Lecture 5: Tracking and Inference Algorithms

CMSC 25910

Spring 2024

The University of Chicago

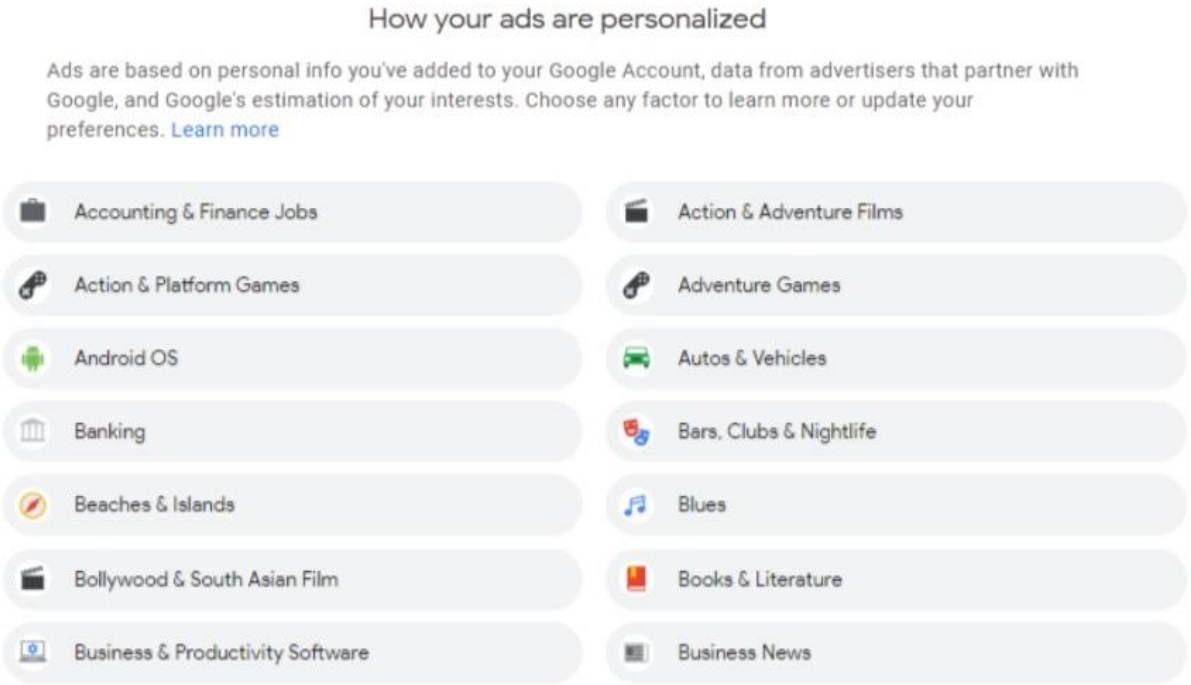
THE UNIVERSITY OF CHICAGO

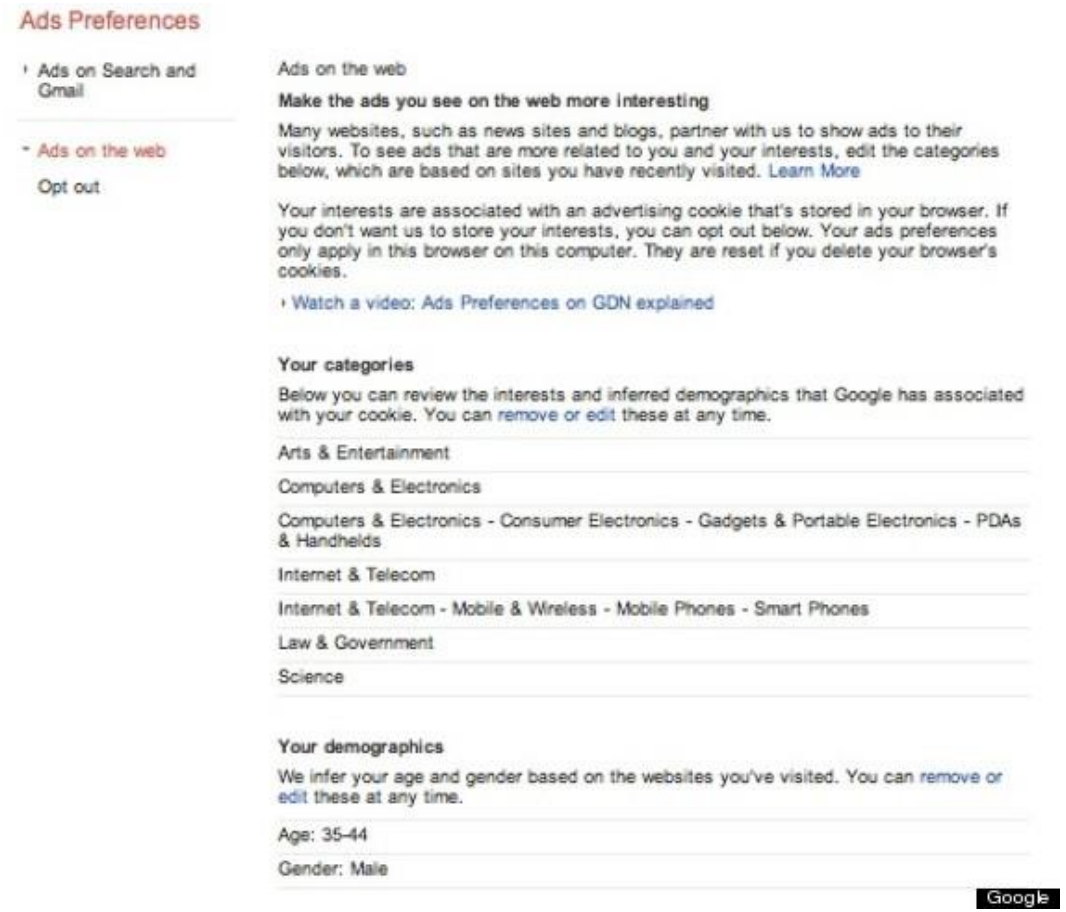
Chartbeat
Google
amazon
facebook
Quantcast
criteo
comscore
2
New Relic.
bluekai

