This paper appears at the 2022 Internet Measurement Conference (IMC '22).

What Factors Affect Targeting and Bids in Online Advertising? A Field Measurement Study

Eric Zeng*
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
ericzeng@cmu.edu

Rachel McAmis
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
rcmcamis@cs.washington.edu

Tadayoshi Kohno
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
yoshi@cs.washington.edu

Franziska Roesner
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
franzi@cs.washington.edu

ABSTRACT

Targeted online advertising is a well-known but extremely opaque phenomenon. Though the targeting capabilities of the ad tech ecosystem are public knowledge, from an outside perspective, it is difficult to measure and quantify ad targeting at scale. To shed light on the extent of targeted advertising on the web today, we conducted a controlled field measurement study of the ads shown to a representative sample of 286 participants in the U.S. Using a browser extension, we collected data on ads seen by users on 10 popular websites, including the topic of the ad, the value of the bid placed by the advertiser (via header bidding), and participants' perceptions of targeting. We analyzed how ads were targeted across individuals, websites, and demographic groups, how those factors affected the amount advertisers bid, and how those results correlated with participants' perceptions of targeting. Among our findings, we observed that the primary factors that affected targeting and bid values were the website the ad appeared on and individual user profiles. Surprisingly, we found few differences in how advertisers target and bid across demographic groups. We also found that high outliers in bid values (10x higher than baseline) may be indicative of retargeting. Our measurements provide a rare *in situ* view of targeting and bidding across a diversity of users.

1 INTRODUCTION

Online advertising is an enormous and complex system, allowing millions of advertisers to reach billions of users across millions of websites, with the capability to target individual users based on their interests, online history, and personal information. On the web, this system is underpinned by a tangled ecosystem of ad tech companies, intermediaries who run the infrastructure for determining which ads are placed on which pages. This model is known as *programmatic advertising*, where for every web page that a user loads, advertisers compete in an automated, real-time bidding auction to determine who gets to place their ads on the page.

The complexity and scale of the online advertising ecosystem makes it difficult for observers outside of the industry to answer empirical questions about how it operates, and how it impacts users' privacy. For example: What information do advertisers use to target ads? How do advertisers decide how much to bid to place ads? And how do factors like a user's demographics and the website the ad appears on affect how users are valued or targeted?

Though prior measurement work has provided some answers on these questions, such as work observing the existence of behavioral targeting and retargeting [8, 21, 29, 30], and measurements of winning bid values from real-time bidding and header bidding auctions [11, 31–33, 35], these studies collect their data through crawler-based experiments, or through field studies with non-representative convenience samples. In the case of crawler studies, statistics like proportions of targeted ads, or bid values, might not be representative of what end users actually experience on the web [25, 44]; or in the case of field studies with limited samples, studies may overlook differences in the user population due to demographics or other factors.

In this paper, our goal is to measure the factors that advertisers use to decide how to target ads, and how much they pay to run those ads, using ecologically valid observations from end users in the wild. We ask the following **research questions**:

(1) How much ad targeting occurs at the individual, demographic, and contextual levels?
(2) How much do advertisers pay to show ads to people, and how do individual, demographic, and contextual factors affect the amount they pay?
(3) How much targeting do users perceive, and do those perceptions relate to bid values?

Estimating the influence of individual, demographic, and website factors on targeting and bid values from user data is challenging, because different users have vastly different browsing habits and histories, and contextual factors like differing ad networks and trackers on websites will affect the ads they see. To control for many of these factors, we scope our study methodology methodology based on the following **measurement goals**:

- *In situ data collection:* To accurately measure behavioral targeting, we aimed to collect data directly from participants' primary browsers, so that the ads that we collect are based on their existing browsing profiles.
- *Demographic representativeness:* Convenience samples of the population, such as friends and colleagues, or unscreened

*Now affiliated with Carnegie Mellon University.

online participant pools, may have skewed demographics — often younger and less tech savvy. Unrepresentative samples can exclude certain demographics and decrease generalizability. Thus, we aimed to recruit a demographically representative sample of participants in the U.S.

- *Control for differences in websites:* In their daily lives, people browse different sets of websites. When comparing ads seen by people in a field study, this makes it difficult to attribute whether differences came from contextual targeting of websites, or behavioral targeting based on past history. To measure differences resulting from behavioral targeting, and control for website-based targeting, we aimed to collect data from a fixed set of websites for all participants.
- *Control for changes over time:* Market conditions, advertising campaigns, as well as user behaviors and preferences, may change over time, affecting results data collected at different times. Thus, we aimed to collect data from small snapshot in time (11 days in Dec 2021) to minimize longitudinal effects.

With these goals, we designed a controlled field measurement study. First, we recruited a representative sample of 286 U.S. participants, asking for demographic information through Prolific.

Participants installed a browser extension that collected the content and winning bid values (via header bidding) of the ads shown to them. All participants visited the same set of 10 websites, to control for differences in topics, popularity, ad networks, and trackers across websites. We also surveyed participants about the perceived level of targeting of a sample of the ads shown to them. In total, we collected 41,032 ads, including 7,117 with winning bid data.

The **contributions** of our measurements include:

- We provide empirical measurements of ad targeting from a representative sample of real users in the U.S., showing large differences in the categories of ads seen on different websites and by different individuals, and minor differences between demographic segments like age and gender.
- We quantify the value of users to advertisers in the wild, using data from header bidding auctions. We observe little to no effect of demographic factors on bid values, but we do find variation in bid values across websites, individuals, ad categories, and ad networks.
- We find that ads with exceptionally high winning bid values (up to 16x higher than average) typically promote products that users previously viewed, providing additional evidence that high bid values correlate with retargeting.
- Our findings complement and concur with findings from prior work measuring targeting and bid values, confirming in the field the same forms of targeting measured by crawlers, and adding evidence that bid values are increasing over time.

2 BACKGROUND

We provide background on how ad auctions in programmatic advertising operate, including real time bidding (RTB) and header bidding. Then, we explain how programmatic ad auctions are the mechanism used to implement targeted advertising.

Real-Time Bidding. Real-time bidding is an method for connecting advertisers, who want to buy ads, to publishers, who are selling spaces on their websites. When a user loads a webpage with an ad, a script on the page will contact one of the website's *demand partners* and request an ad. These demand partners are typically *supply side platforms* (SSP) or *ad networks*, which are entities whose primary purpose is to help websites place ads on their page. Upon receiving a bid request, SSPs will forward the request to an *ad exchange*, which runs an auction where advertisers can bid on the opportunity to run their ad in that slot (usually offered via another intermediary — a *demand side platform* (DSP) [42]). The ad that wins the auction is rendered on the website, and the advertiser pays the website (and intermediaries) the amount they bid [13]. The value of a bid is typically denoted in CPM, or cost per mille, which means the cost to show 1000 impressions of an ad. For example, a typical bid may be $1.50 CPM, or $0.0015 to show the ad to a single user.

Targeting and Bid Strategies. To help decide how much to bid in RTB auctions, bidders are supplied with identifiers for the user, like cookies or fingerprints, which they can use in conjunction with data collected by web trackers and data brokers to find users' interests, browsing behavior, and real world behaviors [42]. Bidders have many strategies for choosing what to target, like targeting visitors of specific websites (contextual targeting) [34], users that appear to be interested in a topic based on past browsing history (behavioral targeting) [10], users that had previously visited their website (remarketing) [42], or people in specific geographical areas (geotargeting) [10]. Determining the exact bid value is an optimization problem where multiple factors are considered to determine the optimal bid value, such as the targeting parameters, budget and strategy of the ad campaign, and how well the ad matches the available information about the website and user [7, 9, 24, 47, 48].

Header Bidding. To complicate matters, websites may partner with more than one company to solicit ads. Websites can make requests to multiple ad networks or SSPs, like OpenX, Criteo, and Google Ads; or run ads via direct orders (a direct agreement with an advertiser). Each of these demand partners run their own RTB auctions, and offer different bids — and some exchanges may not provide a bid at all [33]. To decide on which demand partner to select for a given ad slot, websites previously used a static priority list, known as "waterfalling" [13], but this approach can be suboptimal when demand partners farther down the list offer higher bids.

To optimally decide on which demand partner to pick when filling an ad slot, many websites began using a technique called *header bidding*. Header bidding allows a website to solicit bids from multiple demand partners in parallel, and pick the highest bid from among them. Header bidding auctions often take place in a client-side JavaScript library, such as Prebid.js. A diagram illustrating this process is available in Appendix A.

Header bidding is advantageous for researchers, because it makes bids transparent. In RTB, bids could be observed through win notifications, but these are increasingly encrypted, making bid prices difficult to measure [35]. Header bidding is typically implemented as a JavaScript library (e.g. Prebid.js), which allows researchers to directly view bid responses by querying the header bidding script using an instrumented browser or browser extension.

3 RELATED WORK

There is a rich body of measurement research aiming to bring tranparency to the online advertising ecosystem.

Targeting Measurements. Prior work has measured targeted ads from a variety of perspectives. Most commonly, web crawlers with synthetic profiles or personas are used to measure behavioral targeting and contextual targeting. In the absence of having access to browsers with real user profiles, crawlers visit a curated list of websites to generate a profile that signals interest in a certain topic, and compare ads seen in different profiles. Crawler-based targeting studies have found that certain ad categories, and personas are more heavily targeted than others, such as health, travel, and shopping [8, 29, 30]. A similar study using fine-grained targeting detection also found that health ads were highly targeted in Gmail [27]. However, it is unclear whether measurements conducted using synthetic profiles are representative of real users [25, 44].

Other crawler-based case studies have examined problematic targeting practices, such as gender discrimination in the behavioral targeting of career ads [12], and contextual targeting of misleading political ads on politically partisan websites [46].

Few studies have measured targeting in field studies with real users. Parra-Arnau et al. collected field measurements to validate their targeting detection method, finding that retargeting was common, and that large firms were responsible for most behavioral targeting, but only used a small convenience sample of other researchers and friends [36]. Iordanou et al. developed a privacy-preserving methodology for detecting demographic-based targeting from crowdsourced data from real users, finding that women, older people, and middle income people were more likely to be targeted, but they did not collect data on the content of ads or websites [21].

Our work adds to this literature by investigating targeting based on demographic factors using data from real users, and by comparing the relative impact of contextual, behavioral, and demographic factors on targeting.

Real-Time Bidding and Header Bidding Measurements. Prior work has measured multiple aspects of ad auctions through real-time bidding (RTB) and header bidding (HB).

Most closely related to our work, a number of papers have measured bid values to quantify the value of users and identify the factors that affect bid values. Olejnik et al. and Papadopoulos et al. measured bid values from RTB auctions, using data collected from convenience samples of real users. They found that bid prices can be affected by contextual and longitudinal factors, such as time of day and year, country, ad slot sizes, operating system, website category, ad category, and retargeting [31, 35]. Pachilakis et al. replicates this work to measure differences in bid values over a multi-year scale, they found increases in bid values due to cookie syncing, and analyzed the effect of gender and age, but did not obtain a demographically representative sample [32]. Other studies have measured bid values through HB using crawlers, finding differences due to ad slot sizes and crawling profiles [11, 33].

Other studies used bid responses as a mechanism to measure other phenomena. Cook et al. utilized bid values from HB to learn tracker-advertiser relationships [11]. Iqbal et al. used header bidding as a signal to detect retargeted ads originating from queries to smart assistants [22]. Other measurements of ad auctions examine performance metrics, such as latency of bid responses and the bidding behaviors of ad networks in the auctions [5, 33, 43].

Our work adds to this literature by providing measurements of HB bid values from a demographically diverse sample of real users, providing insight into demographic effects on bid values, and by separating the effects of other factors such as site, demand partner, and individual variation.

Other Related Work. Farther afield, other work has investigated issues with targeted ads on other platforms like Facebook, such as discrimination in ad delivery [4, 20], and targeting of harmful ads [3] and misinformation [37]. Other work has measured the prevalence of web trackers and fingerprinting which enable behavioral targeting on the web [1, 2, 6, 16, 23, 28, 38].

4 FIELD STUDY METHODOLOGY

In this section, we describe the methodology for our field study. As described in Section 1, our overall goal was to investigate how individual, demographic, and contextual factors affected how advertisers targeted and bid on ads. Based on our measurement goals, we scoped our study in the following ways: 1) We collected data from real users' browsers, leveraging their existing browsing profiles to measure behavioral targeting. 2) We recruited a demographically representative sample to improve generalizability. 3) To isolate the factors we aimed to investigate and allow direct comparisons between participants, we controlled for differences in context and browsing habits during data collection by collecting data from a fixed set of websites, at approximately the same point in time.

4.1 Participant Recruitment

We recruited a demographically representative sample of 286 U.S. participants from Prolific. We chose to obtain a representative sample so that we could make comparisons across demographic categories such as age, gender, and ethnicity.

