What Factors Affect Targeting and Bids in Online Advertising?

A Field Measurement Study

Eric Zeng
Postdoctoral Researcher
Carnegie Mellon University
https://ericwzeng.com

In collaboration with:
Rachel McAmis, Tadayoshi Kohno,
Franziska Roesner
University of Washington

Targeted advertising is very opaque

Little public data on how ads are targeted is available

Makes informed decision making on privacy difficult

Images: Vectors Market and Freepik via flaticon.com

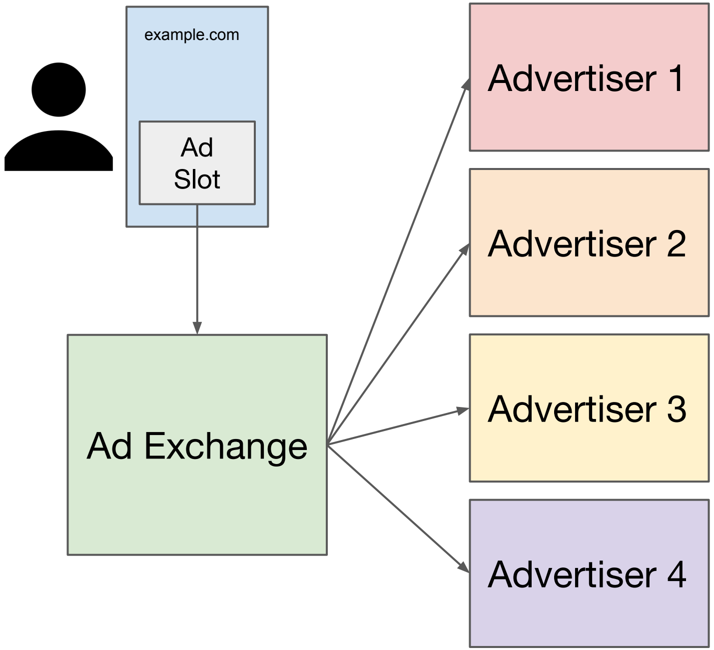
Our study: basic measurements of targeted advertising on the web

- How **prevalent** are behavioral targeting and contextual targeting on the web?
  - Behavioral targeting: targeting of individual users based on interests inferred from browsing behavior
  - Contextual targeting: targeting based on the website the ad appears on
- How do ads differ across **demographic groups** due to behavioral targeting?

3

Measuring bid values in header bidding ad auctions

- **Ad auction:** advertisers bid to place an their ad on a web page/app, conducted in real time for each ad each individual user loads
- **Header bidding:** meta-auction between multiple ad networks, often in the browser



Measurements of bid values in header bidding ad auctions

- Bid values can help reveal which signals advertisers find valuable for targeting
- **How much do advertisers bid** to place ads on the web?
- How do **individual, demographic, and contextual factors** affect bid values?

