4, focusing on one aspect at a time. Related work is presented in Section 5, before Section 6 presents our conclusions.

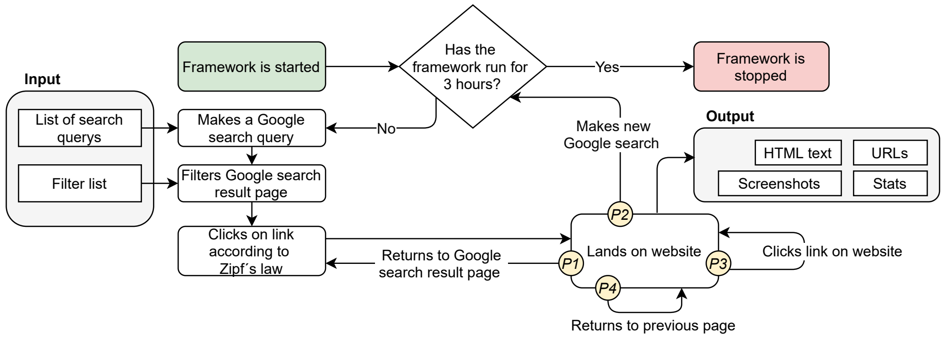


Figure 1: An overview of the framework design.

2 PROFILE CREATION

Profiles with different personas are central in our evaluation design. We next describe how the personas were designed (§ 2.1) and their user behavior was emulated (§ 2.2).

Table 1: Summary of the six different personas.

Tag	Gender	Age	Interest	Occupation	Civil Status	Kids
A	Female	18	Horses, celebrities, gardening	High School	Single	No
B	Female	33	Hair, fashion, DIY	Hairdresser	Single	No
C	Female	51	Movies, stock trading, interior design	Bank worker	Married	Yes
D	Male	21	American football, baseball, medicine	University	Relationship	No
E	Male	37	Cooking, traveling, electronics	History teacher	Divorced	Yes
F	Male	68	Birds, baking, crossword puzzles	Retired	Married	Yes

2.1 Persona design

Characteristics. Six different personas have been created for the purpose of imitating online user behavior. Every persona has been assigned various characteristics: name, gender, age, three main interests, occupation, civil status, and parental status. Table 1 summarizes the key characteristics selected for each of the six personas. The characteristics have been selected to be of relevance to the Google's Ad personalization page, this page's categories, and target aspects that browsers are expected to find relevant when categorizing users and their interests based on their online activity.

Search queries. For each persona, we created 200 search phrases based on their individual key characteristics. The specific search phrases were created by a group of six people, each responsible for one persona. This choice was motivated by each person having their own style of writing search queries. 80% of the search queries were based on the interest categories and the remaining 20% were based on other personal information (e.g., whether they had children, were single, etc.). Furthermore, all personas use English as primary language and (as outlined in Section 3) all profiles were located in English-speaking countries.

Stereotypes. To simplify the data collection and the interpretation of the results, the personas were created to be fairly "stereotypical" in the sense that their interests tend to be commonly linked to their age and gender. These stereotypes are based on our own biases. However, other personality traits such as religion, sexuality, or race have not been included in the design process. Finally, the personas were assigned common names that are widely applicable: (A) Mary Johnson, (B) Jennifer Brown, (C) Patricia Jones, (D) James Davis, (E) John Anderson, (F) Robert Smith.

Google accounts. Six Google accounts were created to match the six personas. Several aspects of persona characteristics had to be revealed to Google when creating these accounts, including the name, birth date, and location. Only six out of 51 VMs utilized these profiles and hence also Google's log in function.

2.2 Framework design

We designed and implemented a Selenium-based framework that emulates the user behavior of a persona and collects a corresponding dataset. Figure 1 presents an overview of the framework. As input, the framework takes the search queries associated with one of the previously created personas. During a user session, the framework then considers one query at a time. The list of search phrases is shuffled every time the framework starts running.

Weighted clickthrough. After making a search query, the framework first filters the Google search result using a list of filter words. This is done to remove irrelevant, commonly appearing links from the search results (e.g., settings, tools, and feedback). Thereafter, one of the available links is clicked, with the links being selected according to a Zipf-distributed probability distribution, meaning that the probability of requesting the n^{th} listed page of a search response is proportional to $1/n$. The choice to use a Zipf distribution was motivated by Zipf-like distributions having been found good to model the relative frequency that users request webpages listed as a result of a search query [13] and the corresponding links' clickthrough rates [20]. These highly skewed distributions capture that the top results for a query see by far the most clicks.

Post-search behavior. After landing on a new webpage, there are four instances that can occur with different probabilities. These instances and their probabilities (*Px*) are explained next.

- Backward-to-search (*P1*): User returns to the original Google results page.
- Forward-to-search (*P2*): User makes a new search query.
- Forward-to-browse (*P3*): User clicks link on the current page.
- Backward-to-browse (*P4*): User returns to the previous page.

Motivated by research by White and Drucker [40], we set these probabilities to *P1*=0.08, *P2*=0.21, *P3*=0.5 and *P4*=0.21.

Think time. At each step, a think time is implemented. This corresponds to the time that the user spends doing some form of navigation, enter data, choosing which link to click, or simply to read a webpage. We randomly chose think times from a smaller set of pre-defined think times and their selection probabilities. Motivated by the median think times of 28.653 sec. observed by Ramakrishnan et al. [38], we defined this set and their selection probabilities such that the think times have a median of 28 sec., they follow a skewed probability distribution, and yet the think times are upper bounded. Specifically, we chose the think times to be 14 sec. with probability 0.45, 28 sec. with probability 0.35, and 56 sec. with probability 0.2.

Session durations and exception handling. A session is intended to run for three hours. If three hours have passed, the emulation is stopped before making a new search. For the sake of simplicity, the framework accepts all pop-ups before continuing with additional tasks. If the framework runs into faulty pages (e.g., pages that do not load, HTML text that cannot be saved, buttons that we cannot click, or unknown exceptions), the recovery mechanism is to simply return to the first-level Google search page and continue with the next search query. In the exceptional case that we run out of search phrases during a session, the framework is setup to simply repeat queries as necessary to ensure that a session reaches its full three-hour duration. However, this was very rare. In most cases, approximately hundred search queries were performed.

Browser specific implementation differences. For Chrome sessions, ChromeDriver is utilized, whereas Firefox sessions use GeckoDriver. There were also some subtle but important browser-dependent differences that needed some care when opening a browser window. First, note that when you open a Chrome window in Selenium it is originally in an automated testing environment. If done this way, cookies and history would therefore not have been saved. For this reason, before the start of every Chrome session, a Chrome debugging window was instead opened manually through the terminal. This ensured that we could maintain consistent cookies and browser history between sessions. When using Firefox, the browser is opened directly through the Selenium framework and cookies are loaded from and saved to a pickle file between sessions to maintain consistency.

3 DATA COLLECTION

3.1 High-level experimental design

The experiments were split into two collection phases: the primary and secondary phase. The two phases lasted for 14 and 7 days, respectively, resulting in a total data collection period of 21 days. The data collection took place between 2021-04-24 to 2021-05-14.