Because online panels are known to have skewed demographics, we used a two-part recruitment method. First we conducted a prescreening survey, open to all U.S.-based Prolific users, where participants provided their age, gender, and ethnicity, primary browser, and whether they used an ad blocker. Optionally, we asked for participants' sexuality, income, and ZIP code.

Next, we filtered out all respondents except those who used either Google Chrome and Microsoft Edge, for compatibility with our extension, and to control for privacy features in other browsers that could affect participants' advertising profiles. We also filtered out participants who reported using ad blockers, which could similarly impact their profiles.

Then, we used stratified sampling to select a representative group of participants. Using G*Power [17], we calculated that we needed a sample size of at least 126 participants to detect medium effect sizes using a linear regression with 10 predictors (our initial modeling approach for analyzing the effect of demographic factors on bid values). We created quotas for each cross-section of the population by age, gender, and ethnicity, based on U.S. demographic data from the 2020 American Community Survey [41], aiming for 300 participants, such that the smallest gender-age-ethnicity subgroups would have contain 1-2 participants. We invited batches of participants to

Table 1: Websites visited by participants in the study.

Website	Topics	Site Rank
businessinsider.com	National and business news	137
weather.com	Weather forecasts and news	288
speedtest.net	Internet performance test	289
usnews.com	National news, college rankings	365
foodnetwork.com	Recipes and cooking content	1016
detroitnews.com	Local newspaper	2904
ktla.com	Local TV news	4626
phonearena.com	Tech news, smartphone reviews	4954
fashionista.com	Fashion and celebrity news	8773
oxfordlearnersdictionaries.com	Online dictionary	8903

a second, private Prolific study, until all quotas were filled. However, we excluded 14 participants post-study due to anomalies in their data, e.g. they used an ad blocker, or could not load particular sites.

4.2 Study Procedure

We ran our field study between December 10-21, 2021. Participants selected for the study were directed to our website with a consent form, and instructions to install our browser extension. Upon installing the browser extension, the extension opened a page asking the participant to sign in with their Prolific user ID, followed by an instructions page.

4.2.1 Website List. After the instructions, participants were redirected to a page showing a list of 10 websites to scan using our extension (Table 1). All participants were asked to visit the same websites to control for contextual targeting, in randomized order to control for ordering effects. We limited the study to 10 websites because our extension required active participation, so we needed to ensure the study did not take too long to complete.

We chose the 10 websites by scanning the top 10,000 websites on the Tranco top sites list, filtering to sites which contained the prebid.js header bidding script, finding 703 sites. Then, we manually evaluated the sites, looking for a set of websites that reliably received bid responses and spanned a range of topics and popularity.

4.2.2 Data Collection. When a participant visited a site on our list, the extension's content script displayed a modal dialog, asking them for permission to start a scan. When the scan was initiated, the extension used CSS selectors from an ad blocker filter list (EasyList) to determine which elements on the page were ad slots.

For each ad, the extension scrolled it into view, and attempted to extract bid metadata from the Prebid.js header bidding script, which is accessible from the global JavaScript context. The extension's content script queried the following APIs: `getBidResponses()` which returns all bids received, `getAllWinningBids()` which returns winning bids for ads which were rendered on the page, and `getAllPrebidWinningBids()` which returns winning bids for ads which won their auction, but the site decided not to run on their page.[1] These calls return bid metadata for all ad slots on the page; so the extension attempted to match bids to the ad currently in view, by checking if the `id` of the ad slot's HTML element matched the `adUnitCode` field in each bid response. If a matching bid for the ad slot was found, the extension took a screenshot of the ad (storing it locally) and sent the header bidding data to the study server. If a bid could not be matched to an ad, then the ad was skipped.

[1]A reason why an ad could win a header bidding auction, but not appear on the page, is that the site has another demand partner that takes precedence over the header bidding result (i.e. waterfall prioritization [13])

After scanning all ads, the extension automatically refreshed the page and collected a second run of data, to increase the sample size of ads collected per site and participant. Thus, each participant loaded 20 pages during the course of the study.

4.2.3 Targeting Perceptions Survey. After visiting all 10 websites, participants were redirected to a survey, where participants rated how targeted they felt by the ads collected. The extension draws a deterministic sample of 8 ads to show the participant; by ranking the ads by winning bid value, and selecting ads at uniform intervals from the lowest to highest value ad. We chose this over random sampling to guarantee that the sample contained ads with a range of bid values. We limited the number of ads in the survey to 8 to reduce participant fatigue and drop out rates.

For each ad in the sample, we asked the participant four questions about their perceptions of the targeting of the ad:

(1) *(Relevance)* "How relevant is this ad to your interests?" (1-5 Scale)
(2) *(Targeting)* "How personalized or targeted is this ad to you?" (1-5 Scale)
(3) *(Likeliness to Click)* "How likely would you be to click on this ad?" (1-5 Scale)
(4) *(Retargeting)* "Have you ever previously clicked on this ad, viewed the product or website featured in the ad, or bought the product in the ad?" (Yes/No/Not Sure)

4.2.4 Data Exclusion. Lastly, we provided a chance for participants to remove any screenshots of ads which they felt might be sensitive, e.g. if they felt that the ad was targeted and the screenshot would reveal unwanted information to us, the researchers. Participants were shown all of the ads we collected (and stored locally), and selected the ones they did not want to upload to our server.

4.3 Labeling Ad Categories

To enabled analysis of targeting, we assigned ads to categories using a mix of automated and manual approaches.

First, we used a topic model to automatically place ads into semantically similar clusters. We first used the Google Cloud Vision API to extract text from ad screenshots. We then used locality sensitive hashing to deduplicate ads. Then, we used the BERTopic topic modeling library [19], which combines several algorithms: the all-MiniLM-L12-v2 language model for generating embeddings, UMAP for dimensionality reduction, and HDBScan for clustering. We also evaluated other topic modeling algorithms, like LDA and GSDMM, but found that BERTopic produced the most qualitatively coherent topics. The topic model produced 311 topics.

We then manually audited the topics, finding overlapping topics, misclassified ads, and generally too many topics for analysis. We manually combined similar topics together into 52 categories of products, such as "medications", "home kitchen and bathroom products", and "electronics". We manually verified each category and moved misclassified ads.

Some ads were not assigned a category, either because the ad was blank, cut off by a popup, or in the middle of loading when the screenshot was taken, or because multiple ads were captured in the image, and we could not determine which ad the header bidding data corresponded to. These ads are excluded from our analysis.

4.4 Ethics

Our study was approved by our institutional review board, which determined that the study qualified for Category 3 Exemption.

Participants agreed to a consent form explaining the risks of the study before starting. Participants were compensated $0.25 for completing the pre-screening survey, and $8.00 for completing the browser extension study, a rate $15.00 per hour by our initial estimates for completion time. Some participants took much longer than expected due to technical issues; in these cases we provided bonus payments to compensate them for the additional time.

We took into consideration users' privacy and safety in multiple aspects of the design of our study and browser extension:

First, we designed the extension to require user input and consent before collecting data: rather than immediately taking control of the browser like a crawler, participants manually visited each site on our list. Then, upon opening a page on the list, the extension asked for permission to start scanning before starting the data collection procedure. For websites not on the list, the content script would not execute at all, so participants could use the site normally.

Second, we were aware that screenshots of ads could inadvertently expose information about participants, if the ads were targeted and revealed something sensitive that they did not want to share. To give participants control over what was shared with us, we added an interface where participants could exclude any screenshots that they found too sensitive before the data was uploaded.

Third, we provided clear instructions for participants to remove the extension at the conclusion of our study, but the extension did not continue to collect any data if participants forgot to remove it.

4.5 Limitations

Our study can only explain factors affecting behavioral targeting to a limited extent, because we do not have ground truth on the targeting parameters used by advertisers to target ads, nor do we have the advertising profiles that ad networks have inferred about participants. Our analysis is able to show correlations between externally observable factors (e.g. participant demographics, website) and the frequency of different categories of ads. Though this does not directly measure how advertisers decide target people, it does show the overall effect of targeting as experienced by different demographics of people, and how it is experienced across websites.

Though we strove to make our participant sample representative by balancing across age, gender, and ethnicity, the size and composition of the sample does not fully capture all of the variation in the U.S. population. Variation among certain individual segments may not be represented fully due to low proportions of certain ethnicities in the U.S. - for example, our sample only contained one Latino male aged 35-44 years old. Additionally, our sample is not balanced across other potentially relevant demographics for ad targeting, such as income or geography, due to practical constraints on the number of participants we could recruit and the number of factors we could stratify simultaneously. Finally, our sample is U.S. centric, and our findings may not generalize to other countries.

We selected a limited set of 10 websites, to control for websites as a variable, and to keep the duration of the study short. However, the small sample size means that certain results may be specific to the sites chosen, such as the overall counts of ads by category, or the overall average bid values.

The sample size of ads with winning bids was smaller than expected, with only 7117 ads. In some cases, we lack the statistical power for certain advanced analyses, such as interactions between factors. For example, we did not have the sample size to analyze an interaction effect between ads categories and a demographic characteristic of a participant, when predicting bid values.

The time period when the ads were collected was approximately 1-2 weeks before Christmas. Bid values may have been higher than usual, due to high demand for advertising during the Christmas shopping season in the U.S.

5 RESULTS

5.1 Dataset Description

5.1.1 Participant Demographics. In total, 286 participants successfully completed data collection for our study. Table 2 shows a summary of the demographic data of our study participants. Our dataset roughly approximates the U.S. population, but skews slightly younger and female. Table 3 shows the distribution of yearly household incomes of our participants, which roughly matches 2019 U.S. Census data. The median household income in our study was between $50,000 and $75,000, while the 2019 ACS median was $65,712 [40]. 267 participants used Google Chrome while 19 used Microsoft Edge.

5.1.2 Ads Overview. We collected 41,032 ads in total, or an average of 143.5 ads per participant, from 20 page loads each.

We were able to extract the winning bid in 25,764 of ads where a header bidding auction took place. Only in 7,117 ads of these ads was the winner actually rendered on the page — websites can choose not to use the winner of the header bidding auction, and instead choose an ad from another ad network to fill the slot instead.

Through topic modeling and manual qualitative analysis, we generated 52 categories describing the content of ads (see Section 4.3). We were able to assign categories to 31,407 ads, 9,625 ads were not assigned a category. Of the rendered winning bids, which we analyze in greater detail later, 5,851 out of 7,117 ads, or 82%, were assigned a category. Ads may not have been assigned categories if we detected anomalies (ads where popups or the extension UI accidentally covered the ad in the screenshot), if the ad was not fully loaded at screenshot time, or if multiple ads were in the screenshot.

In the study, we analyze four overlapping subsets of data:

- *Ads with categories* (31,407 ads). This subset contains the ads which we were able to assign a category to, either manually or automatically. We examine this subset in Section 5.2, where we analyze how the categories are distributed across demographics and sites.
- *Ads with rendered winning bids* (7,117 ads). These are ads for which we obtained the winning bid amount, and confirmed that the ad was rendered on the page. We examine this subset in Section 5.3.

Table 2: Demographics of the 286 participants in our study. All values are provided as percentages.

Gender	Female					*F-All*	Male					*M-All*	Non-binary		*NB-All*	*All*
Age	18-24	25-34	35-44	45-54	55+		18-24	25-34	35-44	45-54	55+		25-34	35-44		
Ethnicity																
Asian or Pacific Islander	2.45	1.05	0.35	0.35	0.00	*4.20*	2.45	2.10	1.05	1.05	0.00	*6.64*	0.00	0.00	*0.00*	10.84
Black or African American	1.75	2.10	1.40	0.70	0.35	*6.29*	0.35	1.75	1.40	0.35	0.00	*3.85*	0.00	0.00	*0.00*	10.14
Hispanic or Latino	4.90	1.40	1.75	0.00	0.00	*8.04*	1.05	2.10	0.35	0.00	0.70	*4.20*	0.35	0.00	*0.35*	12.59
Other	0.00	2.10	0.35	0.35	0.00	*2.80*	1.40	0.70	0.00	0.35	0.00	*2.45*	0.00	0.35	*0.35*	5.59
White or Caucasian	6.99	5.59	7.69	4.55	6.99	*31.82*	2.10	6.99	9.09	5.24	5.59	*29.02*	0.00	0.00	*0.00*	60.84
All	16.08	12.24	11.54	5.94	7.34	*53.15*	7.34	13.64	11.89	6.99	6.29	*46.15*	0.35	0.35	*0.70*	100.00

Table 3: Yearly household income of participants in our study.

