# Lecture 5: Tracking and Inference Algorithms

CMSC 25910 Spring 2024 The University of Chicago







### **Online Tracking**

#### Advertisers want to show you advertisements targeted to your interests and demographics Ads Preferences

	Accounting & Finance Jobs	1	Action & Adventure Films
P	Action & Platform Games	P	Adventure Games
۰	Android OS	-	Autos & Vehicles
	Banking	-	Bars, Clubs & Nightlife
۲	Beaches & Islands	8	Blues
-	Bollywood & South Asian Film		Books & Literature
•	Business & Productivity Software		Business News

How your ads are personalized

Ads are based on personal info you've added to your Google Account, data from advertisers that partner with

Google, and Google's estimation of your interests. Choose any factor to learn more or update your

preferences. Learn more

Ads on Search and Gmail

Opt out

#### Ads on the web

#### Make the ads you see on the web more interesting

- Ads on the web

Many websites, such as news sites and blogs, partner with us to show ads to their visitors. To see ads that are more related to you and your interests, edit the categories below, which are based on sites you have recently visited. Learn More

Your interests are associated with an advertising cookie that's stored in your browser. If you don't want us to store your interests, you can opt out below. Your ads preferences only apply in this browser on this computer. They are reset if you delete your browser's cookies.

Watch a video: Ads Preferences on GDN explained

#### Your categories

Below you can review the interests and inferred demographics that Google has associated with your cookie. You can remove or edit these at any time.

Arts & Entertainment

Computers & Electronics

Computers & Electronics - Consumer Electronics - Gadgets & Portable Electronics - PDAs & Handhelds

Internet & Telecom

Internet & Telecom - Mobile & Wireless - Mobile Phones - Smart Phones

Law & Government

Science

#### Your demographics

We infer your age and gender based on the websites you've visited. You can remove or edit these at any time.

Age: 35-44

Gender: Male

Google

#### **Data-Driven Inferences**





#### You might like dogs!

### **Online Tracking**

- First party = the site you are visiting (address is in the URL bar)
- Third party = other sites (i.e., origins) contacted as a result of your visit to the first party
- First-party tracking (on search engines, shopping sites)
- Third-party tracking (ads on lots of sites)

#### Mechanics of First-Party Online Tracking

• Use cookies, JavaScript, URL parameters to track



#### Mechanics of First-Party Online Tracking

Sponsored 🕜	Desert Living Rock Cactus Cacti 10 Ct ★★☆☆☆ ~ 6	Waimanalo Papaya! Tropical Fruit Tree Seeds Plant ★★★☆☆ ~ 8	Seed Glow
	\$899	\$ <b>10</b> <sup>68</sup>	\$799
	\$5.48 shipping Only 4 left in stock - order soon.	FREE Shipping	✓prin FREE by An Only 7
https://www.amazon.com/Plant-Seeds-Succulent-Ga	arden-Bonsai/dp/B08RRQGR6F/ref=sr_1 9?dchild=1&keyw	ords=rare+plants&gid=1621969916&sr=8-9	_

#### Mechanics of Third-Party Online Tracking



### Details of HTTP Request

Request Headers (735 B) Raw	>
GET / HTTP/2	1
Host: www.uchicago.edu	
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefo	1
<pre>Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8</pre>	
Accept-Language: en-US, en;q=0.5	
Accept-Encoding: gzip, deflate, br	
DNT: 1	
Connection: keep-alive	
Cookie: uchicago-prod_last_visit=1306604446; uchicago-prod_last_activity=1621964446;	
Upgrade-Insecure-Requests: 1	
If-Modified-Since: Tue, 25 May 2021 17:40:36 GMT	
TE: Trailers	
c	> 1
	P