Primary phase. Using the six personas described in Section 2.1, we created a larger set of user profiles. Each profile is determined by the persona (personas A-F), the geographic location of the user (US East, US West, UK), what browser extension the user used (none, AdBlock [6], AdBlock Plus [35], Ghostery [2], CatBlock [1]), what browser the user used (Chrome, Firefox) and whether the user was logged in to the their Google account or not. To ensure clean results, we run each profile inside a separate virtual machine (VM) given a unique public IP address. Furthermore, to keep the cloud computing costs manageable for a long-term experiment with many public IP addresses and since it is complicated to create more than a handful of Google profiles, we decided to not run a full factor experiment ($6 \times 3 \times 5 \times 2 \times 2 = 360$ combinations). Instead, we selected to consider one factor at a time, always including the three base profiles based on personas A, C, and E, which we located in eastern US, and that used Chrome without any extensions. First, to better understand differences between profiles we include results also for personas B, D, and E, which we placed (as a base case) in the UK. Second, we included this full set of profiles also when considering the impact of being logged in or not. Third and fourth, when studying the impact of location and the use of extension, we simply changed the location of the base profiles to the other two locations of considerations or applied one of the four extensions, respectively. Fifth, we repeated all tests comparing the impact of using different personas as well as all experiments with different extensions when using Firefox instead of Chrome. This was made possible since all four extensions worked for both Chrome and Firefox. Finally, we included three additional experiments with profiles B, D, and E. These were included to provide symmetric baseline results for each profile, when combining the results from the two phases of the data collection. Table 2 summarizes the set of VMs (with VMs number as per our internal numbering used in the datasets) and the secondary phase is described next.

Table 2: The virtual machines with their set ups.

VM ID	Browser	Login	Persona	Region	Primary	Second
1,3,5	Chrome	No	A,C,E	East US	None	Ext.
2,4,6	Chrome	No	B,D,F	UK	None	Ext.
7,9,11	Chrome	Yes	A,C,E	East US	None	Ext.
8,10,12	Chrome	Yes	B,D,F	UK	None	Ext.
19-21	Chrome	No	A,C,E	UK	None	Ext.
34-36	Chrome	No	A,C,E	West US	None	Ext.
22-24	Chrome	No	A,C,E	East US	AdBlock	None
25-27	Chrome	No	A,C,E	East US	AdBlock Plus	None
28-30	Chrome	No	A,C,E	East US	Ghostery	None
31-33	Chrome	No	A,C,E	East US	CatBlock	None
13,15,17	Firefox	No	A,C,E	East US	None	Ext.
14,16,18	Firefox	No	B,D,E	UK	None	Ext.
40-42	Firefox	No	A, C, E	East US	AdBlock	None
43-45	Firefox	No	A, C, E	East US	AdBlock Plus	None
46-48	Firefox	No	A, C, E	East US	Ghostery	None
49-51	Firefox	No	A, C, E	East US	CatBlock	None
37	Chrome	No	B	UK	AdBlock Plus	None
38	Chrome	No	D	UK	AdBlock Plus	None
39	Chrome	No	F	UK	CatBlock	None

Secondary phase. At the start of the secondary phase, we “flipped" the extension-related configurations used by each fictive user. In particular, the profiles that used extensions during the primary collection were instead run with these uninstalled during the secondary phase. In contrast, the profiles that did not use any extension during the primary collection, added an extension for the secondary phase. Since we avoided having redundant VMs during the primary phase (which is the focus of most of our analysis!), we could not achieve the same level of coverage for each extension during the second phase. To allow some fair head-to-head comparisons also here, we therefore opted to tie the choice of extensions being added based on the persona used in each experiment. For VMs adding an extension for the secondary phase, we used the following assignments: AdBlock was added to VMs of personas A+D, AdBlock Plus to VMs of personas B+E, Ghostery to VMs of persona C, and CatBlock to VMs of persona F. Finally, as noted above, to allow symmetry in the profile-based experiments looking at the impact of adding or removing extensions, we include VMs 37-39. These VMs ensure that we have a corresponding case where an extension is removed from a VM when there exists an experiment where that extension has been added to the same persona, location, and browser. (Again, we opted to not design for the opposite to hold, since this would require redundant experiments during the primary phase, which is the primary focus of our study.)

System setup and collection details. VMs were created to ensure sandboxed environments and unique, public IP addresses for each VM. Using Microsoft Azure, we setup 51 VMs, one for each profile. The operating system used during the data collection was Microsoft Windows 10. The VMs had 2 vCPUs, 4 GiB memory, and a standard SSD. For the logged in profiles, the users remained logged into the persona's Google account for the entire duration of the data collection. In total, 33 VMs were placed in US East (Washington D.C.), 15 VMs were placed in UK South (London), and three VMs in US West (San Jose). Each VM was adjusted to the corresponding time zone of their location. Furthermore, every session was launched at 04:00 PM (GMT+2) ± 2 hours, each day. As a result of time zone differences, the daily data collection began at 09:00 AM ± 2 hours local time in US East, 06:00 AM ± 2 hours local time in US West, and 03:00 PM ± 2 hours local time in the UK.

3.2 Dataset creation and labeling

Logged data. During each session, log files were created containing information about the visited URLs, extracted HTML texts, and screenshots of the visited webpages. For each session, we also saved summary statistics about the number of clicked links and screenshots, for example, as well as information about potential exceptions. A scroll function was implemented to ensure that multiple screenshots were taken of the visited pages. While the intention with the scroll function is to capture all ads on a page, the implementation was not perfect and some pages allow almost endless scrolling. Our scroll function may therefore have missed some ads, making the observed ads a lower bound of the actual number of ads on a page. Other obstacles such as pop-ups also resulted in some ads being difficult to assess. We do not expect these limitations to impact our conclusions, as our analysis primarily makes relative comparisons. Finally, for the personas that were logged in to their Google account, we took manual screenshots of their Google Ad Personalization page at the end of each session.

Identifying ads. To identify ads, the screenshots were manually evaluated. This process was divided into three steps: (1) The placement of the image suggested that it may be an ad. Frequent ad placements include headers, banners, and side bars. Here, we considered all these placements. (2) The image contained the word "Ads", "Ad", or "Advertisement". (3) The image contained labels of ad choices and/or an option of a closing checkbox. At this point, we note that determining what is an ad compared to what is simply a display on a webpage can be difficult. While we acknowledge that manual labeling can be an error source, we expect that our manual labeling is more accurate than an automated process would have been (including compared to extensions studied in this paper).

Categorization of ads. We next labeled each identified ad. Here, we based our classification on the combined set of 18 interest categories of the six personas. Ads related to other persona traits, such as marital or parental status, were placed in the category "private life" of each persona. Ads that did not meet any of the above categories were labeled as "other". Similar to the identification of ads, we acknowledge that our manual labeling may introduce some errors. At this time, we also note that the interests of some personas are significantly narrower than others.

Finally, we looked closer at ads which we considered part of a general ad campaign. While all ads are in some way part of an ad campaign, the campaigns that we categorized as "general ad campaigns" were not related to a persona's interests. Instead, they were ads that were observed by a majority of personas and across the geographic locations. In most cases these were found to be targeting society as a whole rather than an individual. This typically allowed us to easily label them based on a "common topic".

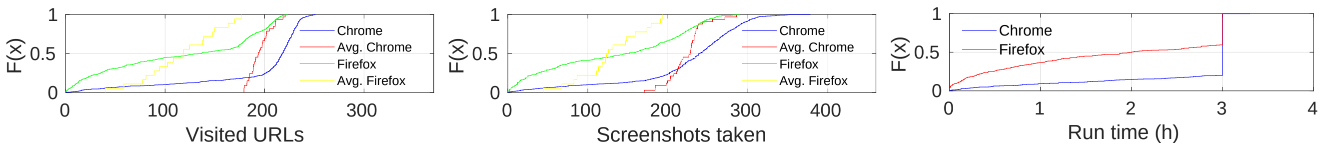
3.3 Summary statistics and limitations

In total, we collected data for 1,071 sessions (51 VMs × 21 days). This resulted in a dataset with 178,650 website visits and 203,175 screenshots. Given the manual effort required to manually annotate and label all the screenshots, we selected to down sample the number of screenshots that we analyzed. For the most part, we therefore focused only on screenshots from days 1, 7, 14, 15, and 21. However, for some VMs, we also did some additional labeling.

When discussing the number of page visits and screenshots recorded, it is important to note that our framework is not intended to be production ready and that Selenium frequently crashes. While we have built recovery mechanisms into the framework, our implementation for Chrome was more successful than for Firefox. This resulted in some noticeable differences in the number of page visits and screenshots during a typical user session with the two browsers. For example, the median number of websites visited during a Chrome session was 218 compared to 136 for Firefox. The empirical Cumulative Distribution Functions (CDFs) of the (a) number of webpages visited, (b) collected screenshots, and (c) time until a session crashes are shown in Figure 2. While the differences are substantial, most of the 1,071 sessions (regardless of browser) have a significant number of page visits, and over a 14-day period (with relatively random crashes) we complete enough page visits to accurately build up comparable profiles. For example, note that the average number of page visits per VM as calculated over a 14-day period varies substantially less than the per-session samples. This is especially the case for Chrome VMs.

The main reason for the differences between the crash rates for our Chrome and Firefox experiments may be that we originally designed the framework for Chrome and then modified it to be used also with Firefox. For the above reasons, our analysis (and our data collection) primarily focuses on the Chrome-based measurements, and we only use the Firefox datasets together with the Chrome datasets and/or to validate our conclusions.

At this time, it is also important to note that we did not select the three-hour session durations to mimic real users. Instead, we selected these session durations to ensure that each VM saw enough page visits to allow trackers and ad providers to build up a profile of the user over the measurement period, and to ensure that we get clear daily snapshots at roughly the same time each day.



(a) Number of page visits/session and average per VM (b) Number of screenshots/session and average per VM (c) Session duration

Figure 2: Summary comparisons of per-session statistics when using Chrome and Firefox.

Finally, and perhaps most importantly, the Firefox dataset appears to have resulted in enough visits to confirm that the conclusions drawn for Chrome also hold for Firefox. Whatever potential bias that the extra Firefox crashes may have resulted in does not appear to have impacted the relative differences observed when comparing the relative personalization experiences by the different profiles and/or the impact that extensions have on personalization.

3.4 Extension discussion

Before presenting our analysis, this section provides a brief overview of the privacy-focused extensions used in the study: AdBlock [6], AdBlock Plus [35], Ghostery [2], and CatBlock [1]. All four extensions are available for both Chrome and Firefox.

Both Adblock and AdBlock Plus are widely popular, free, open-source browser extensions designed to filter content and block ads. To do so, they both (per default) use a blocklist (e.g., EasyList [18]) to determine whether the content should be blocked or not. Furthermore, they use a list of “acceptable ads" that they use to let through ads by accepted advertisers.

In addition to adblocking, Ghostery aims to uphold user privacy by preventing third-party tracking. Ghostery also analyzes the active third-party trackers, the general website performance, and gives users an overview of who is tracking them and where [33].

CatBlock is by far the least popular extension studied here. While the other extensions have relatively higher penetration rates, CatBlock only had 5,000+ Chrome users and 170+ Firefox users at the time of the study. (Numbers from Chrome web store [24] and Firefox browser add-ons [34].) CatBlock also differs in its functionality. In particular, CatBlock replaces online ads with images of cats. The extension was developed by Peckett and Taro starting with existing code from AdBlock (used to identify the ads on a webpage), suggesting that many of its properties are inherited from AdBlock.

4 ANALYSIS

We next present our analysis. Section 4.1 builds a basic baseline using the Google Ad Profiles and analyzes whether being logged in or not affects the ads displayed to a user. Section 4.2 takes a closer look at the differences in the level of targeting observed by different personas, including how the differences depend on their interests and personal life characteristics. Section 4.3 analyzes how the personalization of ads is affected by using different privacy enhancing extensions. Section 4.4 evaluates regional differences in ad targeting. In Section 4.5, we take a closer look at general ad campaigns observed during our study. Finally, Section 4.6 presents a browser-based comparison and demonstrates that our derived conclusions hold for both Chrome and Firefox.

Table 3: Summary when the persona characteristics are added to the Google Ad Personalization page.

	Persona					
	A	**B**	**C**	**D**	**E**	**F**
Interests	Day 2	Day 4	Day 2	Day 4	Day 2	Day 2
Occupation	-	-	Day 2	Day 2	-	Day 6
Parental status	Day 2	Day 2	Day 3	Day 2	-	-
Relationship	Day 8	-	Day 2	-	-	Day 2
Age	Day 2	Day 2	Day 2	Day 2	Day 2	Day 2
Gender	Day 2	Day 2	Day 2	Day 2	Day 2	Day 2
Match	83.3%	66.7%	100%	83.3%	50.0%	83.3%

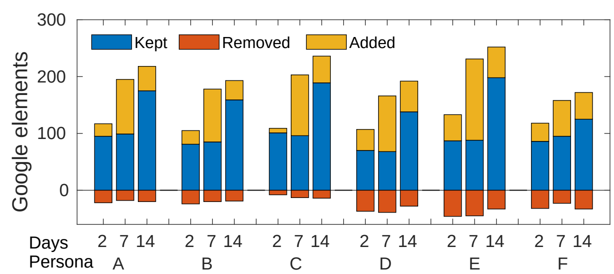


Figure 3: Summary of the changes in the profile elements for logged in users (personas A-F).

4.1 Profile-based baseline comparison

Before studying the level of personalization experienced by each persona, let us first establish a baseline of how quickly a browser may learn the characteristics of each persona.

Google profile vs. persona profile. Here, we used the six VMs (7-12) in which each persona stayed logged in to their Google account for the duration of the experiments. In particular, we studied the elements in the Google profile (i.e., the Google Ad personalization page) of each user and how quickly the personas’ characteristics were learned.

Although some personal attributes are learned by Google via the account creation process (e.g., gender and age), most attributes were learned during the actual measurement campaign. However, also here, we found that Google managed to build a well-matching profile fairly quickly. Table 3 shows what day in the campaign that each type of information was learned. For example, at the latest, Google displays all persona interest categories by day 4. The categories that Google struggled the most with (failed in 50% of the cases) were the occupation (A, B, E) and relationship status (B, D, E). Also parental status proved difficult for two profiles (E, F). Otherwise, almost all characteristics were identified within 3 days, or within 8 days at the latest (relationship status of A). Since we only allocated 20% of the search queries of each profile to capture the non-interest-based characteristics, it is perhaps not surprising that these aspects were harder to identify.

Google elements fluctuation. We use Figure 3 to illustrate how both the number of elements of the Google profile and the churn in the number of such elements differ between the personas. While we see noticeable churn in which elements Google use to characterize a profile throughout the measurement campaign, the relative increase in the number of elements (i.e., added minus removed) slows down over time. Here, we show the changes for day 2, from day 2 to 7, and from day 7 to 14. Note that by day 7 the Google profiles are fairly mature. Beyond this point, the number of added elements (yellow) is only slightly more than the number of removed elements (red) and the total elements (blue+yellow) at days 7 and 14 are relatively similar.

While the differences are small between personas, we note slightly fewer elements for the UK-based personas (B, D, F) and slightly higher churn (e.g., more removed) for males (D, E, F) than females (A, B, C). This may suggest somewhat less knowledge and somewhat more uncertainty of the categorization of these two groups. Since personas D and F represent the intersection of these two groups, we may expect these two profiles to be the hardest to accurately target. As seen in later sections, this is indeed the case.

One reason for the big differences is the number of elements associated with the interest categories of the different personas. Out of the 18 interest categories considered, "Fashion", "Traveling", and "Interior" generate the most Google elements, whereas "Celebrities", "Baseball", and "American Football" generate the least. Having said that, we also kept track of an "Other" category where we placed all elements that did not have any obvious connection to each persona's characteristics. This category saw the biggest churn in what elements were included (both additions and removals) and had the most elements. It appears that the profiles reacted fairly quickly to contents that the users were in contact with. The higher churn of this category is therefore not surprising, since contents related to various topics that make up the "Other" category typically were not revisited as frequently as topics related to the core interests.

Similar differences also show up in the correctness of the profiles. For example, when comparing the match between Google's profiles and the persona characteristics (Table 3), we observed variations between 50.0% and 100.0%, though most experienced a match of at least 83.3%. This shows that Google typically is highly accurate at profiling the six personas, at least when logged in.

Targeting when logged in vs not logged in. To examine differences in the level of personalization experienced by Chrome users logged in versus not logged in, we compared VMs 1-6 and 7-12 over days 6, 7, 13, and 14. As seen in Figure 4 there are no significant differences in the level of personalization experience by the users. In the case of personas D and F, these users consistently see low levels of targeted content. This matches our observation above regarding Google perhaps having the toughest time profiling these personas. The narrowness of these personas' interest categories is the most likely reason for this. It does not help that we decided to place persona D in UK, as his sports interests (baseball and American football) are much less popular in the UK than these sports are in the US. Companies targeting people with these interests may therefore invest less advertising money in the UK.

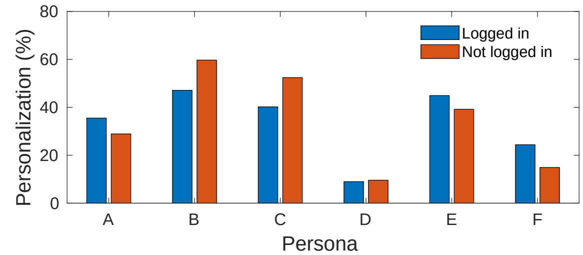


Figure 4: Differences in targeting of ads for logged in and not logged in users. (Personas A-F.)

4.2 Detailed profile comparison

We next look closer at the differences in the level of personalization observed for the personas, including how the targeting depends on the interest categories and personal details of each persona.

Persona targeting. Table 4 shows combined results for days 1, 7, 14 for the VMs for which the user did not use an extension and was not logged in (i.e., VMs 1-6, 13-18). Here, we show the fraction of ads observed by each persona (row in the table) that were associated with one of the – in some cases narrow – interest categories of the six personas (columns) as well as a "private life" category based on each user's other characteristics. For easier readability, we omit the "other" category from this table.

The main observation here is that there is a clear diagonal correlation of ads being related to the interest categories of each persona, resulting in a stronger red color in these cells. However, we also note that several of the columns in Table 4 are mostly white. For these interest categories, none of our profiles observed a noticeable fraction of ads. This may suggest that these categories (e.g., "Birds" and "DIY") see limited advertising budgets or the profiles that match may be located in a region where advertising for that product (e.g.,"American football" and "Baseball" ads in the UK) may be limited. Yet, in all these cases the profile with the persona that best matched this interest always saw the most ads for this category. In fact, for all 18 interests (columns), the profile with a persona (row) that match that interest always was the profile that saw the significantly largest fraction of ads for that interest category. This clearly shows that even for narrow interests, personalized advertising is successful with regard to which personas they are presented.

In general, narrower interest categories exhibited fewer targeted ads. In contrast, we see many targeted ads for the interest categories "Fashion", "Traveling" and "Electronics". We also checked that this was not due to search biases. For example, when comparing word counts from the html texts (VMs 1-6, 13-18; days 1-14), we observed no significant correlation between personalized advertising and the collected HTML-text. Instead, we observed roughly as many extracted words for the least observed ad categories as the most observed ad categories (based on interest). We also observe significant targeting of the "private life" characteristics of each user.

The level of personalization is also impacted by factors such as age and gender. For example, Figure 5 shows that age group 25-34 experiences most targeting (70%) and age group 65+ the least targeting (below 20%). Similarly, as seen in Figure 6, female personas (A, B, C) have been exposed to twice the amount of targeted advertising seen by the male personas (D, E, F). Female personas are also exposed to more ads related to "fashion" than male personas. In

Table 4: Heat map showing the fraction of advertising seen associated with each of the interest categories of personas A-F. We also include a personal life category for each of the six persona.

	A			B			C			D			E			F			
	Horses	Celebrities	Plants	Hair	Fashion	DIY	Movies	Stock-trading	Interior /Home	American Football	Baseball	Medicine	Cooking	Traveling	Electronics	Birds	Baking	Crosswords	Private life
A	0.17	0.04	0.12	0.00	0.10	0.00	0.03	0.01	0.03	0.00	0.00	0.01	0.02	0.02	0.01	0.00	0.00	0.00	0.04
B	0.00	0.00	0.00	0.23	0.42	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.00	0.00	0.02
C	0.00	0.00	0.02	0.00	0.03	0.00	0.13	0.20	0.17	0.00	0.00	0.01	0.02	0.03	0.03	0.00	0.00	0.00	0.06
D	0.00	0.00	0.02	0.00	0.06	0.00	0.04	0.03	0.01	0.03	0.03	0.07	0.00	0.04	0.11	0.00	0.00	0.00	0.10
E	0.00	0.00	0.03	0.00	0.07	0.00	0.03	0.03	0.01	0.00	0.00	0.03	0.20	0.18	0.07	0.00	0.00	0.00	0.02
F	0.00	0.00	0.01	0.01	0.10	0.01	0.02	0.01	0.03	0.00	0.00	0.03	0.06	0.05	0.06	0.01	0.07	0.04	0.04

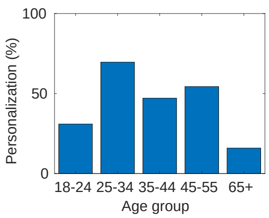


Figure 5: Personalization in the different age categories.

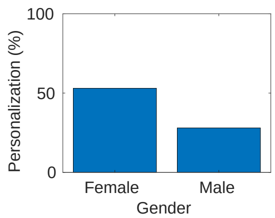


Figure 6: Personalization of two categories of gender.

contrast, male personas are exposed to almost twice the amount of “electronics" advertising than the female personas, and more than twice the amount of “traveling" advertising. This entails that gender is an attribute that is targeted to some extent, although only 2 categories of gender are considered in this paper.

4.3 AdBlock and other extensions

Extension comparisons. Here, persona A, C and E are compared, using the extensions (i) AdBlock, (ii) AdBlock Plus, (iii) Ghostery, and (iv) CatBlock. The two VMs (Chrome + Firefox) of all three profiles (A, C, E) using an extension are compared to the results for the VMs hosting the corresponding profiles without an active extension (i.e., VMs 1+13, 3+15, 5+17) using data for days 1, 7, and 14. While the use of extensions substantially reduced the number of ads seen by the users (on average by 87.1%), we have found that the fraction of personalized ads is higher than this baseline for three of the extensions. Table 5 summarizes these results. Here, we also present p-values based on one-sided two-sample binomial hypothesis testing of the null hypothesis that the fractions of personalization seen by an extension is the same as without an extension. These results show statistical significance (with 95% confidence) for all extensions. For CatBlock we see significantly lower personalization than without, and for the others (AdBlock, AdBlock Plus, and Ghostery) we see higher personalization than without.

AdBlock, AdBlock Plus, and CatBlock all use EasyList. This may explain why some “accepted" ads fall through when using the standard settings. However, it is less clear why AdBlock and AdBlock Plus do so much worse than CatBlock. The findings that CatBlock lets through substantially fewer ads than the other extensions (both when calculated per session or per screenshot) and that it also lets through the smallest fraction of personalized ads, are noteworthy

Table 5: Personalized ads with an extension compared to without. An asterisk (*) is used to indicate that the result is statistically significant with 95% confidence (or higher).