Introduction

Study Design and Methodology

Results – Ad Targeting

Results – Winning Bid Values

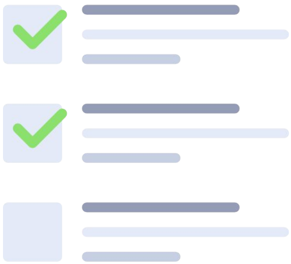
Discussion

Measurement Goals

Measure individual, demographic, and contextual factors in targeted advertising

Field study: collect ads from real users

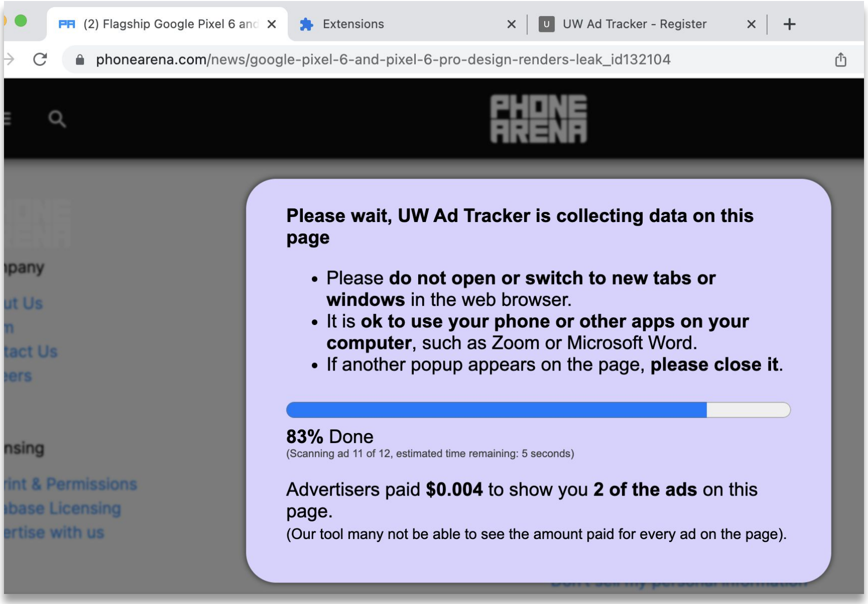
Demographically representative sample (in the U.S)



Control for website effects

Chrome extension for data collection

1. Detects ads on page using EasyList
2. Takes a screenshot of each ad
3. Extracts winning bid values for each ad from header bidding scripts (prebid.js)
4. Auto refreshes page



Screenshot of the browser extension used by participants

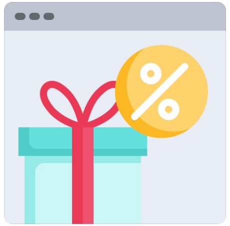
Field Study Protocol

- IRB approved study
- Recruited participants via Prolific
- Part 1: Pre-Screening Survey (n=1460)
    - Participants provided demographic information
    - We screened out ad blocker users, stratified by age/gender/ethnicity
- Part 2: Extension Study (n=286)
    - Install browser extension
    - Visit list of 10 websites
    - Survey + data exclusion

5

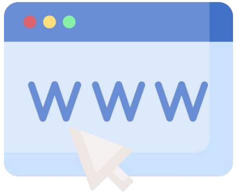
Data analysis

Winning bid value
(some of the time)



Screenshot of ad

Demographic
characteristics



Website ad
appeared on

Extract **ad category** from screenshot

- OCR
- Topic modeling
- Manual auditing of clusters

Analysis techniques

- Targeting: analyze distribution of ad categories
- Bid values: model using linear mixed regressions

Dataset overview

- 41,032 ads
  (143.5 ads / participant)
- 10 websites
  - All used prebid.js
  - Spans a variety of topics and popularity (in Tranco top 10k)
- 52 categories of ads
  - e.g. apparel, healthcare, electronics, travel

- businessinsider.com
- weather.com
- speedtest.net
- usnews.com
- foodnetwork.com
- detroitnews.com
- ktla.com
- phonearena.com
- fashionista.com
- oxfordlearnersdictionary
  .com

Limitations

- Small sample size
  - Data collected from only 10 websites
  - Some demographic segments are small
- Header bidding data is incomplete
  - Websites often ignored winner - only 7,117 ads were “rendered”
- Targeting analysis is limited to correlations
  - No ground truth on targeting parameters

Introduction

Study Design and Methodology

Results – Ad Targeting

Results – Winning Bid Values

Discussion

Clear contextual targeting on some sites

Website	Top categories	% of ads on site
businessinsider.com	B2B Products Careers Credit Cards	26% 21% 13%
phonearena.com	Electronics Phone Service Software	35% 14% 14%
weather.com	Medications Food and Drink	8% 7%
oxfordlearners dictionary.com	B2B products Apparel	15% 10%

