ONLINE TRACKING AND INFERENCING: USING TRANSPARENCY TO OBTAIN USER PERSPECTIVES

Michelle Mazurek, University of Maryland

Tracking and inferencing are pervasive

It looks like you're using an ad-blocker!

Business Insider is an advertising supported site and we noticed you have ad-blocking enabled.

sometimes useful

often creepy



generally poorly understood

If we want:

More uptake of PETs
 More effective regulation
 Practices that respect users' wishes

Occupied property please respect privacy

Then we need:

- More understanding
- More salience
- Less hypothetical imagining!

Our work

Build tools for transparency of tracking/inferencing
 Ground understanding in real, contextual data
 Measure preferences without hypotheticals





Transparency in practice

• Browser extension: web tracking/inferencing

• Twitter data downloads: behavioral advertising Security 2020

• Improving transparency/access going forward

* Many slides designed by Miranda Wei

SOUPS 2021



User-facing tracking controls



built-in browser defenses





But these aren't sufficient

- Some tools show trackers on a given page, but not what this means for you
- Dashboards are generally vague or incomplete
- Instead: Which trackers made which inferences?
 Based on which browsing activities?

Our tool



How it works



Topic modeling

- Use Google ad categories as decision classes
- Use Wikipedia to train a model
- Create a plausible topic match per website visited
- Fully client-side data collection





WIKIPEDIA The Free Encyclopedia







nterests. Based on the pages you visited, Tracking Transparency's simulations have ide he chart below shows the interests suggested by your browsing activity. Click a slice o	ntified 155 topics trackers might think are relevant to you. f the chart to see more details.
	Pets & Animals
	79 Sites88 Trackers79 of the sites you visited were about Pets & Animals. These sites contained a total of 88 trackers.Pets & Animals is a popular interest.
Pets & Animals	Other people are often somewhat comfortable with having their interest in this topic being used to personalize their web experience. More about this interest >
	Filters Recency () ALL 24 HRS 7 DAYS Popularity () ALL LESS MORE

R

Tracking Transparency	Interests	Trackers	Sites	Activity	About	٥

Home > Interests > Pets & Animals

Your Pets & Animals profile

Pets & Animals is a popular interest.

Other people are often somewhat comfortable with having their interest in this topic being used to personalize their web experience.

Which sites were about Pets & Animals?

You visited **75 sites** that may have been about Pets & Animals. Therefore, trackers may have guessed this is relevant to you. *Click on a bar to learn more.*



Which trackers might think you are interested in *Pets & Animals*?

89 trackers may have guessed that you are interested in Pets & Animals. *Click on a bar to learn more.*



R

Tracking Transparency	Interests	Trackers	Sites	Activity	About 🌣



• Who is tracking you?

298 trackers have been present on the sites you visited since installing Tracking Transparency. Your most frequently encountered tracker is Google, which was present on 64.34% of the pages you visited.



Google

24989	1349	148
Pages	Sites	Interests

Google was present on **24989 pages** across **1349 sites** that you visited since installing Tracking Transparency. From those tracking encounters, they may have guessed that you are interested in **148 topics**.

More about this tracker >

Tracking Transparency	Interests	Trackers	Sites	Activity	About 🗶	*
		_	_		 G	

• Your Google profile

What does Google do?

"Google AdSense is a program run by Google that allows publishers in the Google Network of content sites to serve automatic text, image, video, or interactive media advertisements, that are targeted to site content and audience." – Source

Based on your browsing, what would *Google* think your interests are?

Using a machine to assign categories to pages you visit, Google may have guessed that you are interested in **150 topics**. *Click on a link in the wordcloud to learn more about each interest.*



On which sites did Google track you?

Google has tracked you on 1412 sites. Click on a bar to learn more.