Yearly Household Income	Count	%
Less than $25,000	52	18.18
$25,000-$49,999	73	25.52
$50,000-$74,999	43	15.03
$75,000-$99,999	47	16.43
$100,000-$124,000	35	12.24
$125,000-$149,000	11	3.85
More than $150,000	18	6.29

- *Ads with user targeting perceptions* (1,744 ads). These are the ads which participants rated with their perceptions of targeting, and is a strict subset of the above subset. We examine targeting perceptions in Section 5.4.
- *Ads with non-rendered winning bids* (18,916 ads). Ads which have a winning bid amount, but were not rendered on the page. We briefly discuss this subset in Section 5.1.3, but do not use this data for other analyses, because the screenshots captured do not correspond to the bid response.

5.1.3 Overall Winning Bid Values Averaged $5.47 per Thousand Impressions. How much did advertisers bid to show ads on the 10 sites in our dataset? The average winning bid had a mean value of $5.47 and median of $4.16 (IQR=$4.43). However, not all ads that won their header bidding auctions were rendered on the page. For non-rendered ads, the mean bid value was $3.60 CPM, and the median was $2.62 CPM (IQR = $3.25). Figure 1 shows the cumulative distribution functions for winning bids, separating ads that were rendered versus not rendered.

Though most bids won with a value less than $10, there is a substantial long tail of outliers. The top 10% most expensive winning bids were $10.62 CPM or above, and the top winning bid was $89.7 CPM, or nearly $0.09 to show a single ad. A case study of these outliers is available in Appendix D.

5.1.4 Summary of Ad Categories. Next, we summarize the categories of ad by content. Figure 2 shows the number of ads collected in each category, in the subset of all ads with a category (31,407 ads). Ads spanned a large variety of products, ranging from apparel, to home goods, and medications. The most common ads were for electronics (smartphones, computers, accessories), business ads (cloud computing, marketing services, office supplies, etc.), banking and finance ads (ads for mortgages, banks, investments), mixed native ads (a.k.a. content recommendation networks), and travel ads. Other

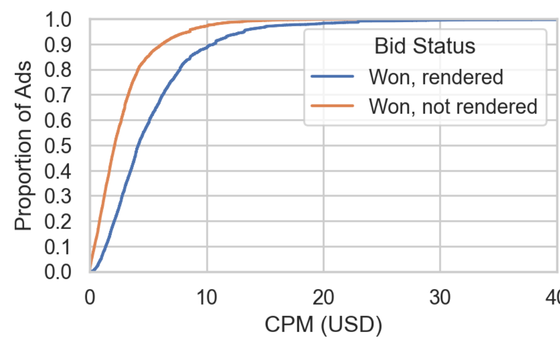

Figure 1: Cumulative distribution function (CDF) of all winning bid values in our dataset. Winning bid values for ads that were actually rendered on the page were higher than those that were not rendered.

notable categories specific to the dates when the measurements were conducted include COVID-19 related ads for vaccines, tests, and PSAs; and holiday-specific ads, such Christmas cards, gift wrap, and holiday sales (measurements were conducted 1-2 weeks before Christmas and other winter holidays in the U.S.).

Note that this distribution of ads by category is biased by the 10 sites we selected for the study; a different configuration of sites may result in a different category distribution. We discuss contextual targeting more in Section 5.2.1. We also observe some differences in the categories of ads in the subset with winning bid data, compared to the subset without bid data — see Appendix B for details.

5.2 How were ads targeted?

Next, we infer the amount of ad targeting in our dataset by analyzing whether categories of ads are correlated with likely targeting categories, such as demographic groups, websites, and individuals. We note that these are not direct measurements of targeting, as our data does not contain ground truth on the targeting parameters used by advertisers or the interest profiles of participants, but these results still serve to quantify the differences in the types of ads people see in the wild.

For demographic and contextual factors of interest, we conducted an omnibus chi-square test of independence, to determine whether there is a significant association between ad category and the factor of interest. We adjusted the resulting p-values for multiple comparisons using the Bonferroni method. To identify which categories were more or less common than expected (based on the overall proportions of ads by category across the dataset) we calculated

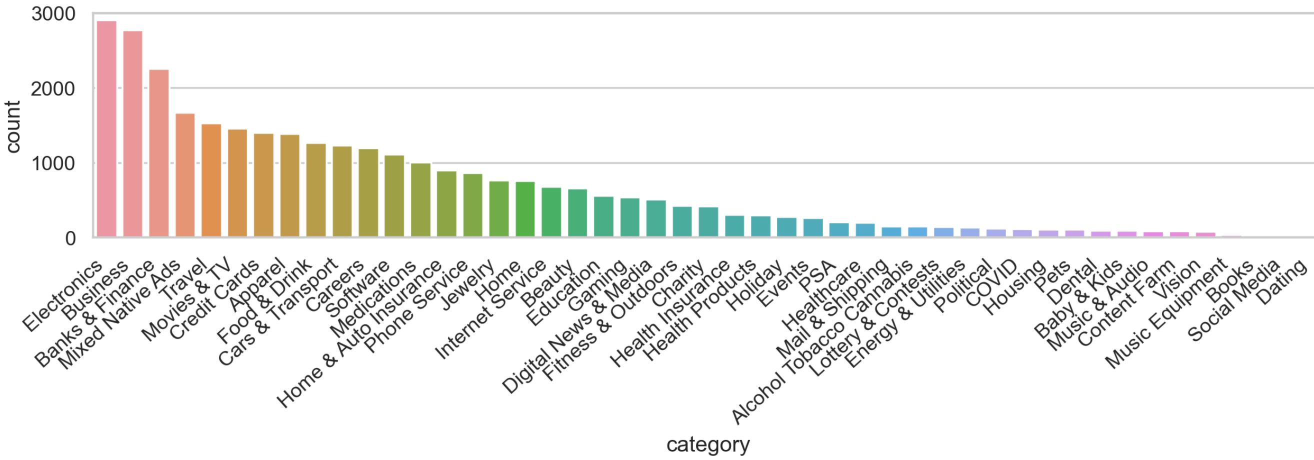

Figure 2: The number of ads in our dataset by category, including ads without winning bids associated with them.

the standardized residuals (a measure of the difference between the observed and expected cell value), and conduct a post-hoc Z-test, with critical values adjusted with the Bonferonni method. For individuals, we use distributional inequality metrics to characterize how each category of ad is distributed across individuals.

5.2.1 Strong Evidence of Contextual (Website-based) Targeting. We find that some categories are more common on specific websites than others, usually when the topic of the ad is relevant to the topic of the website — evidence of contextual targeting. A chi-squared test of independence found a significant association between website and category ($\chi^2(423, N = 31,407) = 37,155.82, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 202 of 470 residuals exceeded the critical value of 3.70 ($p < 0.05$), indicating that a large number of the categories were over- or under-represented on specific sites.

Table 4 shows the percentage of ads from each category on each website, for the 24 most common categories overall. Qualitatively, we find that categories that are more common than expected (in bold) are often related to the website. For example, ads in the "education" category, which contain ads for college programs and online classes, are much more common on usnews.com (11.24%), a website best known for its college rankings. speedtest.net, a tool for measuring internet speeds, had a high percentage of ads for gaming (14.6%) and internet service (20.7%); two topics where bandwidth is important. Business ads, which include marketing services and cloud software, were common on businessinsider.com (25.47%), a business news site.

5.2.2 Targeting Correlations with Demographic Factors. In a small number of ad categories, we identify correlations between the number of ads seen and demographic factors such as age, gender, and ethnicity. We note that these correlations may not be indicative of direct demographic targeting by advertisers, and may capture other targeting strategies instead, such as targeting of interests that correlate loosely with demographics.

Gender. We saw differences in the number of ads seen between genders in a small number of categories. A chi-squared test of independence found a significant association between gender and category ($\chi^2(92, N = 31,407) = 425.72, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 12 of 72 residuals exceeded the critical value of 3.39 ($p < 0.05$). Table 5 shows the percentage of ads by category. We found that women tend to receive more ads for Apparel and Beauty, while men tended to receive more ads for Gaming, Digital News, and Phone Service. We did not have enough non-binary participants to find significant differences.

Ethnicity. We saw significant differences in the number of ads seen between ethnicities in a small number of categories. A chi-squared test of independence found a significant association between ethnicity and category ($\chi^2(184, N = 31,407) = 690.03, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 23 of 235 residuals exceeded the critical value of 3.52. Table 6 shows the percentage of ads by category shown to people by ethnicity. Among the significant examples, Black and Latino participants were shown more Beauty ads, Latino participants were shown more Credit Card ads, White participants were shown more Charity and Home ads, and Asian participants were shown more Education ads.

Age. We saw differences in the number of ads seen across age ranges in a small number of categories. A chi-squared test of independence found a significant association between gender and category ($\chi^2(184, N = 31,407) = 735.93, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 20 of 235 residuals exceeded the critical value of 3.52 ($p < 0.05$). Table 7 shows the percentage of ads by category, across age ranges. 18-24 year olds saw more ads for apparel and travel, and fewer for careers, 25-34 year olds saw more ads for food and drink, 35-44 year olds saw more ads for careers, 45-54 year olds saw more ads for jewlery, and 55+ year olds saw more ads for internet service.

5.2.3 Individual Targeting. Next, we characterize the amount of variation in ads seen by individuals, due to possible behavioral targeting. Theoretically, if there are no differences in the ads seen by different people visiting the same sites, we would expect equal quantities of ads from each category in our study. However, with

Table 4: Percent of ads by category observed on each website (top 24 categories only). Blue/bold cells indicate a significantly higher proportion than expected, and red/italic cells indicate a significantly lower proportion than expected, based on post-hoc Z-tests on the standardized residuals. Darker colors indicate larger differences.

Category	businessinsider.com	detroitnews.com	fashionista.com	foodnetwork.com	ktla.com	oxfordlearnersdictionaries.com	phonearena.com	speedtest.net	usnews.com	weather.com
Apparel	*1.11*	**6.13**	6.11	**8.56**	**6.36**	**9.30**	*0.78*	*0.74*	*2.87*	5.50
Banks & Finance	6.01	7.54	*1.76*	*4.11*	*5.11*	5.85	*1.27*	*3.74*	**28.46**	7.88
Beauty	*0.66*	**3.20**	2.81	2.49	**3.09**	2.28	*0.49*	2.58	2.58	2.23
Business	**25.47**	*5.44*	7.59	*4.00*	*5.46*	**14.85**	*2.82*	10.80	*5.35*	*5.50*
Careers	**21.36**	*0.33*	*0.50*	*0.63*	*0.73*	*0.92*	*0.21*	*0.18*	*2.26*	*1.07*
Cars & Transport	4.11	**5.65**	*1.27*	4.32	**7.99**	3.39	*1.62*	*0.86*	3.42	3.83
Charity	*0.39*	1.43	**3.69**	0.88	1.70	1.66	*0.38*	0.43	**2.44**	1.75
Credit Cards	**13.16**	3.70	6.16	*1.58*	4.66	3.88	*1.22*	*2.27*	*2.15*	*2.67*
Digital News & Media	2.23	*0.19*	**17.55**	*0.53*	*0.10*	*0.00*	*0.07*	*0.06*	*0.73*	*0.92*
Education	*0.49*	*0.81*	*0.33*	*0.70*	1.22	2.53	*0.45*	1.04	**11.24**	1.28
Electronics	*0.98*	*3.70*	8.86	10.67	8.86	*5.61*	**35.49**	8.10	*2.66*	*3.93*
Fitness & Outdoors	*0.13*	*0.69*	0.44	**8.32**	1.18	0.37	*0.71*	*0.31*	1.24	0.85
Food & Drink	*1.07*	*2.76*	4.51	**9.72**	**6.53**	4.19	*1.15*	4.05	2.87	**6.74**
Gaming	*0.30*	*0.96*	0.99	1.89	*0.73*	2.28	*0.33*	**12.94**	1.24	2.28
Home	*0.26*	2.04	1.71	**5.51**	**4.00**	2.83	*0.85*	2.70	2.04	**3.97**
Home & Auto Insurance	*0.94*	*2.04*	*1.16*	3.93	**6.05**	**4.50**	*1.48*	*0.43*	2.40	**5.91**
Internet Service	*0.17*	*0.27*	1.60	*0.84*	*0.73*	*0.49*	*0.64*	**18.47**	**4.11**	**3.30**
Jewelry	**4.09**	*0.60*	**8.58**	**4.32**	2.36	1.11	*0.19*	*0.86*	2.00	2.57
Medications	*0.13*	**6.96**	2.75	2.42	*1.08*	2.16	*1.60*	2.33	2.26	**7.63**
Mixed Native Ads	*0.19*	**19.42**	*0.00*	*0.07*	*0.17*	*0.06*	**9.04**	*0.06*	*0.00*	**8.02**
Movies & TV	*2.98*	**6.73**	**6.99**	3.82	*1.43*	**9.37**	3.74	**10.00**	*2.00*	4.53
Phone Service	*0.39*	*0.23*	2.81	*1.23*	*1.46*	1.48	**13.74**	*1.29*	*0.80*	*1.38*
Software	*0.60*	*1.91*	*1.05*	*0.67*	*1.11*	**6.96**	**14.29**	**7.24**	*1.20*	*1.33*
Travel	**9.70**	*1.81*	*2.42*	**8.98**	**12.83**	3.94	*0.59*	*1.66*	*3.09*	*2.81*

the presence of individual targeting, a few participants may account for a large proportion of the ads in a category.