Top categories make up large % of ads + match site topic

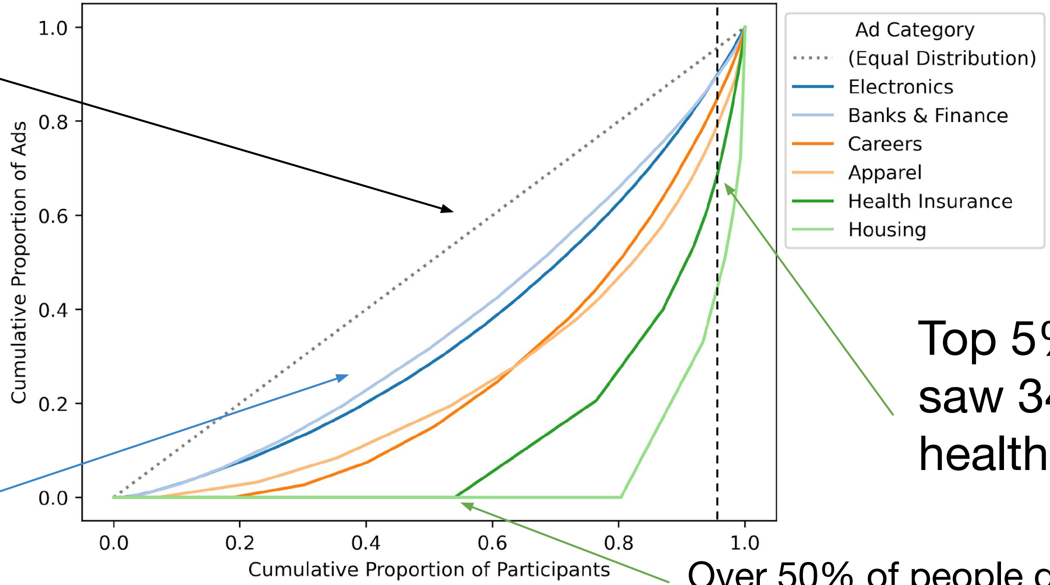
Top categories are smaller, not relevant to site

Behavioral targeting is evident in individuals

Lorenz curve – distribution of ads across individuals

If ads were distributed equally, the line would be **straight**

Everyone saw a similar number of electronics ads and finance ads



Top 5% of people saw 34% of all health insurance ads

Over 50% of people did not see **any** health insurance ads

Behavioral targeting by demographics is less clear

Gender (women vs. men)

- ⬆ Apparel +2.1%
- ⬆ Beauty +1.5%
- ⬇ Gaming -0.9%

Ethnicity (vs. even distribution)

- ⬇ White: Movies and TV -0.4%
- ⬆ Asian: Education +1.5%
- ⬆ Black: Jewelry +1.3%

Age (vs. even distribution)

- ⬆ 45-54: Jewelry +1.4%
- ⬆ 25-34: Food and Drink +0.9%
- ⬇ 18-24: Careers -0.9%

9-16% of ad categories were over- or under-represented across demographic groups

Introduction

Study Design and Methodology

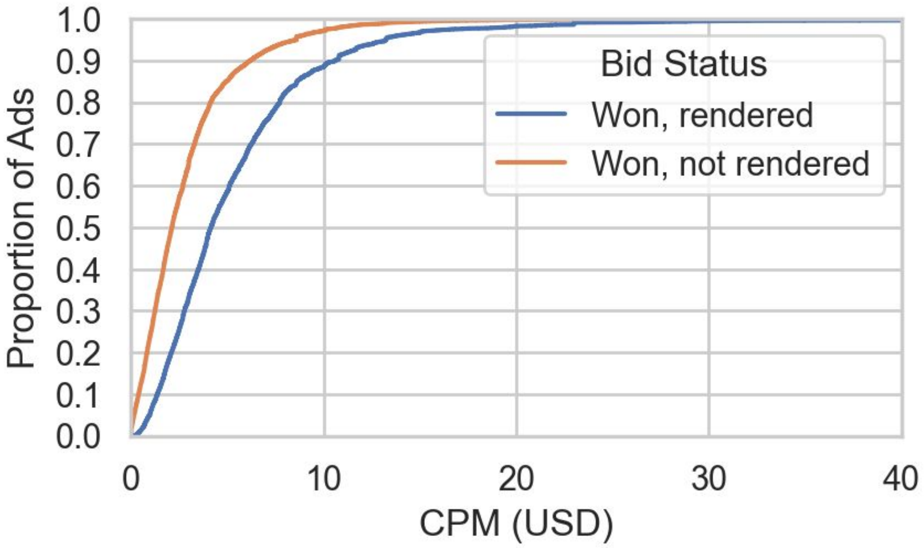
Results – Ad Targeting

Results – Winning Bid Values

Discussion

Bid value summary

- Average winning bid value:
  - Mean: $5.47 CPM
  - Median: $4.16 CPM
- Winners ignored by website had lower bids