Home > Trackers > Google

Evaluation approach

Iterative usability interviews



13 participants 30 minutes \$10 Amazon gift card



Field study

425 participants, 18+, located in the US, 95% HIT approval rating, use Firefox or Chrome regularly

pre-usage survey	use Tracking Transparency for one week	post-usage survey
[15 minutes, \$3]		[20 minutes, \$7]
 demographics estimates of tracking knowledge and attitudes about OBA 	collected telemetry data - # of distinct web pages - # of trackers - inferred topics	 qualitative reactions behavioral intentions estimates of tracking knowledge and attitudes about OBA

Study conditions

†	Static					
control		Iracking Iransparency	Sites	Activity	About	۵
	December	Home > Sites				
•	Browsing	Where w	ere	you tracked?		
1		Since installing this brow	ser exter	nsion, you have visited 38501 different pages on 1585 sites .		٦
comparison		Trackers see which sites fingerprinting. When a tra	you visit acker se	ted through a variety of tracking methods, including third-party cookies, tracking pixels, and browser es that you have visited multiple sites, they can use that activity to link together your interests.		
•		Most recently visited site	es			
		git.io				
A		github.blog				
fully		google.com				
featured		graducom				
•						

Study conditions



Study conditions

Results

Surprised by tracking prevalence

"I was surprised at how much I was tracked by Amazon and Google. I was also surprised to find that I was tracked on over 75% of the pages I visited." (P369)

"That live.com is the site with the most trackers. That is scary considering I only use live.com for email." (P140)

Improved knowledge of tracking

"Everything was pretty surprising and it feels like my privacy has been exposed. I never knew that companies tried to create ads supported for me based on my recent searches" (P175) "It shows my top interest is shopping, which i didn't figure that to be true, since i usually hate shopping. but it made me realize that i do a lot of shopping online now. that's new to me. It's also new that I have 75 potential interests." (P161)

Improved estimates of tracking 160 Actual 120 Standard 80 No Inferences 40 **Connection Graph** Browsing Tracker List \cap Static Pre-Usage Post-Usage

More likely to use blocking tools

Static Browsing Connection Graph Tracker List No Inferences Standard

More likely to use private browsing

Static Browsing Connection Graph Tracker List No Inferences Standard

About the same

https://git.io/trackingtransparency

Takeaways

built a platform to show users data about tracking in their own browsing

tool increased awareness of how inferences are made & ability to quantify tracking

longitudinal visualizations about tracking increased privacy intentions

What Twitter Knows

Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data

Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, Blase Ur

Targeted advertising: expectations

Miranda Wei <u>gender</u>: F <u>location</u>: Seattle, WA <u>interests</u>: cats, ramen, Battlestar Galactica

Basic demographics, interests, location

Reality: organic ketchup & more

Heinz Ketchup @HeinzKetchup_US

This is it, the last week of **#HeinzFantasySandwich**! This week, you're Team Grilled Cheese. You can only draft FOUR players. Who's on your team? **#FantasyFootball #FantasyDraft**

5:30 AM · Sep 26, 2018 · Twitter for Advertisers

targetingType: Tailored
audiences (lists)

targetingValue: NCS_PD_04358_Kraft_Organic and natural ketchup buyers_1_26362226

•••

...

targetingType: Keywords
targetingValue: #parenting

Targeting types

demographic:

characteristics about user and their device(s)

provided by user or inferred by Twitter

targeting type	Uses
language	350,121
platform	32,351
location	31,984
new device	236

Targeting types

demographic:

characteristics of user and their device(s)

 provided by user or inferred by Twitter

psychographic:

user lifestyles, behaviors or attitudes

 provided by user or inferred by Twitter

targeting type	USES
language	350,121
platform	32,351
location	31,984
new device	236

targeting type	Uses
behavior	35,008
interest	25,284

Targeting types

demographic:

characteristics of user and their device(s)

 provided by user or inferred by Twitter

psychographic:

user lifestyles, behaviors or attitudes

 provided by user or inferred by Twitter

advertiser:

user information collected offline ° provided by advertiser

targeting type	Uses
language	350,121
platform	32,351
location	31,984
new device	236

targeting type	USES
follower lookalikes	590,502
conversation	128,005
behavior	35,008
interest	25,284

targeting type	Uses
tailored (list)	113,952
mobile	21,631
tailored (web)	18,016

Potential policy violations?

prohibited to target by race, religion, sex life, health, politics, and financial status

yet...