Figure 3 shows Lorenz curves for each ad category, which describe the level of distributional inequality [26] in who sees ads from each category. If a category of ads were distributed equally across participants, the line would be diagonal; the lower the curve, the more unequally the ads are distributed.

We find that ad categories had varying levels of distributional disparities. Some ads, like Mixed Native Ads, and Electronics ads, were shown roughly equally: the top 5% of participants saw 7.4% and 11% of the ads in those categories (if totally equal, the top 5% would have seen 5% of ads). On the other hand, ads for Charity ads and Fitness ads were much more unequally distributed; the top 5% of participants saw 24.7% and 26% of ads respectively. Though ads that were more common overall were generally more evenly distributed, this was not a perfect correlation: Apparel ads were less evenly distributed than Movies & TV (23% vs. 16% shown to the top 5% of participants), even though both categories contained around 1400 ads.

We also investigate whether behavioral targeting at the individual level might amplify contextual targeting. In Table 8, we compare the percent of ads seen by the top 5% of participants in contextually targeted categories on specific sites, with the percent of ads seen by the top 5% participants over the whole dataset. We find that within websites, ads likely to be contextually targeted were distributed *more* equally than in the overall dataset. Thus, in our sample, we do not see evidence of behavioral-contextual amplification.

5.3 What influences winning bid values?

In an ad auction, bidders consider many factors to determine the value of the ad, including the user's inferred interests, demographics, the website the ad appears on, and the targeting and budget parameters of the ads. To estimate the influence of each of these factors on bid values simultaneously, we used a linear mixed effects model to predict rendered winning bid values (response variable) as a function of the user's age, gender, and ethnicity (fixed effects/explanatory variables), as well as the website the ads appeared on, the bidder, the individual, and the category of the ad (random effects).

We selected our model using the top-down method suggested by Zuur et al. [49]: we started with a full specified model, including all of the above fixed and random effects, as well as other optional demographics we collected (sexuality, income, and children), and other labels we generated, such as whether ads used a native format, and labels based on our contextual targeting results. We did not include interaction effects, like gender and ad category, because we did not have enough data to estimate the number of parameters. We then experimented with removing random effects and

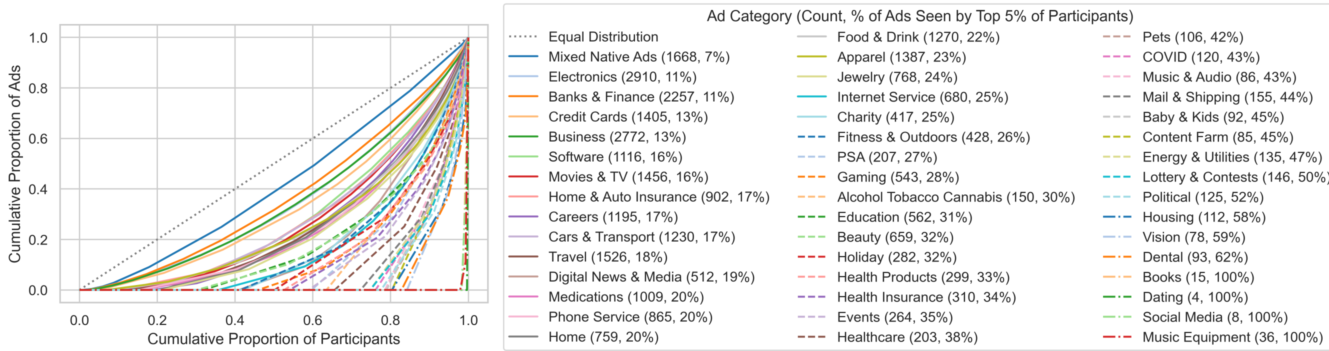

Figure 3: Lorenz curve showing the cumulative fraction of ads as a function of the cumulative fraction of participants, for each category. Curves closer to the diagonal line represent ad categories that are more evenly distributed across participants.

Table 5: Percent of ads observed by category, across genders (top 24 categories). Blue / bolded cells indicate a significantly higher proportion than expected, and red / italic cells indicate a significantly lower proportion than expected.

Gender	Female	Male	Non-binary
Apparel	**5.39**	*3.30*	2.74
Banks & Finance	6.78	7.65	8.22
Beauty	**2.64**	*1.48*	0.91
Business	8.96	8.66	9.13
Careers	3.80	3.78	5.94
Cars & Transport	3.64	4.26	2.74
Charity	1.38	1.28	0.46
Credit Cards	4.50	4.42	5.94
Digital News & Media	*1.38*	**1.90**	3.20
Education	1.74	1.85	1.37
Electronics	8.84	9.71	13.24
Fitness & Outdoors	1.32	1.42	0.91
Food & Drink	4.17	3.92	2.28
Gaming	*1.31*	**2.17**	4.57
Home	2.66	2.13	2.28
Home & Auto Insurance	2.83	2.95	1.37
Internet Service	1.97	2.42	0.00
Jewelry	2.65	2.20	2.74
Medications	3.52	2.89	0.91
Mixed Native Ads	5.27	5.33	6.85
Movies & TV	4.43	4.86	5.48
Phone Service	*2.43*	**3.11**	4.11
Software	3.38	3.78	2.28
Travel	4.69	5.11	1.83

Table 6: Percent of ads observed by category, across ethnicities (top 24 categories). Blue / bolded cells indicate a significantly higher proportion than expected, and red / italic cells indicate a significantly lower proportion than expected.

Category	Asian	Black	Latino	Other	White
Apparel	4.51	3.29	4.68	*2.45*	4.71
Banks & Finance	7.03	7.42	6.86	7.30	7.23
Beauty	1.40	**3.20**	**3.00**	1.95	1.88
Business	8.49	7.55	8.83	9.53	9.02
Careers	3.20	3.04	4.31	4.52	3.87
Cars & Transport	5.02	4.09	3.10	4.01	3.85
Charity	1.07	1.12	0.76	0.72	**1.57**
Credit Cards	4.69	4.32	**5.70**	4.40	4.23
Digital News & Media	2.01	1.47	1.50	1.73	1.61
Education	**3.32**	1.98	1.50	1.78	*1.56*
Electronics	10.50	9.92	9.25	8.70	9.01
Fitness & Outdoors	1.55	1.41	1.18	1.06	1.39
Food & Drink	3.41	4.22	3.47	3.51	4.28
Gaming	1.37	1.86	1.45	**3.29**	1.68
Home	2.34	1.57	1.76	2.06	**2.73**
Home & Auto Insurance	2.71	3.55	2.34	3.40	2.85
Internet Service	2.83	2.56	1.94	1.84	2.06
Jewelry	1.67	**3.77**	2.00	2.23	2.47
Medications	*2.16*	2.56	3.94	2.56	3.41
Mixed Native Ads	4.96	5.44	5.49	5.85	5.26
Movies & TV	5.24	5.60	5.44	5.07	*4.18*
Phone Service	2.95	2.50	3.00	3.40	2.66
Software	2.65	3.65	3.36	2.73	3.80
Travel	6.00	4.64	4.89	**6.97**	*4.50*

fixed effects to improve the fit of models, using the REML Akaike information criterion (AIC) (when removing random effects) and maximum likelihood AIC (when removing fixed effects) to measure the goodness of fit. Our final model included all random effects but only included age, gender, and ethnicity as fixed effects. The final model's REML criterion was 42141.3. We show the raw regression estimates in Appendix C.

5.3.1 Demographics: Advertisers Bid Slightly Higher for Women. Overall, we did not see that rendered winning bid values were strongly affected by demographic factors. Bid values for male participants were estimated to be \$0.58 CPM lower than women. However, we did not detect any effect of age or ethnicity on bid values. A linear mixed model analysis of variance indicated a statistically significant effect on bid values of gender ($F(2, 329) = 3.25, p = 0.040$) but no statistically significant effect of ethnicity ($F(4, 277) = 1.589, n.s.$) or age ($F(1, 281) = 0.085, n.s.$). We also did not detect an effect of optional demographic factors (sexuality, income, children) on bid values; these variables did not improve the fit of the model, and were excluded from the final analysis. Figure 4 shows cumulative

Table 7: Percent of ads observed by category, across age ranges of participants (top 24 categories). Blue / bolded cells indicate a significantly higher proportion than expected, and red / italic cells indicate a significantly lower proportion than expected.

Age Range	18-24	25-34	35-44	45-54	55+
Apparel	**5.46**	*3.54*	3.75	5.11	4.96
Banks & Finance	6.46	7.65	6.85	6.84	8.44
Beauty	2.51	2.00	2.34	1.80	1.40
Business	8.49	8.70	8.89	9.17	9.21
Careers	*2.89*	3.43	**4.58**	4.28	4.28
Cars & Transport	3.81	3.96	3.73	3.93	4.33
Charity	*0.88*	1.38	1.55	1.25	1.67
Credit Cards	4.82	4.58	4.12	4.74	4.06
Digital News & Media	1.56	1.72	1.80	1.20	1.67
Education	1.93	1.79	2.13	1.35	1.33
Electronics	9.50	9.51	9.47	8.82	8.42
Fitness & Outdoors	1.17	1.40	1.39	1.33	1.60
Food & Drink	3.60	**5.00**	3.83	3.76	3.53
Gaming	1.57	1.92	**2.30**	1.38	*0.89*
Home	2.24	2.11	2.27	**3.23**	2.83
Home & Auto Insurance	2.47	2.98	3.33	2.66	2.71
Internet Service	1.89	2.35	2.08	1.50	**3.07**
Jewelry	2.85	2.00	*1.78*	**3.78**	2.59
Medications	3.45	2.85	3.46	3.93	2.39
Mixed Native Ads	5.17	5.27	5.51	5.26	5.32
Movies & TV	5.31	5.24	4.43	*3.36*	3.85
Phone Service	2.93	2.61	2.91	1.95	3.24
Software	3.45	3.92	2.95	3.58	4.09
Travel	**5.95**	4.27	4.98	5.26	*3.56*

Table 8: Percent of ads from a category seen by the top 5% of participants, comparing contextually targeted sites to all websites. The amount of individual targeting does not appear to increase when the ad is also targeted at a particular site.

Category	Website	Top 5% on Site	Top 5% Overall
Business	businessinsider.com	14.22	13.06
Careers	businessinsider.com	13.45	16.85
Electronics	phonearena.com	13.45	11.31
Phone Service	phonearena.com	16.78	19.63
Education	usnews.com	17.44	31.39
Banks & Finance	usnews.com	9.97	11.43
Internet Service	speedtest.net	13.67	24.53

distribution functions for bid values by gender and ethnicity, and a scatter plot of age and bid values.

This finding suggests that in the online advertising markets, no particular demographic groups are in substantially higher or lower demand than others, overall. However, this does not mean that people are not being targeted by demographics. A possible explanation is that there is relatively even advertiser demand for people of all ages, ethnicities, and genders, and demand for one demographic group may be canceled by demand for another.

5.3.2 Individual Variation: Winning Bids Differed Between Participants. Though all participants visited the same set of websites, the same number of times, the mean value of the bids seen by each participant ranged from as low as \$1.15, and as high as \$17.35. The median of the mean bid value for each participant ranged was \$4.96 (IQR = 2.34). Participants' median bid values were slightly lower than the mean; the median of the median values was \$4.39 (IQR = 2.35), indicating that outliers skewed means upwards. The mixed model predicts a slightly smaller amount of variation than the raw averages (by controlling for other factors): the median random intercept for participant was -\$0.23, with an IQR of \$1.61. The variance of the participant random effect was 3.266, which explains 10.1% of the variance in the model.

5.3.3 Website: Winning Bid Values Differed Across Websites. Among the 10 websites in our study, we found differences in the winning bid values. Table 9 shows the average winning bid values for each domain. For example, we saw that speedtest.net had the highest mean winning bid at \$9.95 CPM, while ktla.com had the lowest at \$2.44 CPM. Mixed model estimates for the effect of website range from \$3.66 to -\$2.62. The variance of the website random effect was 3.748, accounting for 11.6% of the total variance. Higher winning bids did not appear to correlate with site rank; for example, phonearena.com had the 7th highest site rank, but the 2nd highest mean winning bid value.

These results suggest that some sites are in higher demand from advertisers than others. Perhaps certain sites signal greater intent to certain types of advertisers; e.g. phonearea.com may have higher demand from phone manufacturers and wireless carriers because visitors are more likely to purchase their products, while news sites like ktla.com may provide little information to most advertisers.

5.3.4 Bidders: Winning Bid Values Differed Across Demand Partners. Bid values varied between the demand partners: the ad networks, supply side providers, or other entities placing the bid on the behalf of the advertiser. Table 9 shows the average winning bid values for each demand partner. Based on estimated intercepts from the mixed model that control for other factors, the highest bidding demand partners were Consumable (mean bid value of \$18.04), TrustX (\$9.42), and District M (\$11.29), while the lowest bidders were NoBid (\$6.76), MediaNet (\$5.38), and TripleLift (\$3.36).

To understand the potential underlying reasons for these differences, we investigated the public facing websites of these bidders. Though many made similar claims about the power and reach of their technology, we noticed some qualitative differences. The highest bidders (Consumable and TrustX), focused their message on "premium" content and advertisers, and improving users' experience, meaning they likely work with higher profile websites and brands, involving higher budgets. The lowest bidders (NoBid and MediaNet), described their products in terms of "maximizing revenue" and filling "unfilled and undervalued inventory", suggesting that their strategy is to win auctions where demand is lowest, and bidding at low amounts.

5.3.5 Ad Categories: Winning Bids Differed Across Ad Categories. How did bid values vary for different categories of ads? Table 9 summarizes winning bid price for ads of each category. The ads with the highest bid values came from the "mail & shipping" category, which included US Postal Service ads and home delivery services (\$13.03), beauty (\$7.27), and medications (\$6.95). Categories with low values included charity (\$2.99), healthcare (\$3.86), and live

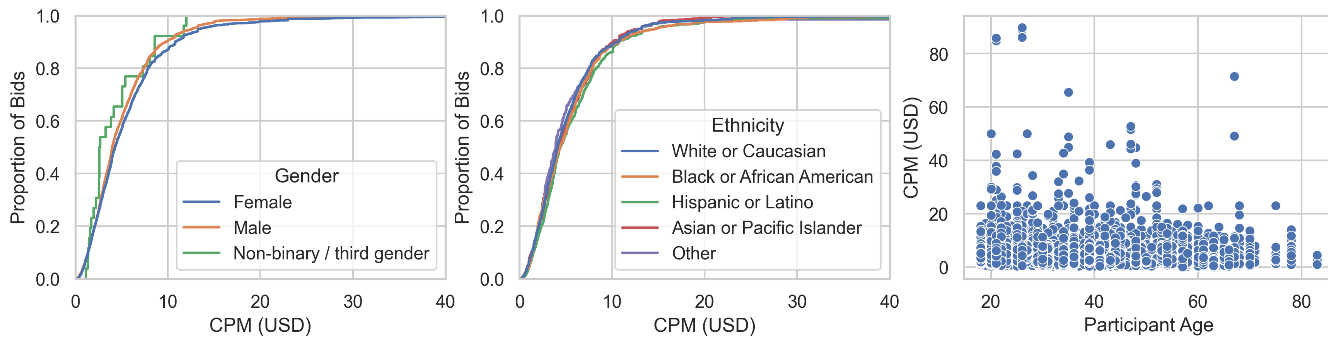

Figure 4: Distribution of bid values across gender, ethnicity, and age. Demographic factors explained little of the differences in bid values; we only detected a significant effect of gender on bid value, with an estimated difference of $0.58 CPM between women and men.

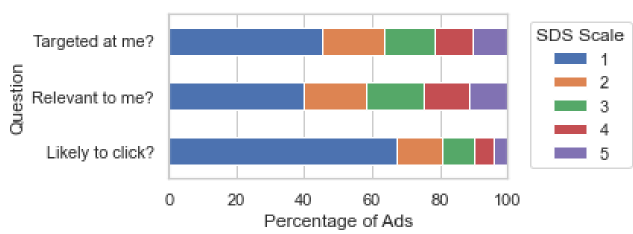

Figure 5: Summary of the targeting perceptions survey. Responses are on a semantic differential scale (i.e., 1 means "not relevant at all", 5 means "very relevant"). Participants said that a majority of ads were not relevant to them, targeted at them, and that they were unlikely to click on them.

events ($3.04). However, the size of the categories suggest that some differences may be due to outlier bids. For example, the mail and shipping category contained only 32 ads (too few to be shown in Table 9), and two outliers with winning bid values over $80, and a standard deviation of 19.61, which suggests that the presence of outliers in a small sample is skewing the overall figures.

5.4 Self-Reported Targeting Perceptions

What proportion of ads did participants themselves perceive as targeted? In this section, we report on results of the self reported targeting perceptions survey. We also investigate whether targeting perceptions correlate with bid values.

Each participant rated a sample of 8 ads that they saw with their perceptions of how targeted each was. We used a deterministic sample of 8 ads with winning bids, uniformly selected across the range of bid values, to ensure we had data on high and low bids for each participant. We received responses for 1746 ads from 286 participants, an average of 6.1 per participant. Some participants were not able to submit responses for all 8 ads for several possible reasons: because the ad screenshots were blank or obscured (215 participants, affecting 449 ads), because they did not receive 8 rendered winning bids in total (16 participants, affecting 61 ads), or because of unknown technical issues with the extension (32 ads).

5.4.1 Most Ads Were Not Relevant to Participants. Figure 5 shows participants' responses to the targeting perceptions survey. Most ads were perceived as not relevant to participants: over 40% of ads received the lowest score of 1 for relevance, targeting and click likelihood, while 10% or less scored the highest score of 5. Comparing the distributions for each question, participants perceived ads as relevant and targeted at similar proportions, but were less likely to click on ads. We also asked participants whether they had previously visited the website of the advertiser or product, which could indicate if the ad was retargeted. Participants responded "Yes" for 18.3% of ads, "No" for 76.6% of ads, and "Not Sure" for 5% of ads. We expected a somewhat even distribution to these responses, because an even number of ads with low and high bid values were sampled, these results still skew towards low relevance, indicating that participants did not perceive much targeting.

5.4.2 Self-Reported Retargeted Ads had Higher Winning Bid Values. Next, we investigate whether participants' targeting perceptions correlate with winning bid values. To determine which factors may be related to bid values, we fit a linear mixed effects model to the subset of 1746 ads with survey responses. Winning bid price was the outcome variable, with fixed effects for perceptions of relevance, targeting, likeliness to click, and retargeting. Additionally, we include the fixed and random effects from the final model in Section 5.3: fixed effects of age, gender, and ethnicity, random intercepts for website, participant, bidder, and ad category. Coefficient estimates are reported in Appendix C.

Ads where participants reported previously visiting the advertiser's site had a median CPM of $4.50 (IQR = $5.08), and ads not perceived as retargeted had a median of $3.90 (IQR = $4.32). A linear mixed model analysis of variance found a statistically significant effect of self-reported visits on winning bid value ($F(2, 1645) = 6.064, p = 0.002$), with an estimated increase of $1.07 for ads with "yes" responses, and $1.45 for "not sure". However, no effect was detected for perceived targeting, relevance, and likelihood to click. Figure 6 shows the CDF for bid values, across participants' responses to whether they visited the advertiser's site.

These findings concur with the findings of Olejnik et al., who found in a crawler-based study that retargeted ads had substantially higher bid values [31].

Table 9: Summary of winning bid values by website, ad category, and demand partner. Estimate refers to the difference from the estimated baseline bid value (random intercept).

	Mean	Std.Dev.	# Ads	Estimate
Website				
speedtest.net	9.95	6.07	508	3.66
businessinsider.com	7.95	6.09	289	2.34
phonearena.com	7.87	3.42	313	0.84
foodnetwork.com	6.03	6.11	873	0.57
weather.com	5.39	5.28	834	-0.17
oxfordlearnersdictionaries.com	5.40	5.75	671	-0.22
fashionista.com	4.88	5.50	369	-1.29
usnews.com	3.83	3.29	589	-1.50
detroitnews.com	4.97	4.96	2033	-1.60
ktla.com	2.44	1.68	638	-2.62
Ad Category (Top 25)				
Medications	6.95	3.17	463	1.14
Beauty	7.27	9.83	184	1.12
Health Insurance	6.37	10.54	73	1.12
Gaming	5.40	6.94	67	0.93
Holiday	6.31	6.47	64	0.67
Jewelry	6.70	6.32	83	0.48
Business	5.80	7.04	428	0.36
Internet Service	6.18	5.95	224	0.29
Banks & Finance	4.19	2.96	366	-0.05
Home	4.63	5.01	177	-0.05
Cars & Transport	5.53	4.03	285	-0.09
Movies & TV	6.43	5.98	293	-0.14
Health Products	4.89	3.14	46	-0.22
Phone Service	6.33	3.96	135	-0.25
Software	4.66	4.68	87	-0.31
Travel	4.94	3.45	131	-0.34
Electronics	5.19	7.61	333	-0.36
Credit Cards	4.92	4.09	172	-0.37
Home & Auto Insurance	4.10	2.88	167	-0.38
Education	4.05	3.58	84	-0.55
Healthcare	3.86	3.70	49	-0.78
Alcohol Tobacco Cannabis	4.21	2.16	70	-0.80
Food & Drink	4.41	3.99	328	-0.86
Apparel	4.90	3.64	326	-0.87
Charity	2.99	2.56	69	-1.89
Demand Partner				
consumable	18.04	20.92	12	5.27
trustx	9.42	12.57	133	3.90
districtm	11.29	7.35	31	1.19
appnexus	7.38	6.62	791	1.15
colossusssp	5.53	6.99	36	0.81
aol	10.85	6.59	25	0.75
pubmatic	7.12	7.25	718	0.55
rubicon	5.87	5.25	684	0.48
sonobi	5.34	2.64	215	0.09
teads	3.71	2.06	376	-0.13
criteo	7.53	5.40	123	-0.22
openx	5.31	3.62	284	-0.55
verizon	2.41	1.29	173	-0.65
kargo	4.51	2.13	157	-0.67
ix	5.05	4.79	803	-0.75
onemobile	4.46	3.29	323	-1.09
pulsepoint	2.43	2.35	30	-1.57
triplelift	3.36	3.47	765	-1.99
medianet	5.38	3.18	132	-2.08
nobid	6.76	2.56	28	-2.33

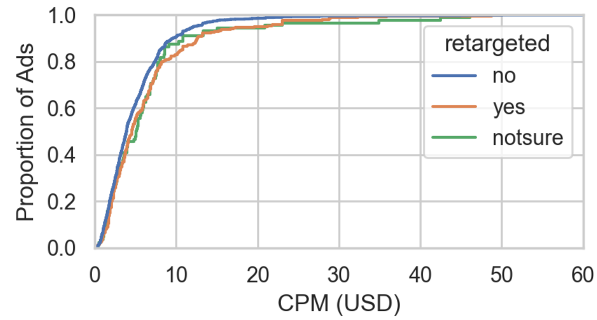

Figure 6: CDF of winning bid value, for ads that participants self reported as retargeted or not retargeted. In aggregate, retargeted ads had higher bid values than non-retargeted ads.

6 DISCUSSION

In our field study, we measured the effect of multiple factors that affect how advertisers bid on and target ads on the web, including site context, user demographics, variation among individual users, and ad networks. By conducting the study in a controlled setting with real users, we show the relative impacts of each of these factors on targeting outcomes. We discuss the implications of these findings, we contextualize the results with prior work in this space, and we describe directions for future work.

6.1 Implications

Website context and retargeting are important as browsing profiles for attributing targeting. Studies of ad targeting, especially crawler-based studies, are often framed using browsing profiles as the primary factor affecting targeting. However, our results demonstrate that other factors, such as the website an ad appears on, and retargeting are also critical for attributing targeting.

Website-based (contextual) targeting was the dominant explanation for targeting in our dataset: websites tended to host ads related to the website's topic, and the distribution of ad categories differed widely between websites, even though the ads were collected from participants with a diverse set of demographics and interests. Retargeting occurred infrequently, but our results suggest that retargeted ads will "override" other targeting factors, because bids for retargeted ads can be over 10x higher than the average.

Based on these results, we propose a heuristic for attributing ad targeting on the web. The factors that affect targeting outcomes, in order of importance, are: (1) previous visits to an advertiser's website (retargeting), (2) the topic of the website the ad is being loaded on, and (3) broadly inferred interests from browsing history.

This suggests that future work on ad targeting on the web should strongly take into account contextual targeting as a factor, and separate contextual targeting from behavioral targeting in their measurements. Additionally, profile based studies should consider attempting to trigger retargeted ads by visiting product pages on e-commerce sites.