Extension	AB.	AB.+	Gho.	Cat.	None
Targeted	64.9%	52.7%	77.8%	13.1%	40.4%
Ads/session	7.72	6.05	4.56	1.67	39.1
Ads/screenshot	0.050	0.038	0.026	0.011	0.219
p-value	<0.0001*	0.0078*	<0.0001*	0.0014*	

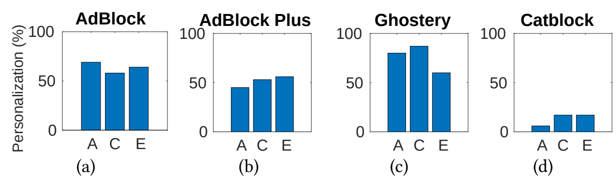


Figure 7: Example differences in the level of personalization experienced by the personas using the same extension.

since CatBlock is substantially less established than the other extensions and has much fewer users. Having used the tool, we note that the blocking (or in the case of CatBlock, replacement of ads with alternative images) can come at the cost of the users' web experience. These results may suggest that targeted ads may be harder to block, at least without deteriorating the user experience.

Another possible reason for the above differences may be that some of the ad providers that provide the most targeted advertisements are more likely to be whitelisted, as they can contribute more revenue to the extensions relative to what the owners of non-targeted ads. While it is hard to fully validate the monetary flows between organizations, this possible explanation is in line with the idea that being exposed to advertisement is a transaction crucial to receiving free content and services online [36].

Finally, we note that Ghostery [2] explicitly claims that third-party trackers are blocked when activating the extension. Yet, a majority of ads found during such sessions are considered to be targeted, suggesting that the ad companies still are able to very successfully classify and target these users. Our results clearly highlight the limited protection that the extensions provide against ad providers building accurate user profiles.

Persona-based variations when using the same extension. To gain further insights and strengthen our conclusions we next compare multiple personas using the same extension. Figure 7

Table 6: Timeline of levels of personalization during primary and secondary collection for users adding (no→ext) or removing (ext→no) an extension. Cases with active extensions are shown using shaded cells.

			Primary		Secondary	
			Day 1	**Day 14**	**Day 15**	**Day 21**
A+E	**AB.**	**ext→no**	6.60%	35.05%	35.19%	34.20%
		no→ext	39.00%	39.95%	52.00%	84.20%
B+D	**AB.+**	**ext→no**	50.00%	41.65%	34.90%	29.65%
		no→ext	11.10%	18.25%	50.00%	50.00%
C	**Gho.**	**ext→no**	85.70%	83.00%	57.50%	40.2%
		no→ext	45.60%	26.70%	50.00%	58.30%
F	**Cat.**	**ext→no**	16.60%	50.00%	7.20%	37.50%
		no→ext	0.00%	33.30%	0.00%	100%

shows the differences in the level of targeted ad content for the three base personas. When comparing different personas using the same extension, it is apparent that most experience similar levels of targeting with slight variations.

Adding or removing extensions. The secondary collection phase was implemented to glean some insights into the effects that adding or removing an extension may have on the personalization experienced by users. Table 6 summarizes the level of personalized advertisement experienced by the six personas for which we have data for both the case when an extension was removed (ext→no) and when the same extension was added (no→ext). Here, personas A+E used AdBlock, personas B+D used AdBlock Plus, persona C used Ghostery, and persona F used CatBlock. When interpreting these results, it is important to remember that personas D and F were the profiles that generally saw the least personalization and that CatBlock lets through by far the least ads.

First, as noted above, the observed ads are more targeted when using an extension than when not using an extension. This makes for interesting example cases, where the fraction of personalized ads increases when adding these privacy-oriented extensions. For example, the average of days 15+21 is higher than for day 14 for all four extensions for no→ext rows). Similarly, a client removing the extension on average see lower personalization (but more ads) on average on days 15+21 than they did on day 14 when using the extension. Second, a more subtle observation is that users removing their extensions (ext→no) on average see more ads on day 15 (just after removing the extension) than the user with the same persona but that did not use any extension during the primary phase (no→ext) did on day 1. This again shows that the ad providers have been able to learn and build useful profile information while the extension is active. While perhaps not surprising, this goes against the privacy-focused nature of these extensions and emphasizes that (at least) the extensions studied here have not been able to fully protect the users from revealing information to third parties.

Finally, we looked closer at the six logged in users that had an extension activated during the secondary collection phase. While we did not monitor these Google profiles on a daily basis during this phase (as we did days 1-14), interestingly, we found that the process of building the profiles continued also over this period. For example, comparing the Google profiles 10 days after the collection finished with the profiles on day 14, we found that on average 180 elements had been kept and 20 had been removed. While another 50 elements had been removed, we expect that this may have been due to the accounts being inactive 10 days since the collection finished. To put this in perspective, we kept these accounts inactive for another 25 days, and at that time another 130 elements had on average been removed from the profiles (and very few had been added). This result is interesting in its own right, since it may indicate that Google's ad personalization page may be relying on a fairly frequent update frequency. This was also seen in the significant daily churn observed for our daily snapshots days 1-14.

Table 7: Regional differences in the average number of ads observed and the level of personalization.

Region	Ads/session	Ads/screenshot	Targeting
UK	12.8	0.055	59.7%
US East	38.8	0.143	46.0%
US West	39.1	0.164	46.0%

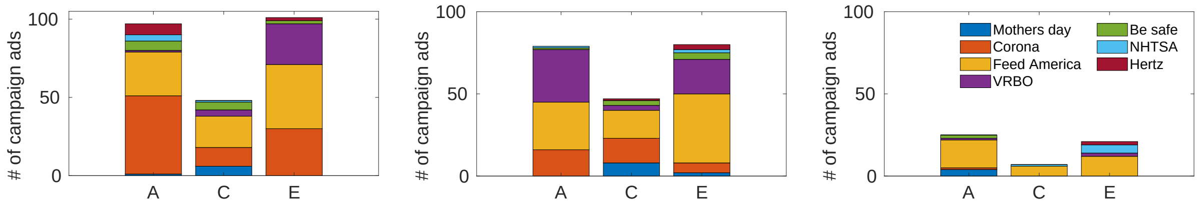
4.4 Regional comparisons

We have observed significant geographic differences in (1) how many ads a person is exposed to and (2) how much personalization is observed. For a fair head-to-head analysis, we used the profiles associated with our baseline personas (A, C, E) when located in US East, US West, and UK. This makes the greatest common set, consisting of VMs 1, 3, 5 (US East), 19, 20, 21 (UK) and 34, 35, 36 (US West). We again used the combined set of ads from days 7+14.

Table 7 summarizes these results. Although the screenshots per session were similar (± 16%) for days 7+14, we include both per-session and per-screenshot statistics. The profiles located in UK experienced approximately a third as many ads per screenshot (and ads per session) as the six US-based profiles. For example, the UK-based profiles observed on average 0.055 ads per screenshot compared to 0.143 and 0.164 ads per screenshot for the US East and US West users, respectively.

In contrast, we observe somewhat higher personalization for the UK-based profiles (59.7%) compared to the two US-based locations (both 46%). This is interesting when put in the context of the results in Section 4.2, where we observed somewhat lower personalization for the UK-based base profiles (personas B, D, F) than we observed for the US-based base profiles (personas A, C, E). Again, we believe that the lower personalization scores for these profiles (especially D and F) are due to them having very narrow interests or being located in a market where they represent a relatively small market share (e.g., American football in the UK). Similarly, we believe that the somewhat higher targeting when located in the UK may partially be situational. For example, perhaps some of the topics that are of interest to personas A, C, and E (e.g., horses) may be associated with bigger markets in the UK than in the US. The most significant geographic difference is instead that UK-based users see much fewer ads than the US users.

When discussing these results, we note that we (similar to many users) leverage Google to find pages related to the users' interests. Since Google itself takes locality into account when directing users to different webpages, we expect that the observed differences can



(a) US West (b) US East (c) UK

Figure 8: Differences in campaigns of three different regions, for persona A, C and E.

Table 8: Regional differences in the most visited TLDs.

Region	.com	.uk	.org	Other	Total
UK	70.5%	21.1%	4.20%	4.30%	12,610
US East	89.5%	0.40%	5.90%	4.10%	11,717
US West	88.3%	0.80%	7.10%	3.80%	12,593

be contributed to both differences in (1) the webpages visited and (2) the ads displayed when visiting a common set of webpages.