### Details of HTTP Request

Request Headers (735 B) Raw
GET / HTTP/2
Host: www.uchicago.edu
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefo
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US, en;q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Connection: keep-alive
Cookie: uchicago-prod_last_visit=1306604446; uchicago-prod_last_activity=1621964446;
Upgrade-Insecure-Requests: 1
If-Modified-Since: Tue, 25 May 2021 17:40:36 GMT
TE: Trailers
< >

#### **Details of HTTP Cookies**



Image taken from https://networkencyclopedia.com/http-cookie/

#### Details of HTTP Request

Request Headers (735 B) Raw	
<pre>GET / HTTP/2 Host: www.uchicago.edu User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefo Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8 Accept-Language: en-US,en;q=0.5 Accept-Encoding: gzip, deflate, br DNT: 1</pre>	
Connection: keep-alive	
Cookie: uchicago-prod_last_visit=1306604446; uchicago-prod_last_activity=1621964446; Upgrade-Insecure-Requests: 1	
If-Modified-Since: Tue, 25 May 2021 17:40:36 GMT TE: Trailers	
<	>

### Details of HTTP Response

 Response Headers (1.078 KB) Raw HTTP/2 200 OK date: Tue, 25 May 2021 18:00:35 GMT content-type: text/html; charset=UTF-8 server: Apache x-frame-options: SAMEORIGIN expires: Mon, 26 Jul 1997 05:00:00 GMT pragma: no-cache vary: Accept-Encoding set-cookie: uchicago-prod last visit=1306605629; expires=Wed, 25-May-2022 18:00:29 G set-cookie: uchicago-prod\_last\_activity=1621965629; expires=Wed, 25-May-2022 18:00:2 set-cookie: uchicago-prod tracker=%7B%220%22%3A%22index%22%2C%22token%22%3A%2226944a set-cookie: uchicago-prod csrf token=deleted; expires=Thu, 01-Jan-1970 00:00:01 GMT; set-cookie: uchicago-prod csrf token=86d47d8690aa7646e1628dd095cd5b464db16bd3; expir last-modified: Tue, 25 May 2021 18:00:29 GMT content-encoding: gzip x-varnish: 10696657 9201444 age: 5 via: 1.1 varnish (Varnish/5.2) accept-ranges: bytes X-Firefox-Spdy: h2

### Details of HTTP Response

<ul> <li>Response Headers (1.078 KB)</li> </ul>	Raw 💽
HTTP/2 200 OK	
date: Tue, 25 May 2021 18:00:35 GMT	
content-type: text/html; charset=UTF-8	
server: Apache	
x-frame-options: SAMEORIGIN	
expires: Mon, 26 Jul 1997 05:00:00 GMT	
pragma: no-cache	
vary: Accept-Encoding	
<pre>set-cookie: uchicago-prod_last_visit=1306605629; expires=Wed, 25-May-2022</pre>	18:00:29 G
<pre>set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-20</pre>	022 18:00:2
<pre>set-cookie: uchicago-prod_tracker=%7B%220%22%3A%22index%22%2C%22token%22%3</pre>	3A%2226944a
<pre>set-cookie: uchicago-prod_csrf_token=deleted; expires=Thu, 01-Jan-1970 00</pre>	
<pre>set-cookie: uchicago-prod_csrf_token=86d47d8690aa7646e1628dd095cd5b464db10</pre>	6bd3; expir
last-modified: Tue, 25 May 2021 18:00:29 GMT	
content-encoding: gzip	
x-varnish: 10696657 9201444	
age: 5	
via: 1.1 varnish (Varnish/5.2)	
accept-ranges: bytes	
X-Firefox-Spdy: h2	

#### uchicago.edu $\rightarrow$ youtube.com

	Insp ل	pector Ŋ Cons	ole D Debugger <b>{}</b> Style Editor	Performan	ce 🛈	Memory 1	Vetv	ork 🗄 Storage 🛉 Accessibility 🎬 Application 🛛 🌒 13 🗖 🗊 🚥
Ô	🗑 Filter	URLs					All	HTML CSS JS XHR Fonts Images Media WS Other Disable Cache No Throttling +
Status	Met	Domain	File	Initiator	Туре	Transferred	Size	Headers Cookies Request Response Timings Security
200	GET	www.uchica	/	document	html	11.41 KB	39	
204	POST	www.youtub	atr?ns=yt⪙=embedded&cpn=ho5PKBh-	base.js:1023 (	html	604 B	0 B	Status 200 OK ?
200	GET	A www.youtub	P-xlixF7B2U?autohide=1&fs=1&autoplay=	subdocument	html	21.81 KB	51	Version HTTP/3
200	GET	Cdn.hypemar	uchicagowww?width=1169&paginate=tru	a5b5e5.js:3 (s	html	128.06 KB	12	Transferred 21.81 KB (51.50 KB size)
200	GET	Cdn.hypemar	popUpModalEndpoint	a5b5e5.js:3 (s	html	10.99 KB	10	Referrer Policy strict-origin-when-cross-origin
								Response Headers (642 B)     Raw      Rev
					date: Tue, 25 May 2021 18:00:36 GMT strict-transport-security: max-age=31536000			
								<pre>permissions-policy: ch-ua-full-version=*, ch-ua-platform=*, ch-ua-platform-version=*, ch-u content-encoding: br server: ESF x-xss-protection: 0 alt-svc: h3-29=":443"; ma=2592000,h3-T051=":443"; ma=2592000,h3-Q050=":443"; ma=2592000,h3</pre>
								content-encoding: br server: ESF x-xss-protection: 0
								<pre>content-encoding: br server: ESF x-xss-protection: 0 alt-svc: h3-29=":443"; ma=2592000,h3-T051=":443"; ma=2592000,h3-Q050=":443"; ma=2592000,h3 &lt;</pre>

	🗘 Insp	ector D Conso	ole D Debugger <b>{}</b> Style Editor	Performan	ice 🕄	🕽 Memory 1	Net
Û	🗑 Filter	URLs				11 9 0	Al
Status	Met	Domain	File	Initiator	Туре	Transferred	Size
200	GET	A www.uchica	/	document	html	11.41 KB	39
204	POST	Www.youtub	atr?ns=yt⪙=embedded&cpn=ho5PKBh-	base.js:1023 (	html	604 B	0 B
200	GET	A www.youtub	P-xlixF7B2U?autohide=1&fs=1&autoplay=	subdocument	html	21.81 KB	51
200	GET	Cdn.hypemar	uchicagowww?width=1169&paginate=tru	a5b5e5.js:3 (s	html	128.06 KB	12
200	GET	Con.hypemar	popUpModalEndpoint	a5b5e5.js:3 (s	html	10.99 KB	10

```
    Request Headers (621 B)

                                                                                     Raw
GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&e
Host: www.youtube.com
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox/88.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US, en; q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Alt-Used: www.youtube.com
Connection: keep-alive
Referer: https://www.uchicago.edu/
Cookie: VISITOR INFO1 LIVE=dACtKPaJViQ; PREF=tz=America.Chicago&f4=4000000; YSC=p2jSvxCMeI
Upgrade-Insecure-Requests: 1
TE: Trailers
```

Request Headers (621 B)	Raw 🔵
GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showi Host: www.youtube.com	nfo= <mark>0&amp;hd=1</mark> &e
<pre>User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q Accept-Language: en-US,en;q=0.5</pre>	
Accept-Encoding: gzip, deflate, br DNT: 1	
Alt-Used: www.youtube.com Connection: keep-alive Referer: https://www.uchicago.edu/	
Cookie: VISITOR_INFO1_LIVE=dACtKPaJViQ; PREF=tz=America.Chicago&f4=4000000; YSU Upgrade-Insecure-Requests: 1 TE: Trailers	C=p2jSvxCMeI

Request Headers (621 B)	Raw 🔵
GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&	showinfo=0&hd=1&e
Host: www.youtube.com	
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/2010	Charles A.
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,	*/*;q=0.8
Accept-Language: en-US,en;q=0.5	
Accept-Encoding: gzip, deflate, br	
DNT: 1	
Alt-Used: www.youtube.com	
Connection: keep-alive	
Referer: https://www.uchicago.edu/	the state of the second state of the second
Cookie: VISITOR_INFO1_LIVE=dACtKPaJViQ; PREF=tz=America.Chicago&f4=400000	0; YSC=p2jSvxCMeI
Upgrade-Insecure-Requests: 1	
TE: Trailers	

### Putting It Together

- (Unless browser is blocking it) third party gets its cookies
- (Unless browser is blocking it) third party sees "referer" [sic]
- First party can choose to send info to third party via URL parameters (not a violation of Same Origin Policy!)
- Third party sees this information for many first parties

## Mechanics of Cookie Syncing

- JavaScript / images from advertising networks loaded as part of your page
  - In iframes
  - Or sometimes not
  - Why does this matter?
- Let's discuss: what can an ad network learn, and how?

### Mechanics of Cookie Syncing



Figure 1: Example of advertiser.com and tracker.com synchronizing their cookieIDs. Interestingly, and without having any code in website3, advertiser.com learns that: (i) cookieIDs userABC==user123 and (ii) userABC has just visited the given website. Finally, both domains can conduct server-to-server user data merges.

From Papadopoulos et al. "Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask," in *Proc. WWW*, 2019.

### **Browser Fingerprinting**

- Use features of the browser that are relatively unique to your machine
  - Fonts
  - GPU model anti-aliasing (Canvas fingerprinting)
  - User-agent string
  - (Often not) IP address (Why not?)

### **Browser Fingerprinting**

- Use combination of device features as an identifier
- <u>https://coveryourtracks.eff.org/</u>



# Alternatives to Cookies for Tracking / Profiling

- Federated Learning of Cohorts
- Clusters users based on their browsing activity and assigns a cohort ID
  - Uses SimHash for clustering
  - Clusters intended to contain 1,000s of users
- Criticisms include fingerprintability, ability to tie cohort to PII, and collapse of different browsing contexts
- (Abandoned in early 2022)

 $\mathbf{C}$ 

#### ○ A https://www.privacyaffairs.com/google-floc/

#### Selecting Interest-based Ads Using FLoC

- Browsers use a FLoC service to get the mathematical model, consisting of many calculated "cohorts." In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.
- 2. Using that FLoC Model algorithm, your browser calculates your cohort.
- 3. Let's say you visited the site of an advertiser abc.com that sells kitchen appliances. Then that site requests the cohort ID from your browser.
- 4. If you visited additional pages of the advertiser, like searching kitchen utensils, it would record those interests.
- 5. Advertisers record these cohort activities periodically and share that information with the ad tech company that helps to deliver advertisements.
- 6. In the same manner, let's say you visited a publisher site that sells ad space; it will also request your cohort ID.
- 7. Then the publisher site requests advertisements relevant to that cohort from the ad tech company.
- 8. The ad tech company combines the data received from the advertiser company about the cohort's interests and data from the publishing company.
- 9. Next, the ad tech company chooses suitable ads according to the interests of the cohort.
- 10. The publisher site then displays the selected advertisement relevant to the interests of the cohort.



#### Google's FLoC Is a Terrible Idea

BY BENNETT CYPHERS | MARCH 3, 2021



About

Update, April 9, 2021 : We've launched <u>Am I FLoCed</u>, a new site that will tell you whether your Chrome browser has been turned into a guinea pig for Federated Learning of Cohorts or FLoC, Google's latest targeted advertising experiment.

The third-party cookie is dying, and Google is trying to create its replacement.

No one should mourn the death of the cookie as we know it. For more than two decades, the third-party cookie has been the lynchpin in a shadowy, seedy, multi-billion dollar advertising-surveillance industry on the Web; phasing out

#### Join Our Newsletter! Email updates on news, actions, events in your area, and more. Email Address

Issues Our Work Take Action Tools Donate Q

Postal Code (optional) Anti-spam question: Enter the three-letter abbreviation for Electronic Frontier Foundation:

Image taken from https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea

### Google's FLoC (Brave's Perspective)

BROWSER CREATORS ADVERTISERS COMPANY BLOG BRAVE SEARCH DOWNLOAD Why Brave | Apr 12, 2021 | Community, Security & Privacy

> <u>Peter Snyder</u>, Senior Privacy Researcher at Brave <u>Brendan Eich</u>, CEO and co-founder of Brave

#### A Step in the Wrong Direction

FLoC is a recent Google proposal that would have your browser share your browsing behavior and interests by default with every site and advertiser with which you interact. Brave opposes FLoC, along with any other feature designed to share information about you and your interests without your fully informed consent. To protect Brave users, **Brave has removed FLoC in the Nightly version of both Brave for desktop and Android. The privacy-affecting aspects of FLoC have never been enabled in Brave releases; the additional implementation details of FLoC will be removed from all Brave releases with this week's stable release. Brave is also disabling FLoC on our websites, to protect Chrome users learning about Brave.** 

Image taken from https://brave.com/why-brave-disables-floc/

FLoC service	Browser	Advertiser a site that pays to advertise its products shoestore.example	Publisher a site that gets paid to display ads dailynews.example	Adtech a company that provides services to deliver ads adnetwork.example	Publisher a site that gets paid to display ads dailynews.example
00%01 #1277 #1077 #1878 #1378 #1378 #1378 #1378 #1378 #1378 #1378 #1378		Shoe Store			
Create FLoC model	Calculate browser cohort	Observe cohort activity	3 Observe visitor cohorts	Select ads relevant to cohort	Display relevant ads
The FLoC service creates a mathematical model with thousands of "cohorts". Each cohort corresponds to thousands of web browsers with similar recent browsing histories.	A user's browser, on their device, uses the FLoC model to calculate which cohort corresponds most closely to its own browsing history.		A publisher observes the cohorts of visitors to its site and shares this with its adtech platforms.* handle observing the cohort he advertiser or publisher.	An adtech platform selects ads appropriate for the user's cohort.	The publisher displays ads that are relevant to the user based on their browser's cohort.

Image taken from https://arstechnica.com/gadgets/2021/04/everybody-hates-floc-googles-tracking-plan-for-chrome-ads/

#### Selecting Interest-based Ads Using FLoC

- Browsers use a FLoC service to get the mathematical model, consisting of many calculated "cohorts." In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.
- 2. Using that FLoC Model algorithm, your browser calculates your cohort.
- 3. Let's say you visited the site of an advertiser abc.com that sells kitchen appliances. Then that site requests the cohort ID from your browser.
- 4. If you visited additional pages of the advertiser, like searching kitchen utensils, it would record those interests.
- 5. Advertisers record these cohort activities periodically and share that information with the ad tech company that helps to deliver advertisements.
- 6. In the same manner, let's say you visited a publisher site that sells ad space; it will also request your cohort ID.
- 7. Then the publisher site requests advertisements relevant to that cohort from the ad tech company.
- 8. The ad tech company combines the data received from the advertiser company about the cohort's interests and data from the publishing company.
- 9. Next, the ad tech company chooses suitable ads according to the interests of the cohort.
- 10. The publisher site then displays the selected advertisement relevant to the interests of the cohort.

#### Image taken from https://www.privacyaffairs.com/google-floc/

### Google's Topics API



#### Google's Topics API

 $\mathbf{C}$ 

🔿 🔒 https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking

VERGE 😏 TWITTER

TER **f** FACEBOOK

Your browser will store these topics for three weeks before deleting them. Google says that these categories "are selected entirely on your device" and don't involve "any external servers, including Google servers." When you visit a website, Topics will show the site and its advertising partners just three of your interests, consisting of "one topic from each of the past three weeks."

As noted on the <u>Topics API GitHub</u> page, there are currently about 350 available topics in its advertising taxonomy (although Google plans on adding anywhere from "a few hundred" to "a few thousand" eventually). Google says Topics won't include any "sensitive categories" like race or gender. And if you're using Chrome, the company is building tools to let you view and delete topics, as well as turn off the feature.

# Studying User Reactions to (Future) Hyper-Targeting

#### Study: Banner Ads



#### Taylor, treat Ryan to a date night this week in Memphis.

We know you LOVE Thai restaurants. Use SUPEReats.co to reserve a table at one of the 7 near you for a deal!

www. 🔗 🗡

www.SUPEREATS.co

#### Personalized



See https://www.blaseur.com/papers/robotext-full.pdf
## Study: Robotext Ads



### Generic

### Personalized





# **Anti-Tracking Tools**

## **Firefox Browser Default Settings**



## Firefox Browser Strict Settings



## Do not track

- W3C standard
- User checks a box
- Browser sends a "do not track" header to website
- Website stops "tracking"
- W3C working group tried to define what "stopping tracking" meant

000			Privacy
General Tabs Co	ntent Applicati	ons Privacy	Security Syr
Tracking			_
Tell web site	es I do not war	nt to be trac	ked
History			
Firefox will:	Remember his	tory	:
Web sites you	member your bro visit. to <u>clear your rec</u>		_
Location Bar			
When using the	e location bar,	suggest: 🗍	History and I

## Are Tools to Stop Tracking Effective?

- Browser privacy settings
  - Cookie blocking
  - P3P
  - Tracking Protection Lists
  - Do Not Track
- Browser add-ons
- Opt-out cookies
- Digital Advertising Alliance (DAA) AdChoices icon and associated opt-out pages



# **Transparency Tools**

## **Current Transparency Tools**































Ads are based on personal info you've added to your Google Account, data from advertisers that partner with Google, and Google's estimation of your interests. Choose any factor to learn more or update your preferences. Learn more

18-24 years old	Female	
Cooking & Recipes	Shopping	

## Actual Targeting is Far More Specific

Hobbies & Leisure>Special Occasions>Holidays & Seasonal Events>Halloween & October 31st Hobbies & Leisure>Special Occasions>Holidays & Seasonal Events>Islamic Holidays Hobbies & Leisure>Special Occasions>Holidays & Seasonal Events>Jewish Holidays Hobbies & Leisure>Special Occasions>Holidays & Seasonal Events>New Year Hobbies & Leisure>Special Occasions>Holidays & Seasonal Events>Thanksgiving Hobbies & Leisure>Special Occasions>Holidays & Seasonal Events>Valentine's Day Hobbies & Leisure>Special Occasions>Weddings Hobbies & Leisure>Water Activities Hobbies & Leisure>Water Activities>Boating Hobbies & Leisure>Water Activities>Diving & Underwater Activities Hobbies & Leisure>Water Activities>Surf & Swim Home & Garden Home & Garden>Bed & Bath Home & Garden>Bed & Bath>Bathroom Home & Garden>Bed & Bath>Bedroom Home & Garden>Bed & Bath>Bedroom>Bedding & Bed Linens Home & Garden>Bed & Bath>Bedroom>Beds & Headboards Home & Garden>Bed & Bath>Bedroom>Mattresses Home & Garden>Domestic Services Home & Garden>Domestic Services>Cleaning Supplies & Services Home & Garden>Gardening & Landscaping Home & Garden>HVAC & Climate Control Home & Garden>Home Appliances Home & Garden>Home Appliances>Water Filters & Purifiers Home & Garden>Home Furnishings Home & Garden>Home Furnishings>Clocks Home & Garden>Home Furnishings>Lamps & Lighting Home & Garden>Home Furnishings>Rugs & Carpets Home & Garden>Home Furnishings>Sofas & Chairs Home & Garden>Home Improvement Home & Garden>Home Improvement>Construction & Power Tools Home & Garden>Home Improvement>Doors & Windows Home & Garden>Home Improvement>Flooring Home & Garden>Home Improvement>House Painting & Finishing Home & Garden>Home Improvement>Plumbing Home & Garden>Home Improvement>Roofing Home & Garden>Home Storage & Shelving





## Design of Tracking Transparency (Research)



- Web browser extension
- Locally records tracking
  - Page metadata, page contents, trackers

D.

Profiles potential interests from that page

https://git.io/trackingtransparency



### **Potential Interests**



### Trackers

### 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.



> See all 1412 sites

### Tracking Transparency v2





# 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









Southampton

## Possible Expectation of Targeted Advertising

... using information about demographics, preferences, and activity to target ads

### **Expectation:**





## Reality: "Organic Ketchup Buyer" + More

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** 

HEIN2 Heinz Ketchup

@HeinzKetchup US



targetingType: Tailored audiences
(lists)

targetingValue: NCS\_PD\_04358\_Kraft\_Organic and natural ketchup buyers\_1\_26362226

•••

...

targetingType: Keywords

targetingValue: #parenting

## Methods Enabled by Subject Data Access Rights

Settings		← Your Twitter data	
@_weimf		Account >	
Account	>	Account history >	
Privacy and safety	>	Apps, devices & information >	
Notifications	>	Account activity >	
Content preferences	>	Interests and ads data >	
General		Download an archive of your data	
Display	>	You can request a ZIP file with the information that we believe is most relevant and useful to you. You'll get an email sent to mirandaaa.wei@gmail.com when it's ready to be downloaded.	
Data usage	>	desktop web browser. Learn more	
Accessibility	>	Twitter Request archive	
About Twitter	>	Periscope Request archive	

### Contents of ad-impressions.js

```
"ad" : {
 "adsUserData" : {
   "adImpressions" : {
     "impressions" : [ {
       "deviceInfo" : {
         "osType" : "Ios",
         "deviceId" : "#############",
         "deviceType" : "iPhone X"
       },
       "displayLocation" : "SearchTweets",
       "promotedTweetInfo" : {
         "tweetText" : "RT @SpotifyBrands: Young people's digital lives are subtly shifting
                       culture. Discover more with our global trends report.",
         "urls" : [],
         "mediaUrls" : []
       },
       "advertiserInfo" : {
         "advertiserName" : "Spotify",
         "screenName" : "@Spotify"
       },
       "matchedTargetingCriteria" : [ {
         "targetingType" : "Events",
         "targetingValue" : "Back to School 2019"
       }, {
         "targetingType" : "Age",
         "targetingValue" : "18 to 49"
       }. {
         "targetingType" : "Locations",
         "targetingValue" : "United States"
       }, {
         "targetingType" : "Platforms",
         "targetingValue" : "iOS"
```

### **Study Protocol**



### 231 Participants and Their Ads



Per participant (means)

- 6.6 years on Twitter
- 1,046.6 ads in last 90 days

### Across all participants

- 240,651 ads with at least one targeting type
  - e.g., location targeting
- 45,209 unique instances
  - e.g., Chicago

## Less-discussed Targeting Types

 Follower lookalikes: people who don't necessarily follow an account, but have similar interests/demographics to that account's followers

242,709 times, e.g., @uchicagocs

- Behavior targeting: specific inferences or actions on Twitter 35,088 times, e.g., US - Household income: \$30,000-\$39,000
- Tailored audience (list): upload lists containing hashed PII 113,952 times, e.g., Lifetime Suppression [Installs] (Device ID)

## Potential Violations of Twitter's Policies

TOS prohibits targeting on 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"

### **Current Ad Explanations**



G

#### Why am I seeing this ad?

One reason you may be seeing this ad is that **Postmates** wants to reach people interested in **Health news and general info**. There may be other reasons you're seeing this ad, including that **Postmates** wants to reach **people above the age of 18 and located here: Phoenix AZ, US**.

You can view and manage information connected to your account that Twitter may use for ads purposes. See your Twitter data.

Twitter also personalizes ads using information received from partners as well as app and website visits. You can control these interest-based ads using the "Personalize ads" setting.

#### 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.

## **Speculative Ad Explanations**

#### Why am I seeing this ad?

You may be seeing this ad because Postmates wants to reach people similar to people who follow @chrishemsworth; and @BarackObama.

You may also be seeing this ad because Postmates has added your Twitter username or email to a list of people who they want to reach. You may have been added if you visited their webpage, used their mobile app, or signed up for their mailing list.

You may also be seeing this ad because Postmates wants to reach people on the following audience lists: Suppression (Installs All Time) (email); Suppression (Installs All Time) (Device Id); and Email Suppression List (May 2018)

You may also be seeing this ad because Postmates wants to reach people interested in Health news and general info. Your interest profile is based on your tweets and retweets, pages and people you follow, websites you visit, and more.

You may also be seeing this ad because Postmates wants to reach people in the following demographics: ages 18 and up; Phoenix AZ, US; and Female.

You can view or manage account information used for ad purposes. Go to the "Personalize ads" setting to control internet-based ads.

#### Why am I seeing this ad?

Some of the targeting types used to target this ad to you were:



Tailored audiences: Postmates can add your name, your Twitter username, or your email to a list of people they want to reach.



Suppression (Installs All Time) (email) Suppression (Installs All Time) (Device Id) Email Suppression List (May 2018)



Follower Look-alikes: Postmates can target people who are similar to people who follow a person or page on Twitter.

### detailed visual

detailed text



### **Speculative Ad Explanations**

### 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.

These inferences are based on your Twitter profile and online activities, such as your tweets and retweets, people and pages you follow, and websites you visit, as well as data that third parties have provided about you.

You can view or manage account information used for ad purposes. Go to the "Personalize ads" setting to control internet-based ads.

### Why am I seeing this ad?

### control

One reason you may be seeing this ad is that **Postmates** paid for an ad on this site.

You can view or manage account information used for ad purposes. Go to the "Personalize ads" setting to control internet-based ads.

### creepy

### Speculative Ad Explanations: Results



Figure 5: Participants' level of agreement to questions about ad explanations.

## **Clarifications From Twitter**



# The Politics of Transparency

## The Politics of Transparency



Who is paying for online advertising to target you? Who's showing you ads because you're married, or you're a liberal or work in the oil industry or some other category? Are they saying different things to different people? Help hold them accountable by installing our browser plugin, which sends us the ads you see on Facebook and YouTube, without compromising your privacy.

#### **o** Install for Chrome

### Install for Firefox

## The Politics of Transparency

### How it works

Ad Observer is a tool you add to your Web browser. It copies the ads you see on Facebook and YouTube, so anyone can see them in <u>our public database</u>. (See documentation <u>here</u>.) If you want, you can enter basic demographic information about yourself in the tool to help improve our understanding of why advertisers targeted you. However, we'll never ask for information that could identify you.

It doesn't collect your personal information. We take your privacy very seriously. <u>You can view</u> the extension code on Github.

#### See Mozilla privacy review.

### What we collect

- The advertiser's name and disclosure string.
- The ad's text, image, and link.
- The information Facebook provides about how the ad was targeted.
- When the ad was shown to you.
- Your browser language.

### What we don't collect

Anything personally identifying, including:

- Your Facebook ID number.
- Your name, birthday, friend list, etc.
- How you interacted with ads or anything else.

## The Politics of Transparency



https://www.wsj.com/articles/facebook-cuts-off-access-for-nyu-research-into-political-ad-targeting-11628052204