Direct targeting of demographic segments is uncommon. Age, gender, and ethnicity had either weak correlations or no correlation with targeting and bid values. We observed no significant differences in how much advertisers bid across demographic groups, and relatively few differences in how ad categories were distributed across demographic groups compared to other factors.

However, prior work has found evidence of targeting based on demographics factors [4, 12, 39]. We propose the following explanations for the apparently low amount of demographic targeting in our dataset: First, Google and other ad networks do not allow direct targeting by race and ethnicity, considering them to be "sensitive categories" [18]. Second, our ad categories may not have been granular enough capture some cases of demographic targeting: for example, we did not consider women's and men's apparel as separate categories. Third, advertisers might not be deliberately targeting by demographic groups — instead, they may be targeting certain interests or audiences (e.g. outdoors enthusiasts, car shoppers), which may be loosely correlated with demographics.

6.2 Comparison to Prior Work

Value of a User. Our study finds higher winning bid values than past studies on real-time bidding and header bidding. We observed a median winning bid value of around $4.16 CPM, which is higher than prior work. Prior crawler-based studies conducted between 2019-2020 measured median bids in header bidding ranging from <$0.10 CPM [33] (including non-winning bids), to $2.00 CPM [11]. RTB studies also found lower winning bids, ranging from $0.36 CPM [31] (2013) to $0.273 CPM [35] (2017). Some methodological factors may explain these differences: the ten sites in our study were relatively high ranked, demand for ads may have been high during our study, due to the December holiday shopping season, and bid values for real users with extensive browsing profiles may be higher than for synthetic profiles or stateless crawls. We also speculate that bid prices are rising over time, which concurs with other recent measurements [32].

Differences in Bid Values. We concur with other results finding that women receive higher bids than men overall, but did not observe statistically significant effect of age [32]. Our finding that self-reported retargeting was associated with substantially higher bids aligns with other studies finding a link between previous visits to sites and higher bid values [31, 33]. Our results on the average bid values of different demand partners differed in rank order differed from the header bidding study of Pachilakis et al. [33], suggesting that bidding behaviors of individual advertisers may not be stable over time or specific collection methodologies.

6.3 Future Work

Are privacy-preserving targeting APIs necessary? The major web browser vendors (Apple, Google, Mozilla) have been considering proposals to limit web tracking through mechanisms like third-party cookies and replace them with more privacy preserving APIs. For targeting, vendors have proposed separate APIs for behavioral targeting (e.g. FLoC, Topics API [15] in Chrome) and retargeting (e.g. FLEDGE [14]). Our work suggests that profile-based behavioral advertising is only one of several factors influencing actual targeting outcomes, alongside retargeting and contextual advertising. This raises the question, how important is behavioral advertising for advertisers? And if it is not as important as other factors, can browsers remove support for it altogether to advance user privacy? Future work should investigate the effectiveness of behavioral advertising in isolation from other targeting methods, to inform whether new APIs for behavioral targeting like Chrome's Topics API are necessary for effective ads.

Are winning bid values related to the quality of ads? What is the economic model behind low-quality, misleading, or other ads that are bad for user experience? Prior work [45] has shown that such ads are common, especially on news websites. Though this work is some of the first to study the content of ads in conjunction with bid values, we does not address the question of ad quality directly, as we did not find many examples of low quality ads with header bidding metadata. Though our data suggests that some SSPs, like NoBid, specialize in filling cheap, low-demand ad slots, we have no data on the incentives for low quality ads themselves. Do they mainly fill low-demand ad slots? Or do they outbid other ads? Future work may require mechanisms besides header bidding to measure the value of these ads.

7 CONCLUSION

To provide transparency on the practices of the online advertising ecosystem, we conducted a field study to measure the influence of individual, demographic, and contextual factors on ad targeting and winning bid values, using data collected from 286 participants on 10 websites. We found that the website an ad appears on, retargeting, and individual behavioral profiles had the most influence on targeting outcomes, while user demographics were at best weakly correlated with targeting. Similarly, we find that differences in bid values are primarily explained by the website and variation among individuals. Our findings suggest that contextual targeting plays a major role in targeting on the web, and that the most valuable signals to advertisers for relevance are the website of the ad slot and previous visits by the user to the advertisers' site. We recommend that browsers and regulators consider these factors when evaluating the necessity of privacy-invasive cross-site tracking and behavioral targeting for effective advertising.

ACKNOWLEDGMENTS

We thank our shepherd, Emiliano De Cristofaro, as well as our anonymous reviewers for their thoughtful feedback on our work. We also thank Ariana Mirian, Mako Hill, and Jamie Morgenstern for providing feedback and advice on earlier versions of this work. We thank Pardis Emami-Naeini and the UW Statistical Consulting Service for their advice on study design and analysis. This work was supported in part by a Consumer Reports Digital Lab Fellowship and the National Science Foundation under grant CNS-2041894 and CNS-156525.

REFERENCES

[1] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., and Diaz, C. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2014), CCS '14, Association for Computing Machinery, p. 674–689.

[2] Acar, G., Juarez, M., Nikiforakis, N., Diaz, C., Gürses, S., Piessens, F., and Preneel, B. Fpdetective: Dusting the web for fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security* (New York, NY, USA, 2013), CCS '13, Association for Computing Machinery, p. 1129–1140.

[3] Ali, M., Goetzen, A., Sapiezynski, P., Redmiles, E., and Mislove, A. All things unequal: Measuring disparity of potentially harmful ads on facebook. In *6th Workshop on Technology and Consumer Protection* (2022), ConPro '22, IEEE.

[4] Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction 3*, CSCW (Nov. 2019).

[5] Aqeel, W., Bhattacherjee, D., Chandrasekaran, B., Godfrey, P. B., Laughlin, G., Maggs, B., and Singla, A. Untangling header bidding lore. In *Passive and Active Measurement* (2020), A. Sperotto, A. Dainotti, and B. Stiller, Eds., Springer International Publishing, pp. 280–297.

[6] Bahrami, P. N., Iqbal, U., and Shafiq, Z. FP-Radar: Longitudinal measurement and early detection of browser fingerprinting. *Proceedings on Privacy Enhancing Technologies 2022*, 2 (2022), 557–577.

[7] Cai, H., Ren, K., Zhang, W., Malialis, K., Wang, J., Yu, Y., and Guo, D. Real-time bidding by reinforcement learning in display advertising. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (New York, NY, USA, 2017), WSDM '17, Association for Computing Machinery, p. 661–670.

[8] Carrascosa, J. M., Mikians, J., Cuevas, R., Erramilli, V., and Laoutaris, N. I always feel like somebody's watching me: Measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies* (New York, NY, USA, 2015), CoNEXT '15, Association for Computing Machinery.

[9] Chen, Y., Berkhin, P., Anderson, B., and Devanur, N. R. Real-time bidding algorithms for performance-based display ad allocation. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2011), Association for Computing Machinery, p. 1307–1315.

[10] Chen, Y., Pavlov, D., and Canny, J. F. Large-scale behavioral targeting. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2009), KDD '09, Association for Computing Machinery, p. 209–218.

[11] Cook, J., Nithyanand, R., and Shafiq, Z. Inferring tracker-advertiser relationships in the online advertising ecosystem using header bidding. *Proceedings on Privacy Enhancing Technologies 2020*, 1 (2020), 65–82.

[12] Datta, A., Tschantz, M. C., and Datta, A. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *ArXiv abs/1408.6491* (2014).

[13] Despotakis, S., Ravi, R., and Sayedi, A. First-price auctions in online display advertising. *Journal of Marketing Research 58*, 5 (2021), 888–907.

[14] Dutton, S. FLEDGE API. https://developer.chrome.com/docs/privacy-sandbox/fledge/, 2022. Google Chrome Developers.

[15] Dutton, S. The Topics API. https://developer.chrome.com/docs/privacy-sandbox/topics/, 2022. Google Chrome Developers.

[16] Englehardt, S., and Narayanan, A. Online tracking: A 1-million-site measurement and analysis. In *ACM Conference on Computer and Communications Security* (2016), CCS '16.

[17] Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior research methods 41* (11 2009), 1149–60.

[18] Google. Personalized advertising. https://support.google.com/adspolicy/answer/143465?hl=en, 2022. Google Ads Policies.

[19] Grootendorst, M. BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics. https://doi.org/10.5281/zenodo.4381785, 2020. v0.7.0.

[20] Imana, B., Korolova, A., and Heidemann, J. Auditing for discrimination in algorithms delivering job ads. In *Proceedings of the Web Conference 2021* (New York, NY, USA, 2021), WWW '21, Association for Computing Machinery, p. 3767–3778.

[21] Iordanou, C., Kourtellis, N., Carrascosa, J. M., Soriente, C., Cuevas, R., and Laoutaris, N. Beyond content analysis: Detecting targeted ads via distributed counting. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies* (New York, NY, USA, 2019), CoNEXT '19, Association for Computing Machinery, p. 110–122.

[22] Iqbal, U., Bahrami, P. N., Trimananda, R., Cui, H., Gamero-Garrido, A., Dubois, D., Choffnes, D., Markopoulou, A., Roesner, F., and Shafiq, Z. Your echos are heard: Tracking, profiling, and ad targeting in the Amazon smart speaker ecosystem. arXiv.

[23] Iqbal, U., Englehardt, S., and Shafiq, Z. Fingerprinting the fingerprinters: Learning to detect browser fingerprinting behaviors. In *2021 IEEE Symposium on Security and Privacy (SP)* (2021), pp. 1143–1161.

[24] Jin, J., Song, C., Li, H., Gai, K., Wang, J., and Zhang, W. Real-time bidding with multi-agent reinforcement learning in display advertising. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (New York, NY, USA, 2018), CIKM '18, Association for Computing Machinery, p. 2193–2201.

[25] Jueckstock, J., Sarker, S., Snyder, P., Beggs, A., Papadopoulos, P., Varvello, M., Livshits, B., and Kapravelos, A. Towards realistic and reproducible web crawl measurements. In *Proceedings of the Web Conference 2021* (New York, NY, USA, 2021), WWW '21, Association for Computing Machinery, p. 80–91.

[26] Lazovich, T., Belli, L., Gonzales, A., Bower, A., Tantipongpipat, U., Lum, K., Huszar, F., and Chowdhury, R. Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics, 2022.

[27] Lecuyer, M., Spahn, R., Spiliopolous, Y., Chaintreau, A., Geambasu, R., and Hsu, D. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2015), CCS '15, Association for Computing Machinery, p. 554–566.

[28] Lerner, A., Simpson, A. K., Kohno, T., and Roesner, F. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *25th USENIX Security Symposium* (2016).

[29] Liu, B., Sheth, A., Weinsberg, U., Chandrashekar, J., and Govindan, R. Adreveal: Improving transparency into online targeted advertising. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks* (New York, NY, USA, 2013), HotNets-XII, Association for Computing Machinery.

[30] Nath, S. Madscope: Characterizing mobile in-app targeted ads. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (New York, NY, USA, 2015), MobiSys '15, Association for Computing Machinery, p. 59–73.

[31] Olejnik, L., Minh-Dung, T., and Castelluccia, C. Selling off privacy at auction. In *Network and Distributed System Security (NDSS) Symposium* (2014).

[32] Pachilakis, M., Papadopoulos, P., Laoutaris, N., Markatos, E. P., and Kourtellis, N. YourAdvalue: Measuring advertising price dynamics without bankrupting user privacy. *Proceedings of the ACM on Measurement and Analysis of Computing Systems 5*, 3 (dec 2021).

[33] Pachilakis, M., Papadopoulos, P., Markatos, E. P., and Kourtellis, N. No more chasing waterfalls: A measurement study of the header bidding ad-ecosystem. In *Proceedings of the Internet Measurement Conference* (New York, NY, USA, 2019), IMC '19, Association for Computing Machinery, p. 280–293.

[34] Panwar, A., Onut, I.-V., and Miller, J. Towards real time contextual advertising. In *Web Information Systems Engineering – WISE 2014* (2014), B. Benatallah, A. Bestavros, Y. Manolopoulos, A. Vakali, and Y. Zhang, Eds.

[35] Papadopoulos, P., Kourtellis, N., Rodriguez, P. R., and Laoutaris, N. If you are not paying for it, you are the product: How much do advertisers pay to reach you? In *Proceedings of the 2017 Internet Measurement Conference* (2017), pp. 142–156.

[36] Parra-Arnau, J., Achara, J. P., and Castelluccia, C. Myadchoices: Bringing transparency and control to online advertising. *ACM Transactions on the Web 11*, 1 (mar 2017).

[37] Ribeiro, F. N., Saha, K., Babaei, M., Henrique, L., Messias, J., Benevenuto, F., Goga, O., Gummadi, K. P., and Redmiles, E. M. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2019), FAccT '19, Association for Computing Machinery, p. 140–149.