(Bid values are denoted in CPM – cost per 1000 impressions)

Winning bid values differ across

across ad categories

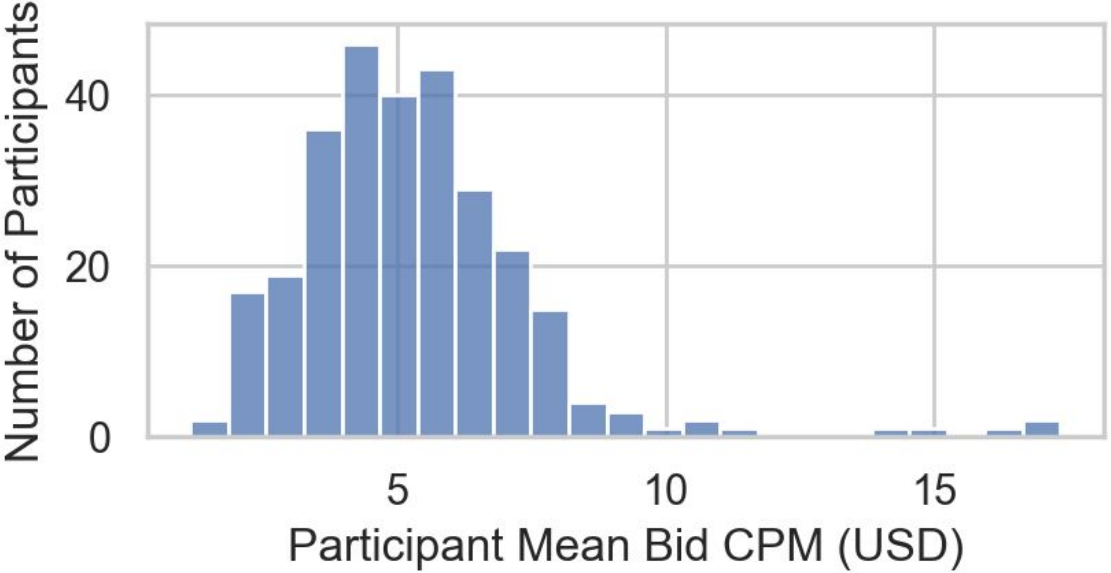
Ad Category	Avg. Bid	Estimated Intercept
Medications	$6.95	+$1.14
Beauty	$7.27	+$1.12
Credit Cards	$4.92	-$0.37
Healthcare	$3.86	-$0.78
Charity	$2.99	-$1.89

and between websites

Website	Avg. Bid	Estimated Intercept
speedtest.net	$9.95	+$3.66
businessinsider.com	$7.95	+$2.34
foodnetwork.com	$6.03	+$0.57
weather.com	$5.39	-$0.17
ktla.com	$2.44	-$2.62

(Bid values are denoted in CPM – cost per 1000 impressions)

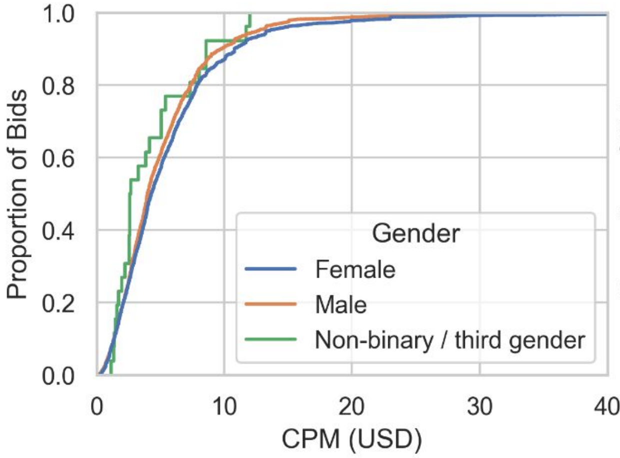
Winning bid values vary between individuals…