While we did not try to separate how much of the differences may be contributed by each aspect, we expect both aspects to have contributed to the differences. For example, while we did observe fewer ads for the UK-based profiles than the US-based profiles in the few cases we had URLs with common screenshots (4 vs 8 ads in the 13 URLs with common screenshots between UK and US East users, and 7 vs 14 ads in the 16 URLs with common screenshots shared between UK and US West users), these differences only make up a small part of the overall differences. This is in big part due to there being a very small overlap in the visited pages observed from the different regions. For example, we only observed 3 URLs for which we had screenshots for all 3 regions, and another 10+13+12 URLs for which we had matching screenshots from 2 out of 3 regions. Here, it is also important to note that we observed significant regional differences in the visited pages. For example, when looking at the top-level domains visited by the different users (Table 8) we found that UK users were significantly more likely to visit .uk domains and less likely to visit .com domains (often hosted in the US) than the corresponding US profiles.

4.5 Campaigns

Some of the differences between US and UK users were also visible when looking at general ad campaigns spanning across all regions and personas. When studying ad campaigns, we used the same nine profiles (3 personas $\times$ 3 regions) as we used for the regional analysis (§ 4.4) and note that all evaluated campaigns were reoccurring over the entire 14-day-long primary phase (2021-04-24 to 2021-05-07).

Figure 8 summarizes the results. First, note that the UK-based users (again) see much fewer campaign ads than the corresponding US-based users. It is therefore interesting to see that several US-targeted ad campaigns still are visible in the UK. In fact, for the UK-based users, the dominating ad campaign (responsible for most campaign ads) is the US-oriented ad campaign for “Feeding America”. The UK-based users A and E were also targeted with Mother’s Day ads to a similar or larger extent than their US counterparts, despite UK’s Mother’s Day (March 14, 2021) happening well before the measurements, whereas the US Mother’s Day (May 9, 2021) occurred two days after the primary collection phase completed.

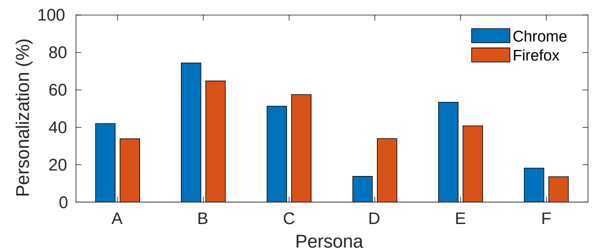


Figure 9: Comparison of the personalization for all personas when using Chrome and Firefox.

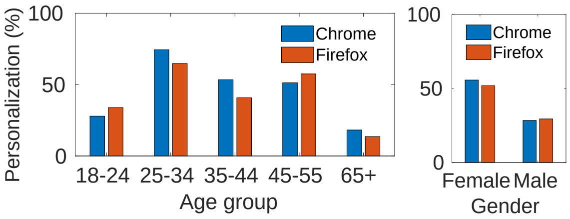
Table 9: Ads/screenshot when using different browsers.

	A	B	C	D	E	F
Chrome	0.20	0.12	0.15	0.10	0.17	0.24
Firefox	0.28	0.16	0.32	0.11	0.21	0.26

One persona that stood out was persona C (Patrica Jones). She consistently saw the least ad campaigns, but still saw noticeably more Mother’s Day ads when she was located in the US (where Mother’s Day was about to take place in a few days) than the other profiles. This was interesting to us since she is the only mother in our persona set, suggesting that the Mother’s Day campaign may have (at least initially) targeted mothers rather than those expected to celebrate their mothers. One reason for the lower volume of general campaign ads is that she may have been exposed to relatively more personal targeting than persona A and E. (See Figure 4.)

In addition to the above mentioned “Feeding America” campaign, US-based users were heavily pursued by two other ad campaigns: “Corona: Wear a mask" and “VRBO ". Like “Feeding America”, the “Corona” campaign clearly focused on current societal crises, whereas “VRBO" is an American vacation rental marketplace (that we classified as a campaign as it was observed across all profiles and regions). That such a campaign shows up here too is perhaps not surprising, as we did find that both “traveling" and (especially) “fashion" ads typically were seen across all personas. These two categories together with “electronics" were the only categories for which we observed ads for all personas. (See columns with all non-zero values in Table 4.)

Secondary phase: Despite a significant reduction in the ad volumes observed when using extensions, the above profiles observed ads for all the above-mentioned campaigns also during the secondary collection phase (after having added extensions). Furthermore, the top campaigns “Feeding America” (49 %) and “Corona: Wear a mask” (29 %) remained the dominating ad campaigns.



(a) Age group (b) Gender

Figure 10: Browser-based personalization comparison.

4.6 Browser differences

Persona comparison. While all profiles observed more ads per screenshot with Firefox than with Chrome (Table 9), the personalization experienced by the different profiles is relatively independent of whether we use Chrome or Firefox. For example, in both cases, personas D and F see the lowest level of personalization and persona B the highest level of personalization. These results are summarized in Figure 9 (comparing VMs 1-6 and VMs 13-18) and the findings are consistent also with what was seen when logged in to Chrome (see Figure 4). Also the two attributes gender (Figure 10(a)) and age (Figure 10(b)) showed similarities in the level of personalization between the two browsers.

Extensions. We have found that the relative level of personalization that users experience with the different extensions relative to when they do not use an extension is similar with the two browsers. For example, regardless of browser, CatBlock blocks the most ads and is the only extension of the four tested that reduces the amount of personalization experienced, whereas the other three extensions are less effective (compared to CatBlock) and result in higher ad personalization levels than if no extension would have been used.

Having said that, we did observe that the effectiveness of the different extensions sometimes differed depending on which browser they were used with. For example, with AdBlock, Chrome had a personalization level of 43.9% compared to 77.8% with Firefox (this difference was significant at the 95% confidence; p=0.0099). For the other extensions, the differences were smaller and non-significant: 55.6% vs 49.8% (AdBlock Plus; p=0.2721), 70.2% vs 85.5% (Ghostery; p=0.0635, and 11.3% vs 14.8% (CatBlock; p=0.4067). Other than these smaller (mostly non-significant) differences, the results appear consistent across the two browsers.

5 RELATED WORK

Related work broadly falls into one of three topics: (i) user tracking and advertisement, (ii) browser extensions, and (iii) user modeling.

User tracking and advertisement. Many users are unaware of the great extent that information about their online activities is being collected, aggregated, and used by various parties. Motivated by this unawareness and the impact that it has on user privacy, several studies have therefore focused on increasing the awareness about this form of privacy leakage online [30, 31]. For example, Malandrino et al. [31] implemented and tested different privacy-focused extensions aimed at helping the user make informed decisions and limit the spread of private information. In the study they also set up accounts on various (first-party) websites, added detailed information (e.g., full name, date of birth, email address, political views, sexual orientation, education and general interests) and studied how information was leaked and how much different third parties could learn. In contrast, we create user profiles and study the impact that different factors have on the level of ad personalization that these clients experience.

Others have studied the third-party tracking in the wild [5, 19, 37] or evaluated the tracking ability of a tracking service [16]. Much of this work crawls various webpages and profile the coverage of different third-party trackers [5, 37]. Interesting related work here include work by Degeling and Nierhof [16], who evaluates the tracking service Bluekai profiling ability during automated browser sessions emulated using links found posted on Reddit. While the study provides interesting light into how online profiles are tracked over time (e.g., with regards to interest, location and profession), the study does not consider the impact that these profiles have on the ad personalization experienced by the users.