[38] Roesner, F., Kohno, T., and Wetherall, D. Detecting and Defending Against Third-Party Tracking on the Web. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)* (2012).

[39] Sweeney, L. Discrimination in online ad delivery. *Communications of the ACM 56*, 5 (2013), 44–54.

[40] United States Census Bureau. 2019 American Community Survey statistics for income, poverty and health insurance available for states and local areas. https://www.census.gov/newsroom/press-releases/2020/acs-1year.html, 2020.

[41] United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs, 2020.

[42] Wang, J., Zhang, W., and Yuan, S. Display advertising with real-time bidding (RTB) and behavioural targeting. *Foundations and Trends in Information Retrieval 11* (2016), 297–435.

[43] Yuan, S., Wang, J., and Zhao, X. Real-time bidding for online advertising: Measurement and analysis. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising* (New York, NY, USA, 2013), ADKDD '13, Association for Computing Machinery.

[44] Zeber, D., Bird, S., Oliveira, C., Rudametkin, W., Segall, I., Wollsén, F., and Lopatka, M. The representativeness of automated web crawls as a surrogate for human browsing. In *Proceedings of The Web Conference 2020* (New York, NY, USA, 2020), Association for Computing Machinery, p. 167–178.

[45] Zeng, E., Kohno, T., and Roesner, F. Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites. In *Workshop on Technology and Consumer Protection* (2020), ConPro '20.

[46] Zeng, E., Wei, M., Gregersen, T., Kohno, T., and Roesner, F. Polls, clickbait, and commemorative $2 bills: Problematic political advertising on news and media websites around the 2020 u.s. elections. In *Proceedings of the 21st ACM Internet Measurement Conference* (New York, NY, USA, 2021), IMC '21, Association for Computing Machinery, p. 507–525.

[47] Zhang, W., Yuan, S., and Wang, J. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2014), KDD '14, Association for Computing Machinery, p. 1077–1086.

[48] Zhao, J., Qiu, G., Guan, Z., Zhao, W., and He, X. Deep reinforcement learning for sponsored search real-time bidding. In *Proceedings of the 24th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining (New York, NY, USA, 2018), KDD '18, Association for Computing Machinery, p. 1021–1030.
[49] Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M., et al. *Mixed effects models and extensions in ecology with R*, vol. 574. Springer, 2009.

A HEADER BIDDING DIAGRAM

Figure 7 illustrates the header bidding process.

B HEADER BIDDING AD CATEGORIES

Table 10: Difference in the size of ad categories between ads with winning bid data, and ads without (15 largest categories shown). The residuals column shows the standardized residuals between the two subsets; residuals larger than ±3.28 indicate significant differences ($p < 0.05$).

	% of Ads with Winning Bid	% of All Other Ads	Residuals
Medications	7.24	1.80	24.19
Internet Service	3.50	1.50	10.78
Food & Drink	5.13	3.10	8.05
Apparel	5.09	3.49	6.10
Cars & Transport	4.45	3.11	5.42
Movies & TV	4.58	3.83	2.79
Banks & Finance	5.72	6.23	-1.54
Business	6.69	7.72	-2.84
Credit Cards	2.69	4.06	-5.21
Software	1.36	3.39	-8.60
Electronics	5.20	8.49	-8.84
Travel	2.05	4.59	-9.28
Careers	0.48	3.83	-13.73
Mixed Native Ads	0.09	5.47	-18.79

We compare the proportion of ads in each category between the subset of ads with rendered winning bids, and all other ads in Table 10. We find that the proportions of certain categories differ substantially while others are approximately equivalent. For example the rendered winning bid dataset has more medication ads (7.24% vs 1.8%), about the same number of banking and finance ads (5.72% vs. 6.23%), and substantially fewer career (0.48% vs. 3.83%) and native ad widgets (0.09% vs. 5.47%). This suggests that the demand partners that advertisers prioritize over header bidding may have qualitatively different ad campaigns in their inventory than the demand partners in header bidding auctions.

C REGRESSION ANALYSIS OF DEMOGRAPHIC EFFECTS ON BID VALUES

Table 11 shows the average winning bid values across age, gender, ethnicity, and income. We used linear mixed models to test whether these demographic factors had a significant effect on bid values in Sections 5.3 and 5.4. Tables 12 and 13 show the fixed effects estimates and random effects structures for those regressions. Overall, we did not find that that demographic factors (the fixed effects) had a significant impact on bid values, with the exception of a small effect of gender.

Table 11: Average CPM for winning bids, across the demographic categories of our participants.

Demographic	Mean	Median	Count
Gender			
Female	3.89	2.78	14127
Male	3.26	2.41	11735
Non-binary	2.26	1.03	171
Age			
18-24	3.71	2.71	6337
25-34	3.41	2.44	6738
35-44	3.57	2.7	6263
45-54	4.10	2.86	3299
55+	3.30	2.45	3396
Ethnicity			
Asian or Pacific Islander	3.23	2.40	2378
Black or African American	3.81	2.66	2118
Hispanic or Latino	4.45	3.14	2093
Other	3.25	2.61	232
White or Caucasian	3.48	2.55	15803
Household Income			
Less than $25,000 per year	3.59	2.66	4602
$25,000-$49,999 per year	3.42	2.51	6671
$50,000-$74,999 per year	3.63	2.56	3953
$75,000-$99,999 per year	3.58	2.63	4218
$100,000-$124,000 per year	3.73	2.70	3337
$125,000-$149,000 per year	3.70	2.51	954
More than $150,000 per year	3.91	2.84	1658

D CASE STUDY: EXTREME OUTLIERS IN BID VALUES

Though the average bid value was $3.55 CPM, we observed many examples of bids an order of magnitude higher, as high as $89.00 CPM. What explains these extremely high bids? In this section, we perform a case study of the ads that we observed in this range, to try to understand what may explain these bid prices. We examine the subset of ads with a winning bid values greater than $20 CPM, which encompasses 127 ads, or the top 1.8% of ads by bid value. This subset of ads came from 66 participants.

Outliers were distributed among individuals roughly evenly; the data is not dominated by one or more individuals. The mean number of outliers seen by an individual was 1.9, 92% participants saw 1-3 outliers, making up 81% of the data, and five participants had 8, 7, 5, 4, and 4 ads.

Individual Examples. First, we describe example ads from individual participants, to illustrate what these outliers ads look like.

Participant 639 had the highest bid values in the dataset, with two ads with bids of $89.75 and $89.09 CPM each. Both ads appeared on oxfordlearnersdictionaries.com, and were ads from Microsoft for Intel-based laptops with Windows 11.

Participant 719 had 7 ads in the outlier subset, with values ranging from $44.96 to $65.57 CPM. All ads were for the same product —

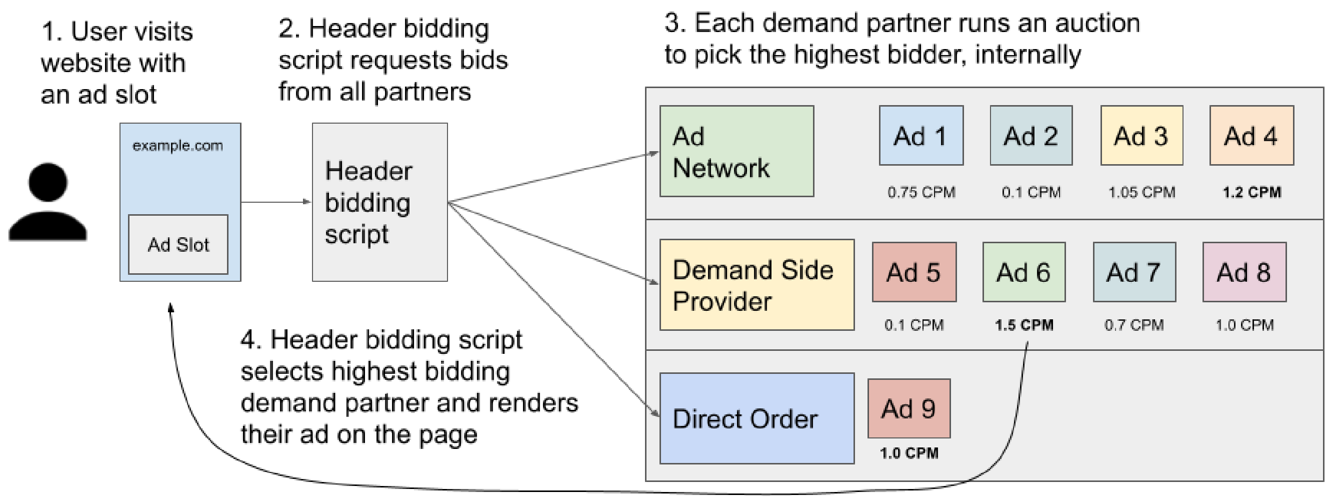

Figure 7: A diagram of a header bidding auction.

Table 12: Fixed effects estimates and random effects structures for a linear mixed model with winning bid values as the outcome variable, fixed effects of age, gender and ethnicity, and random effects for website, individuals, bidder, and ad category. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom. Male participants received slightly lower winning bid values (-$0.58 CPM). 38% of the variance that demographics did not account for are explained by variation in websites, bidders, individual participants, and ad categories, though 62% of the variance remains unexplained.

Fixed Effects Effect	Estimate	Std. Error	t	Pr(>|t|)
Intercept	5.903	0.889	6.639	>0.000***
Age	0.003	0.010	0.291	0.771
Gender — Male	-0.582	0.250	-2.324	0.021*
Gender — Nonbinary	-2.068	1.716	-1.205	0.229
Ethnicity — Asian	-0.231	0.427	-0.542	0.588
Ethnicity — Black	0.315	0.424	0.743	0.458
Ethnicity — Latino	0.746	0.392	1.900	0.059
Ethnicity — Other	0.807	0.567	1.424	0.156
Random Effects Groups	Effect	Variance	Std.Dev.	
Website	Intercept	3.748	1.936	
Bidder	Intercept	3.639	1.908	
Participant	Intercept	3.255	1.807	
Ad Category	Intercept	1.719	1.311	
Residual		19.977	4.470	

Table 13: Fixed effects estimates and random effects structures for a linear mixed model with winning bid values as the outcome variable, and fixed effects of demographic factors and targeting perceptions. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom. Ads that participants perceived as retargeted had higher winning bid values (+$1.36 CPM).

Fixed Effects Effect	Estimate	Std. Error	t	Pr(>|t|)
Intercept	5.998	1.058	5.668	<0.001***
Age	-0.008	0.012	-0.676	0.500
Gender — Male	-0.391	0.300	-1.306	0.193
Gender — Nonbinary	-1.857	2.368	-0.784	0.433
Ethnicity — Asian	-0.483	0.510	-0.947	0.345
Ethnicity — Black	0.321	0.505	0.636	0.525
Ethnicity — Latino	1.213	0.484	2.506	0.013*
Ethnicity — Other	-0.707	0.667	-1.060	0.290
Retargeted — Yes	1.074	0.386	2.783	0.005**
Retargeted — Not Sure	1.447	0.566	2.557	0.011*
Perceived Relevance	-0.073	0.163	-0.451	0.652
Perceived Targeting	0.239	0.163	1.468	0.142
Likely to Click	-0.212	0.157	-1.354	0.176
Random Effects Groups	Effect	Variance	Std.Dev.	
Website	Intercept	5.891	2.427	
Bidder Code	Intercept	2.587	1.608	
Participant	Intercept	2.307	1.519	
Ad Category	Intercept	1.530	1.237	
Residual		22.583	4.752	

a perfume from Yves Saint Laurent — and all appeared on detroitnews.com. The participant reported that the ad was something they visited the website for previously, and responded with the maximum score for targeting perception, relevance, and likeliness to click. These pieces of evidence strongly suggest that these ads were targeted at the particular individual.

Participant 535 had four ads with bids ranging from $44.31 to $52.80, all appearing on foodnetwork.com, and all from SurveyMonkey, an online survey platform. The participant reported that

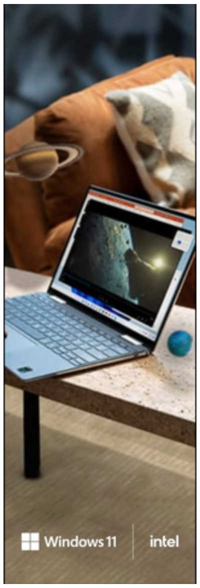

Figure 8: Ad seen by Participant 639 two times. This is the highest valued ad in our dataset, at $89.75 CPM, or almost $0.09 for a single impression.