- keyword: "unemployment," "Gay,"
 "#AfricanAmerican," "latinas"
- conversation: "Liberal Democrats (UK)"
- tailored list: "YYYY account status: balance due,"
 "Christian Audience to Exclude," "LGBT Suppression List"

Comparing ad explanations

Control: Because advertiser bought an ad
 Current: Facebook, Twitter style

Why am I seeing this ad?

You're seeing this ad because you're on a list **Postmates** wants to reach on Twitter. When the list was uploaded, Twitter did not learn any new identifying information about you.

Your Data

Postmates Learn more about Postmates

- March 17, 2019
 Postmates uploaded a hashed list. Twitter matched your information with information on that list.
- April 17, 2019
 You saw this ad from Postmates

There may be other reasons you're seeing this ad, including that **Postmates** wants to reach **people who are ages 18 and up, are located in Phoenix AZ, US, and are Female**. This information is based on your Twitter profile and where you've connected to the internet.

Comparing ad explanations

- Control: Because advertiser bought an ad
 Current: Facebook, Twitter style
- Speculative: Detailed text, detailed visual

Comparing ad explanations

- Control: Because advertiser bought an ad
- Current: Facebook, Twitter style
- Speculative: Detailed text, detailed visual
- Speculative: Creepy

creepy

Why am I seeing this ad?

You saw this ad on April 17, 2019 at 11:05 AM on the Twitter app from a(n) Android device, IP address ###.###. (Phoenix AZ, US).

You are seeing this ad because Postmates used your information, such as your email address or phone number, to find you on Twitter.

You are also seeing this ad because Postmates has made the following determinations about you:

- Your information on Twitter was matched with external lists called Suppression (Installs All Time) (email), Suppression (Installs All Time) (Device Id), and Email Suppression List (May 2018)
- You have a lot in common with people who follow @chrishemsworth, and @BarackObama
- You are interested in Health news and general info.
- You are participating in the conversation about Fitness on Twitter.
- · You are ages 18 and up, and are Female.
- You are located in or around Phoenix AZ, US.

Takeaways

• Data access enables transparency!

- \circ But is not really all that transparent
- Concerning targeting types are understudied
- How to enforce policy when skirting is trivial?
- Ad explanations should be more detailed
 - People liked "creepy"!

During the Twitter study ...

Data downloads have lots of useful info
But they are also undocumented, hard to interpret
Confusion between rights of access, portability

How should we reimagine data downloads for people?

How do users react to the format and content of data downloads?

What information do they find important? What practical uses do they imagine?

How should data downloads be redesigned to support transparency and other goals?

Method: Focus groups and co-design

Focus group details

• Two sessions for each of six platforms:

• Amazon (orders), Facebook, Google (search), Spotify, Uber, YouTube

• 3-5 participants per session

- Only one student; only one IT expert per session
- Remotely via Hangouts
- Extensive privacy protections for participants

Results

Surprised by the level of detail

I knew Google is recording everything. It's just that seeing this in front of me and all the data that has been collected over all the years, it's like a rude realization. (GA4)

It was like reading a book about myself but not written by myself. (FB1)

Uses and misuses

Track my own privacy exposure
Especially in case of data breach
Misuse by law enforcement?

Inferences and synthesis

What's interesting to me is how my online behavior is affecting how this company and all the affiliates see me. And in what category, say, they put me or don't put me. (FA2)

- Desire to synthesize different kinds of data to find patterns
- e.g., Examine spending data or site usage over time

Disorganized, hard to parse

 Hard to find specific info across multiple files
 Difficulty with JSON files

Most of the interesting data is stored in these files, that as a non-specialist, I can't read. ... We're effectively illiterate when it comes to reading this additional data that they've been collecting. (YA5)

Interactivity desired

• Sort, filter, prioritize

- High-level overview with click/hover to zoom in
- Ability to verify, question, delete inband

Organize content

Allow exploration, filtering, interactivity

Recommendations: Design

For companies and for third-party organizations

Support aggregation and inferencing (simple scripting?)

Direct manipulation for participation, erasure

Implications for policy

Differentiate	Mandate	Clarify
access from portability	comprehensibility (e.g., via README)	required contents (especially inferences)

Overall takeaways

- Transparency and personalization help users better understand how their data is collected and used
- Concrete examples lead to meaningful preferences
- Much more work needed to make transparency tools truly usable and useful