Perhaps most closely related our work are works by Carrascosa et al. [14] and Barford et al. [9]. Both these works build personas and evaluate the level of personalized advertisement experienced by the personas. Carrascosa et al. [14] used a subset of the related websites provided by Google's Ad Words (as example webpages for those ad words) to create a set of narrow lists of websites focused on a particular topic or two, and then visited these webpages several times. Using 51 "regular" and 21 "sensitive" personas (each visiting at least 10 pages) they could identify interesting differences in how targeted advertising varies depending on the economic value associated with the users' interests and also compared the level of personalization observed from Spain and the US. Rather than using a fixed set of websites, we instead build a user model based on existing research on user modeling and use this to drive the web access patterns of each user. This allow us to capture regional differences in what users actually see when accessing the web that would be missed by Carrascosa et al. [14]. Our reuse of the same persona across several VMs (e.g., logged in vs not, several regions, server browsers, and several extensions) significantly differentiate our work from theirs.

Barford et al. [9] generated web traffic for each persona by going to each webpage on the Alexa top-50 lists of several webpage categories (these lists have since been discontinued by Alexa) and then clicked every advertisement link identified on these pages to (automatically) identify the website behind each ad. This behavior is not realistic for regular users, and it is expected that clicking advertisement links can bias the data. In contrast, we used hand crafted user profiles that generated more realistic user traffic and manually labeled each advertisement without clicking any advertisement links. Furthermore, similar to Carrascosa et al. [14], Barford et al. [9] did not control and/or study many of the factors studied in this paper.

Extensions. There are privacy and security risks associated with using the extensions. For example, Gulyás et al. [25] show how extensions can contribute to the uniqueness of a user, contributing to making users more recognizable to fingerprinting techniques used to identify users visiting a webpage. Multiple store-approved extensions have also been found to breach user integrity [22]. The main reasons for this are that browser extensions often have had the capability to modify and observe all browsing activity of a user [12], including the capability to insert code which can retrieve information such as cookies and form inputs from webpages [15].

In an effort to minimize the collection of sensitive user information, browsers have limited the data that extensions can collected through regulations [11, 28]. The expectation is that an extension should limit the data collection as much as possible and that it should not be allowed to collect any user data without explicitly saying so.

Alrizah et al. [8] present a study based on a crowdsourcing process of EasyList, the block list used by several of the extensions studied here. This study analyzed the update history of EasyList over a nine-year period, found concerning cases where EasyList incorrectly blocked legitimate content or EasyList editors failed to block content given by advertisers, and demonstrated how long websites are kept in the list. Malloy et al. [32] studied the geographic differences in both the adblock usage and the fraction of ads blocked by these extensions.

Others have focused on the performance of the adblockers. For example, Garimella et al. [21] investigated how the users' privacy in impacted by the use ad-blockers and the mechanism to counter them. Using a list of 30,000 URLs, they compared performance of the browsers in six different environments, with and without adblockers. With exception of Ghostery, they found that the ad blockers on average reduced the total data transferred by 25-33%, blocked between 60-80% of different privacy related parameters they considered (e.g., "track", "user-id" and "user-cookie"). More recently, Borgolte and Feamster [12] studied the user perceived performance when using adblockers and other privacy-focused extensions with Google Chrome and Mozilla Firefox. None of these studies considered the personalization of the ads displayed to the users.

User modeling. While our focus in this work is not on user modeling, we do leverage research from this domain in the creation of our user models. First, and most importantly, we build our model based on the post-search behavior of users modelled by White and Drucker [40]. In their work they divided and characterized the post-search behavior into four subcategories: backward-to-search, forward-to-search, forward-to-browse, backward-to-browse. Here, we emulate this user behavior in in our Selenium-based framework. Furthermore, we base our think times on research by Ramakrishnan et al. [38], who studied the think times of real users, and the used Zipf-like post-search clickthrough probabilities are motivated by several characterization works [7, 13, 20].

Incorporating these more realistic user models of the user behavior into our data collection process significantly differentiates our work from previous persona-based works studying ad personalization (see discussion above about works by Carrascosa et al. [14] and Barford et al. [9]). Furthermore, by automating this process we can run experiments with several different user profiles with the same persona (repeated for several example personas) from different locations, using different browsers, browser extensions, and when logged in vs when not.

6 CONCLUSIONS

This paper presents a profile-based evaluation of the targeted advertising experienced by different users, including those that try to protect their integrity using privacy enhancing extensions. Using the emulation tool developed in the project, we performed a 21-day longitudinal measurement campaign where we applied a carefully designed experimental methodology that allows us to evaluate the impact that many different factors have on the ad personalization perceived by six (or in some cases three) carefully handcrafted user personas. In these experiments, we collected daily data from 51 VMs, which in total visited 178,650 websites. From these website visits, we collected 230,175 screenshots, many of which we have manually gone through to identify and classify ads. Tools and datasets will be made available with the paper.

Using the datasets, we studied how the personalization changes over time, starting from the day that a new profile is put online, how the level of personalization changes when adding or removing an extension, and how the results differ depending on the profile's persona (e.g., interest, occupation, age, gender, etc.), geographic location (US East, US West, UK), what browser extension they use (none, AdBlock, AdBlock Plus, Ghostery and CatBlock), what browser they use (Chrome, Firefox), and whether they are logged in to their Google account or not. Here, we briefly discuss some of our findings and highlight differences in the success that today's ad providers achieve for different users.

In general, our conclusions appear to be consistent regardless of whether the users used Chrome or Firefox. For example, regardless of browser, we observed significant differences between the level of personalization achieved for different profiles. Here the most targeted individuals typically were interested in topics that may see bigger advertisement spending (e.g., fashion, travel, and electronics), were female, and were in the early-to-middle part of their careers (age groups 25-55). We also found that the level of personalization quickly ramped up (e.g., Google profiles had a good match within just a few days) and that the level of personalized advertising experienced by the users did not appear to be impacted by whether the users were logged in to Google or not, suggesting that the there is sufficient tracking for the advertisement companies to achieve highly successful ad targeting, regardless. This may perhaps be one contributing reason why Google is not planning to develop new ways to track users once they phase out third-party tracking [10] but instead will work with the web community to "create web technologies that both protect people's privacy online and give companies and developers the tools to build thriving digital businesses to keep the web open and accessible to everyone, now, and for the future [23]". If a more transparent technology can be developed that achieve a similar success rate as the current standards, then Google (and other advertisement providers) can achieve higher user trust while still being able to provide an attractive service for advertisers. Google also has the competitive advantage that they have the most popular browser (Chrome) and people's high reliance on their search engine (also used in our study) influences which pages users actually visit (which itself – as we show here – may impact the level of personalization a user experience).

While the usage of extensions helped reduce the number of ads that users were exposed to, the ads that slipped through the filters provided significantly higher average personalization than the average personalization experienced by users not using an extension. It was also clear that companies can successfully build user profiles during the time an extension is used. Interestingly, the least established extension, CatBlock, let through substantially fewer ads than the other extensions and was the only one of the extensions that reduced the fraction of personalized ads. With the other (more established) extensions, we observe a statistically significant increase in the fraction of personalized ads, raising question whether it simply is harder to block personalized ads (at least without deteriorating the user experience) or whether some of the ad providers that provide the most targeted ads may be whitelisted.

The main regional differences that we observed were with regards to the volume of ads and the exposure to ad campaigns. For example, UK-based users saw only 32% of the number of ads seen by their US-based counter parts and most of the identified ad campaigns were US focused (e.g., "Feeding America", "Mother's Day") or primarily targeted US users (e.g., "Corona: Wear a mask", "VRBO"). Even the most observed campaign in the UK, was US-focused ("Feeding America"). Otherwise, the UK-based users saw slightly higher personalization. Overall, we believe that this work provides an important profile-based characterization of the current level of targeting that ad providers achieve for different users. Given the changes that may happen as Google and the web community sets out to reshape the tracking and advertisement landscape, we also believe that this study serves as an important reference point of the current level of personalization that may be compared against (e.g., using the framework presented here) after Google eventually phase out support for third-party cookies [23].