Figure 9: Ad seen by Participant 719 seven times, with bid values $44.96-$65.57 CPM.

they hadn't been to the SurveyMonkey site before, and only rated it with a 2 for perceived relevance, targeting, and likeliness to click.

Participant 414 had four ads with bids ranging from $21.74 to $31.00. All ads were from Jewelry Television, a TV channel specializing in selling jewelry, three appearing on businessinsider.com and one appearing on speedtest.net. The participant reported going to this site in the past, and scored the relevance, targeting, and likeliness to click 4, 5, and 4.

Targeting Survey Responses to Outliers. Participants perceived the ads in this subset to be more targeted than the remainder of the dataset, but not overwhelmingly so. 40 of 127 ads had relevance survey responses. The average SDS scores for relevance, targeting perception, and likelihood to click were 2.65, 2.65, and 1.73 respectively, compared to 2.36, 2.22, and 1.66 for all other ads (scores range from 1-5). 40.0% of participants said they had visited the website of the ad previously, compared to 14.1% for ads outside this subset.

Though these values suggest that ads with significantly higher bid value are more likely to be perceived to be targeted by participants than others, around half ads in the dataset are still not seen as targeted. Because the data is self-reported, we cannot know for sure whether this is because the ads were not targeting individuals, or if they were simply poorly targeted for their actual interests.

Repeat Ads. In many cases, the same advertiser would show multiple high-value ads to the same participant. We manually inspected the advertiser of these ads, and found that 24 of 33 participants who received more than one outlier received multiple ads from the same advertiser. Often times, these repeat ads appeared on the same website.

Demographics. The subset of participants in the outlier subset were skewed younger and more female than the overall sample of participants. 66% of participants were female; 60% were white; 30% were aged 18-24, 29% were aged 25-34, 21% were aged 35-44, 11% were aged 45-54, and 9% were aged 55+.

Website and Ad Category. Outliers appeared on some sites more than others. weather.com, speedtest.net, and detroitnews.com, hosted 39, 33, and 18 ads each, while fashionista.com, phonearena.com, and usnews.com only hosted 2 ads each. Outliers cover a range of topics: for example, beauty (7), business (11), electronics (6) gaming (3), health insurance (4), home (8) movies and TV (12, the maximum), etc. No particular category is notably overrepresented.

Demand Partner. We observed Pubmatic and Rubicon had substantially more ads in the outlier subset than all other demand partners. 43 ads were from Rubicon, (34%), and 40 were from Pubmatic (31%). The remaining demand partners had between 1 and 9 ads in the subset. This suggests that these two demand partners are more aggressive in their bidding strategies.

E DATA COLLECTION EXTENSION SCREENSHOTS

Figures 10-14 show screenshots of the user interface of the browser extension that participants used to collect data. Ads and website content are blurred for copyright considerations.

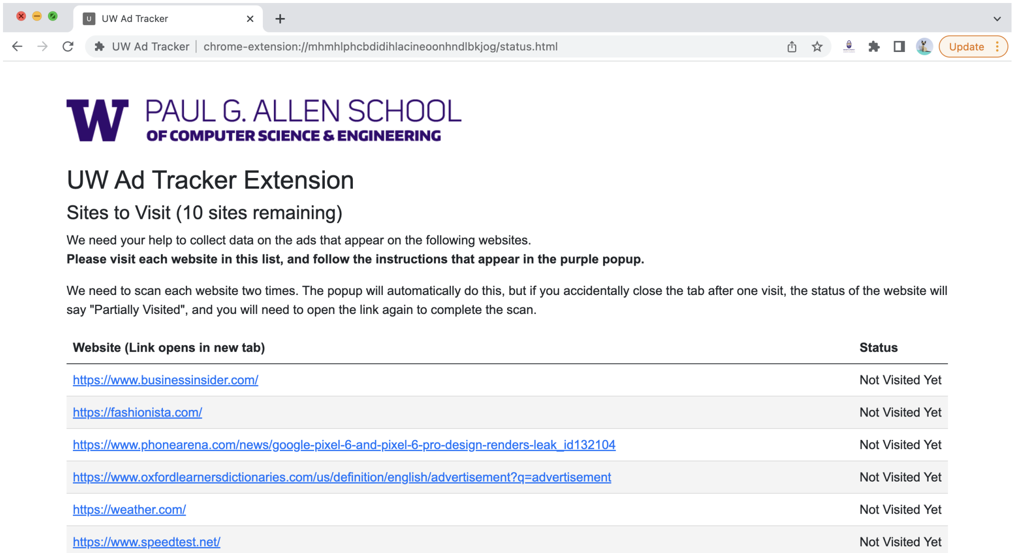

Figure 10: After registering the data collection extension, participants are instructed to visit all 10 websites in this list.

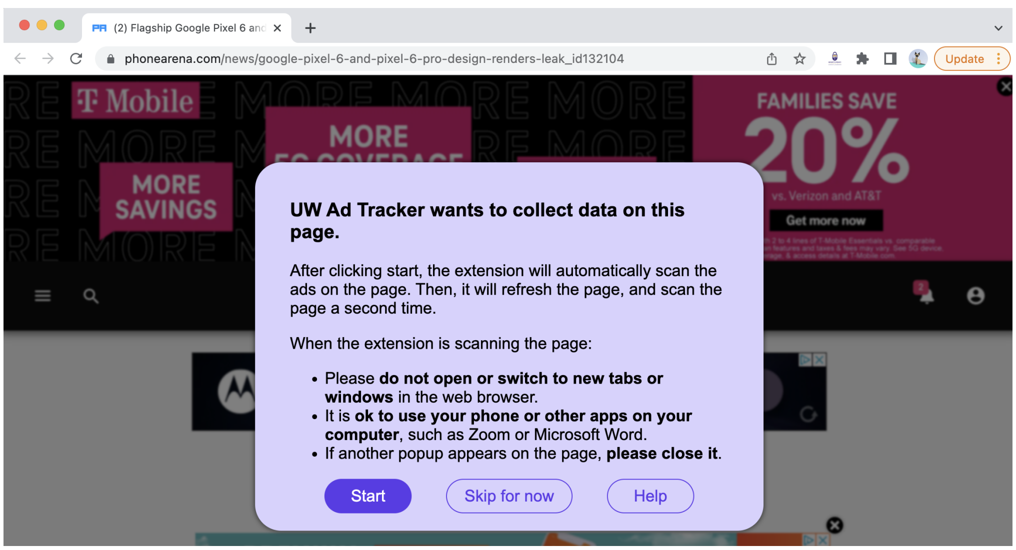

Figure 11: On visiting a site from the list, participants are asked for permission to collect data.

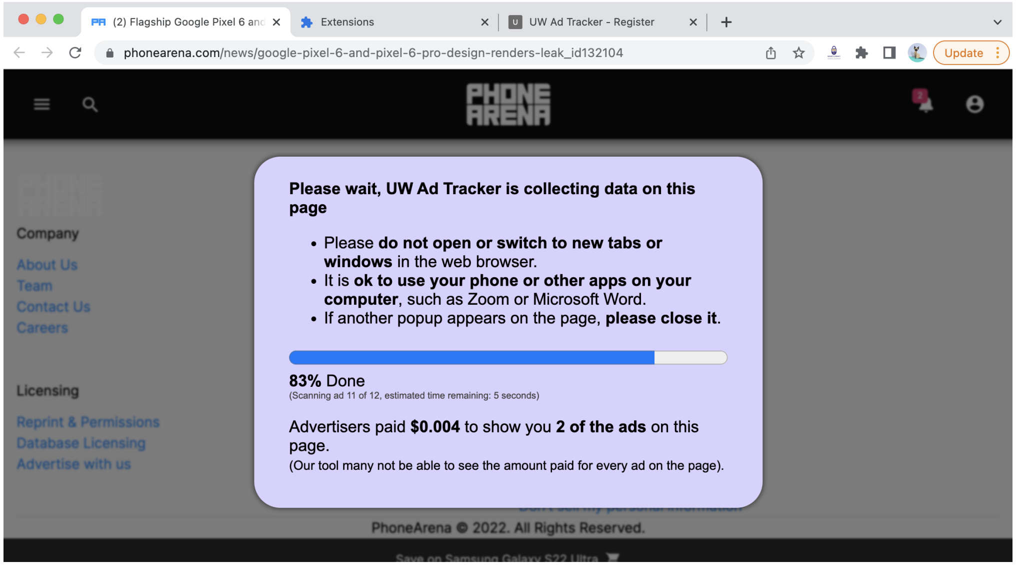

Figure 12: The extension scans the page from top to bottom, one ad at a time. During this time the participant is instructed to not navigate from the page or open other tabs, which interferes with the screenshot process.

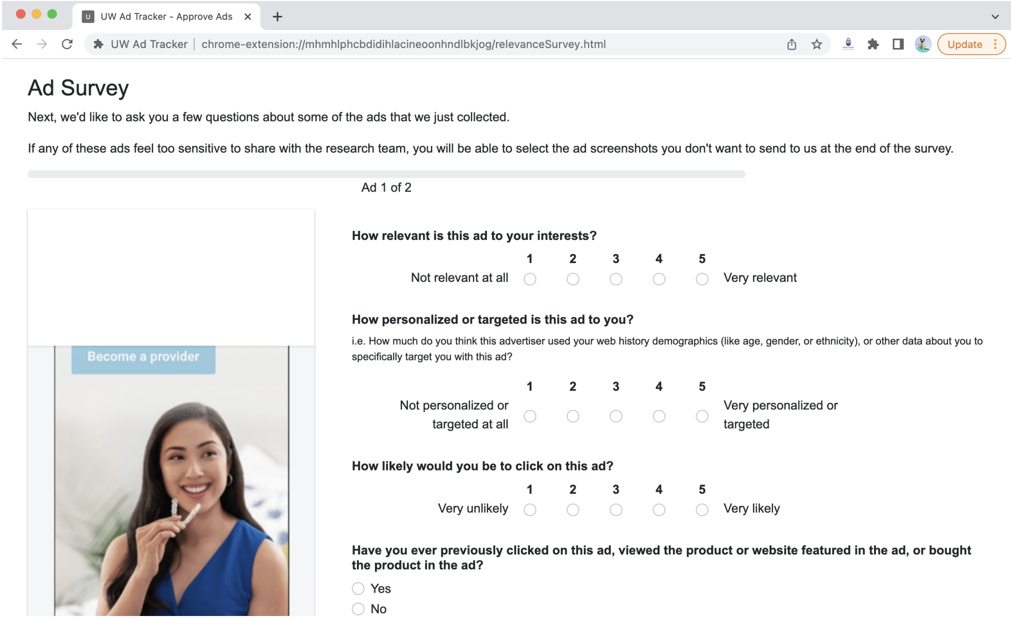

Figure 13: After all data is collected, for a sample of their ads, participants are asked about how targeted they perceived the ad to be.

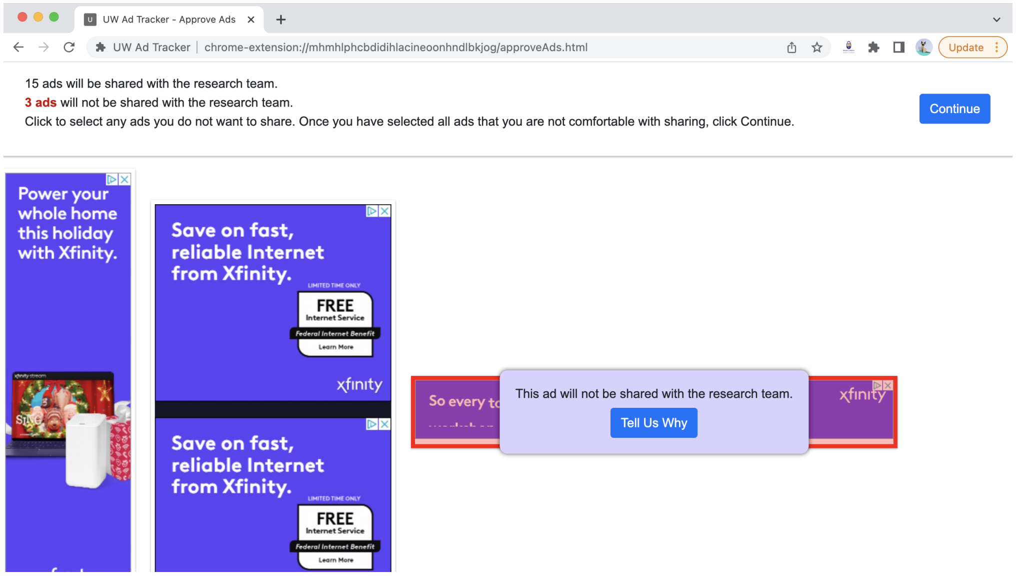

Figure 14: Lastly, participants can opt out of sending any screenshots that they did not want to share with us.