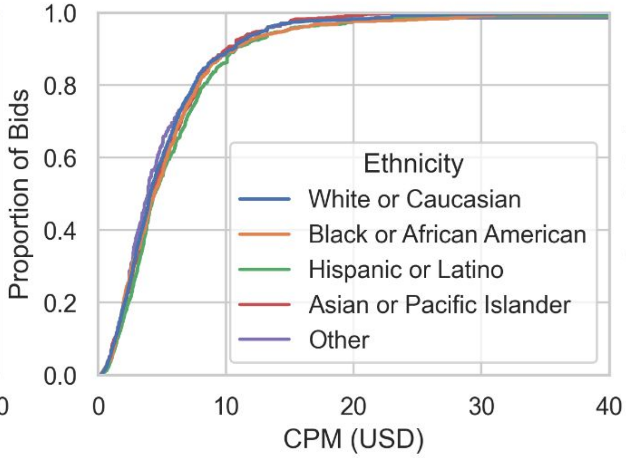


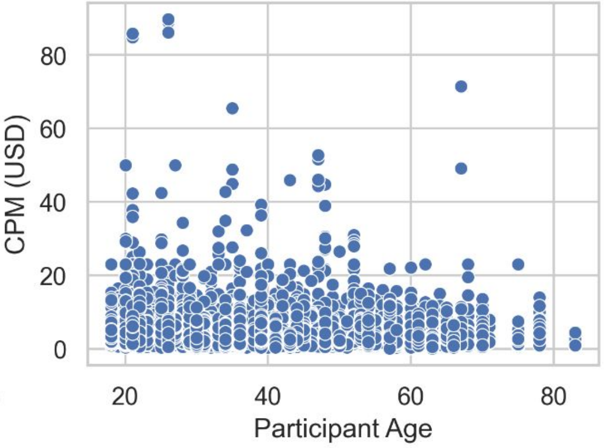
Mean	$4.96
Median	$4.39
IQR	$2.35

(Bid values are denoted in CPM – cost per 1000 impressions)

…but do not appear to differ across demographic groups





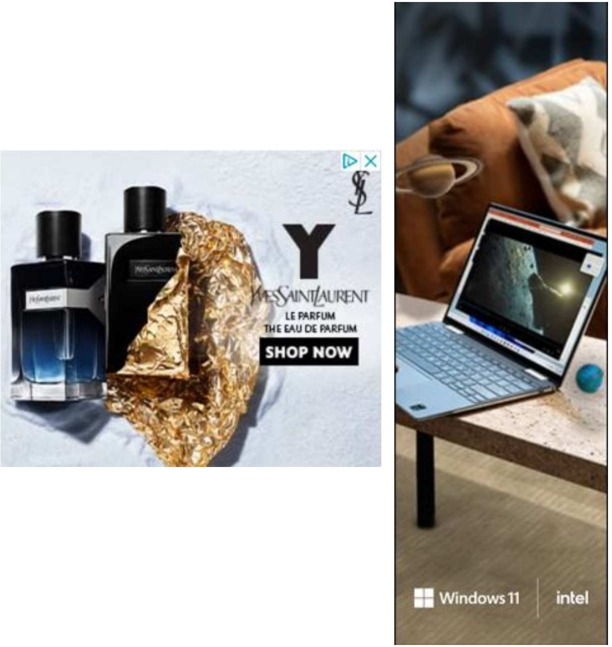


(Bid values are denoted in CPM – cost per 1000 impressions)

High bid values indicate retargeting

Retargeted ads: when you visit a site, and get ads from that site later

- 18% of ads may have been retargeted (participant self-report)
- Bids for (likely) retargeted ads were $1.07 more than others
- Outlier values: $52.80-$89.75 CPM



Ads with the highest bids in our dataset.

Introduction
Study Design and Methodology
Results – Ad Targeting
Results – Header Bidding
Discussion

Discussion

- Alternatives to behavioral targeting on the web are prevalent, and valued by advertisers
  - What would a web with only contextual targeting and retargeting look like? Do we need Google’s FLoC/Topics?
- Demographic disparities in targeting are hard to detect
- Need more transparency from ad tech

Thanks for listening!

Eric Zeng
Postdoctoral Researcher
Carnegie Mellon University

Contact

ericzeng@cmu.edu
ericwzeng.com


Rachel McAmis


Yoshi Kohno


Franzi Roesner

Carnegie Mellon University
CyLab
Security and Privacy Institute

SECURITY & PRIVACY
RESEARCH LAB
UNIVERSITY of WASHINGTON

W PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING