(Do Not) Track Me Sometimes: Users’ Contextual Preferences for Web Tracking

William Melicher, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu*, and Pedro Giovanni Leon

Carnegie Mellon

* Qualcomm®
What Is Online Tracking?

*Cookies* are small tokens that store website state

- Used for: logging in, shopping carts, **tracking**

User requests web page

User

![Cookie Image]

Your user ID is 1234

1st Party
What Is Online Tracking?

Later...

User loads another web page on the same domain

My user ID is 1234

1st Party
What Is Online Tracking?

Later...

User loads another web page on the same domain

My user ID is 1234

Load an ad

Ad for CNN; ID for an advertiser: 5678

Ad

1st Party

3rd Party
What do experts think about online tracking?

**Proponents say:**
- Targeted (better) ads, customized content, social widgets, shopping recommendations
- Revenue used to provide free services online

**Opponents say:**
- Privacy concerns
- Third parties can build detailed profiles about users
- Can happen without users’ knowledge
But what do *users* think?
Current Understanding of Users’ Views

- 65% to 79% have serious privacy concerns
- Users’ preferences are complex
- Prior work: hypothetical situations

How do you feel about tracking…

…on a shopping website?

vs

…when you were shopping for heartburn medicine on Thursday on amazon.com?
Research Questions

• Users’ tracking preferences in the context of their own web history:
  – What harms and benefits do users care about?
  – What situational factors affect users' comfort with tracking?

• Do current tools address users’ needs?

• How can we improve current tools?
Interview Methodology

- 35 semi-structured interviews using participants’ browsing history
- $15 for an interview lasting 1 hour
- Craigslist, posters, university research participant pool
Methodology: Interviews

Send filtered web history

Prepare interview

Conduct interview

• Variety of situations:
  – News, weather, shopping, search, financial services, etc.
  – 1\textsuperscript{st} and 3\textsuperscript{rd} party tracking

• General and situational preferences
Methodology: Example Situation

• Benefits of tracking?
• Harms of tracking?
• Are you comfortable with tracking?

1. nytimes.com
The New York Times - Breaking News on Wed, Jan 14 07:05 PM
Analysis

• Researchers collaboratively developed codebook

• 2 coders independently coded a test set and discussed differences

• Coders independently coded the entire set
Results

• Perceived outcomes of tracking
  – Perceived as harmful or beneficial
  – Overt or hidden

• Situational factors
Example Perceived Outcomes: Overt

• Targeted ads
  – Beneficial: more useful, relevant
  – Harmful: annoying, others might see

• Feel “stalked”

• Customized websites
  – Beneficial: saves time, more relevant
  – Harmful: “filter bubble”

• Possible legal repercussions

% participants

- 74
- 60
- 69
- 31
- 14
- 23
Example Perceived Outcomes: Hidden

• Company revenue
  – Beneficial: provides for free services
  – Harmful: feel used by companies

• Price discrimination
  – Beneficial: special sales, coupons
  – Harmful: maybe higher prices

• Data linked to identity
  – Harmful: privacy invasive

% participants
31  23
31  14
34
Outcomes vs. Comfort

- Perceived harms/benefits $\not\Rightarrow$ comfort
- Less comfortable with harms
- Hidden outcomes $\rightarrow$ least comfortable
Situational Preferences

What about specific page visits made users more or less comfortable?

- Sensitive contexts: less comfortable with third party tracking than first
- What kind of information is tracked
- Sharing with other 1\textsuperscript{st} parties
- Trust in the tracking party
- Lack of awareness of tracking
- Lack of consent to tracking
- Visit frequency to website
Tool Evaluation

• Use findings from interviews to evaluate tools
  – Do tools limit perceived harms of tracking?
  – Do tools allow benefits from tracking?
  – Do tools have selective controls based on situational factors users care about?
Do Tools Meet Needs: Outcomes
# Do Tools Meet Needs: Outcomes

<table>
<thead>
<tr>
<th>None</th>
<th>Harms</th>
<th>Benefits and harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✔</td>
<td>✓ ✔</td>
</tr>
<tr>
<td>× × × × × ×</td>
<td>× ×</td>
<td>× ×</td>
</tr>
<tr>
<td>×</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>×</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Ads**: ✔ ✔ ✔ ✔ ✔ ✔
- **Price Discrimination**: × × × × × × × ×
- **Customization**: × ✔ × × × × ✔
- **Feel “stalked”**: ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔
- **Revenue**: ✔ ✔ ✔ ✔ × × ×
- **Legal action**: × × × × × × ×
- **Linked to ID**: × ✔ ✔ × × × ✔
## Selective Tracking Controls

<table>
<thead>
<tr>
<th>Situational Factor</th>
<th>ABP</th>
<th>Ghost</th>
<th>Brave</th>
<th>Tor</th>
<th>Adblock Plus</th>
<th>Cookies</th>
<th>Mask</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracker is trusted</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Lack of awareness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Lack of consent</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Sharing with 1&lt;sup&gt;st&lt;/sup&gt; parties</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Visit frequency</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has personal info</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has social info</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has search info</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has shopping info</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has financial info</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has correspondence</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>No volunteered info</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
Current Tools

✓ Adequately address perceived harms

✗ Do not allow benefits

✗ Provide few controls based on situational factors
Is it feasible to predict the user’s preference for tracking?
Situational Preference Prediction

• Use machine learning methods to classify websites according to preference
• Experimented with:
  – AdaBoost, SVM, Generalized Linear Effects Models
Situational Preference Prediction

• Use machine learning methods to classify websites according to preference

• Experimented with:
  – AdaBoost, SVM, Generalized Linear Effects Models
Ideal
Prediction Accuracy

% of good tracking allowed

% of bad tracking allowed
(Do Not) Track Me Sometimes

- Explored users’ *in-context* preferences
  - Based on actual browsing history
  - Found outcomes, situational factors that matter
- Evaluated current tools
  - Tools don’t adequately address users’ needs
- Hope for automated preference enforcement

*William Melicher*, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu*, and Pedro Giovanni Leon