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
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?
Measuring bid values in header bidding ad auctions

- **Ad auction:** advertisers bid to place their ad on a webpage/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?
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
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
Data analysis

Winning bid value (some of the time)

Demographic characteristics

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

Web sites:
- 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

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

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

Top categories are smaller, not relevant to site
Behavioral targeting is evident in 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 and between websites

<table>
<thead>
<tr>
<th>Ad Category</th>
<th>Avg. Bid</th>
<th>Estimated Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medications</td>
<td>$6.95</td>
<td>+$1.14</td>
</tr>
<tr>
<td>Beauty</td>
<td>$7.27</td>
<td>+$1.12</td>
</tr>
<tr>
<td>Credit Cards</td>
<td>$4.92</td>
<td>-$0.37</td>
</tr>
<tr>
<td>Healthcare</td>
<td>$3.86</td>
<td>-$0.78</td>
</tr>
<tr>
<td>Charity</td>
<td>$2.99</td>
<td>-$1.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Website</th>
<th>Avg. Bid</th>
<th>Estimated Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>speedtest.net</td>
<td>$9.95</td>
<td>+$3.66</td>
</tr>
<tr>
<td>businessinsider.com</td>
<td>$7.95</td>
<td>+$2.34</td>
</tr>
<tr>
<td>foodnetwork.com</td>
<td>$6.03</td>
<td>+$0.57</td>
</tr>
<tr>
<td>weather.com</td>
<td>$5.39</td>
<td>-$0.17</td>
</tr>
<tr>
<td>ktla.com</td>
<td>$2.44</td>
<td>-$2.62</td>
</tr>
</tbody>
</table>

(Bid values are denoted in CPM – cost per 1000 impressions)
Winning bid values vary between individuals…

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

Mean $4.96
Median $4.39
IQR $2.35
...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
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