REFERENCES

[1] [n.d.]. CatBlock. https://getcatblock.com/
[2] [n.d.]. Ghostery. https://www.ghostery.com/
[3] [n.d.]. Privacy Badger. https://privacybadger.org/
[4] 2021. Who's Most Targeted and Does My New Adblocker Really Help: A Profile-based Evaluation of Personalized Advertising. In *Proc. ACM CCS Workshop on Privacy in the Electronic Society (ACM WPES @CCS)*.
[5] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *CCS'14: Proceedings of the ACM Conference on Computer and Communications Security.* https://doi.org/10.1145/2660267.2660347
[6] AdBlock. [n.d.]. About AdBlock. https://getadblock.com/
[7] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. 2006. Learning User Interaction Models for Predicting Web Search Result Preferences. In *SIGIR '06: Proceedings of the annual international ACM SIGIR conference on Research and development in information retrieval.* https://doi.org/10.1145/1148170.1148175
[8] M. Alrizah, S. Zhu, X. Xing, and G. Wang. 2019. Errors, misunderstandings, and attacks: Analyzing the crowdsourcing process of ad-blocking systems. In *Proceedings of the Internet Measurement Conference.* 230–244.
[9] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan. 2014. Adscape: Harvesting and Analyzing Online Display Ads. In *WWW'14: Proceedings of the International Conference on World Wide Web.* https://doi.org/10.1145/2566486.2567992
[10] BBC. 2021. Google promises to drop personalised ad tracking. BBC News. https://www.bbc.com/news/technology-56267425
[11] Alexandre Blondin. 9 Dec, 2020. Making Chrome extensions more private and secure. *Chrome - safety and security* (9 Dec, 2020).
[12] K. Borgolte and N. Feamster. 2020. Understanding the Performance Costs and Benefits of Privacy-focused Browser Extensions. In *WWW '20: Proceedings of The Web Conference.* https://doi.org/10.1145/3366423.3380292
[13] L. Breslau, Pei Cao, Li Fan, Graham Phillips, and S. Shenker. 1999. Web Caching and Zipf-Like Distributions: Evidence and Implications, In INFOCOM '99: Proceedings of the Conference on Computer and Communications Societies. *Proceedings - IEEE INFOCOM*, 126 – 134. https://doi.org/10.1109/INFCOM.1999.749260
[14] J.M Carrascosa, J. Mikians, R.Cuevas, V. Erramilli, and N. Laoutaris. 2015. I Always Feel like Somebody's Watching Me: Measuring Online Behavioural Advertising. In *Proceedings of the ACM Conference on Emerging Networking Experiments and Technologies.* https://doi.org/10.1145/2716281.2836098
[15] Q. Chen and A. Kapravelos. 2018. Mystique: Uncovering Information Leakage from Browser Extensions. In *CCS '18: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* https://doi.org/10.1145/3243734.3243823
[16] M. Degeling and J. Nierhoff. 2018. Tracking and Tricking a Profiler: Automated Measuring and Influencing of Bluekai's Interest Profiling. In *WPES'18: Workshop on Privacy in the Electronic Society.* 13. https://doi.org/10.1145/3267323.3268955
[17] Disconnect. [n.d.]. No more trackers. https://disconnect.me/
[18] EasyList. [n.d.]. Overview. https://easylist.to
[19] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *CCS '16: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* 1388–1401. https://doi.org/10.1145/2976749.2978313
[20] M. J. Freedman. 2010. Experiences with CoralCDN: A Five-Year Operational View. In *NSDI '10: Proceedings of the USENIX Symposium on Networked Systems Design and Implementation.* https://doi.org/10.5555/1855711.1855718
[21] K. Garimella, O. Kostakis, and M. Mathioudakis. 2017. Ad-blocking: A Study on Performance, Privacy and Counter-measures. In *Proceedings of the ACM on Web Science Conference.* 259–262. https://doi.org/10.1145/3091478.3091514
[22] C. Giuffrida, S. Ortolani, and B. Crispo. 2012. Memoirs of a browser: a cross-browser detection model for privacy-breaching extensions. In *ASIACCS '12: Proceedings of the ACM Symposium on Information, Computer and Communications Security.* https://doi.org/10.1145/2414456.2414461
[23] Vinay Goel. 2021. An updated timeline for Privacy Sandbox milestones. Google blog. https://blog.google/products/chrome/updated-timeline-privacy-sandbox-milestones/
[24] Google. 2021. Chrome web store. https://chrome.google.com/webstore/category/extensions
[25] G. Gulyás, D. Somé, N. Bielova, and C. Castelluccia. 2018. To Extend or not to Extend: On the Uniqueness of Browser Extensions and Web Logins. In *WPES'18: Proceedings of the Workshop on Privacy in the Electronic Society.* https://doi.org/10.1145/3267323.3268959
[26] Raymond Hill. [n.d.]. uBlock Origin. https://chrome.google.com/webstore/detail/ublock-origin/cjpalhdlnbpafiamejdnhcphjbkeiagm?hl=en-GB/
[27] Leslie K. John, Tami Kim, and Kate Barasz. 2018. Ads That Don't Overstep. *Harward Business Review* (Jan/Feb. 2018), 62—-69.
[28] P. Kewisch, Rebloor, A. Wagner, A. Tasy, J. Villalobos, W. Bamberg, and Kmaglione. [n.d.]. Add-on Policies. https://extensionworkshop.com/documentation/publish/add-on-policies/
[29] B. Krishnamurthy and C. E. Wills. 2010. On the leakage of personally identifiable information via online social networks. *Computer Communication Review (CCR)* 40 (2010), 112–117. https://doi.org/10.1145/1672308.1672328
[30] A. Majumder and N. Shrivastava. 2013. Know Your Personalization: Learning Topic Level Personalization in Online Services. In *WWW '13: Proceedings of the 22nd international conference on World Wide Web.* https://doi.org/10.1145/2488388.2488464
[31] D. Malandrino, A. Petta, V. Scarano, L. Serra, R. Spinelli, and B. Krishnamurthy. 2013. Privacy Awareness about Information Leakage: Who knows what about me?. In *WPES '13: Proceedings of the Workshop on Privacy in the Electronic Society.* https://doi.org/10.1145/2517840.2517868
[32] M. Malloy, M. McNamara, A. Cahn, and P. Barford. 2016. Ad blockers: Global prevalence and impact. In *Proceedings of the Internet Measurement Conference.* 119–125.
[33] G. Merzdovnik, M. Huber, D. Buhov, and N. Nikiforakis. 2017. Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools. In *EuroS&P'17: IEEE European Symposium on Security and Privacy.* https://doi.org/10.1109/EuroSP.2017.26
[34] Mozilla. 2021. Firefox browser add-ons. https://addons.mozilla.org/en-US/firefox/
[35] AdBlock Plus. [n.d.]. Surf the web with no annoying ads. https://adblockplus.org/
[36] E. Pujol, O. Hohlfeld, and A. Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. In *IMC '15: Proceedings of the Internet Measurement Conference.* 93–106. https://doi.org/10.1145/2815675.2815705
[37] J. Purra and N. Carlsson. 2016. Third-party tracking on the web: A Swedish perspective. In *IEEE Conference on Local Computer Networks (LCN).* IEEE, 28–34. https://doi.org/10.1109/LCN.2016.14
[38] R. Ramakrishnan, V. Shrawan, and P. Singh. 2017. Setting Realistic Think Times in Performance Testing - A Practitioner's Approach. In *ISEC '17: Proceedings of the Innovations in Software Engineering Conference.* https://doi.org/10.1145/3021460.3021479
[39] Statista. 2021. Digital advertising spending worldwide from 2019 to 2024. https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/ Accessed July 16, 2021.
[40] R. White and S. Drucker. 2007. Investigating Behavioral Variability in Web Search. In *WWW '07: Proceedings of the international conference on World Wide Web.* https://doi.org/10.1145/1242572.1242576