Chartbeat
Google
facebook
criteo
amazon
Quantcast
comscore
New Relic
bluekai

Online Tracking

- Advertisers want to show you advertisements targeted to your interests and demographics





Data-Driven Inferences

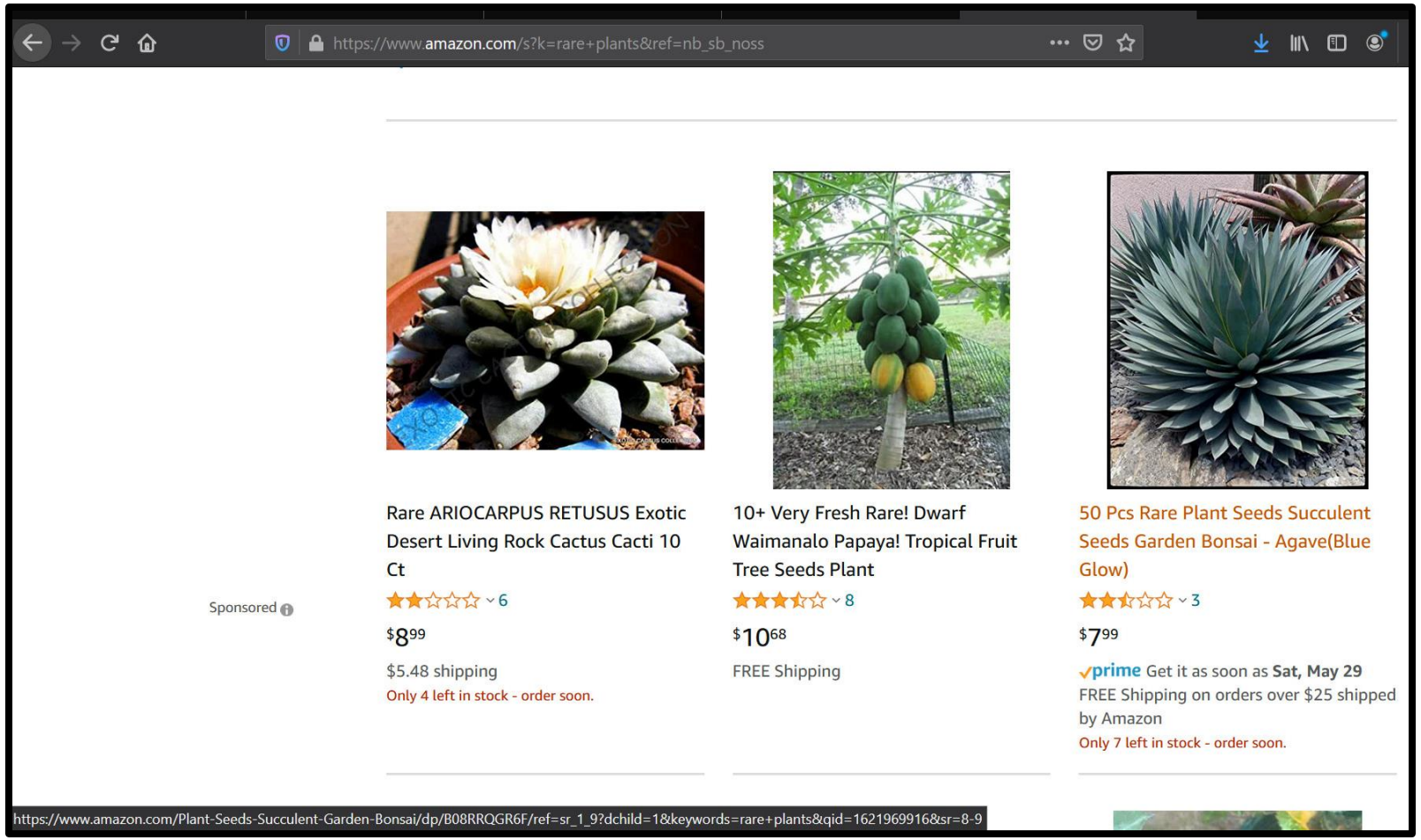
You might like dogs!

Online Tracking

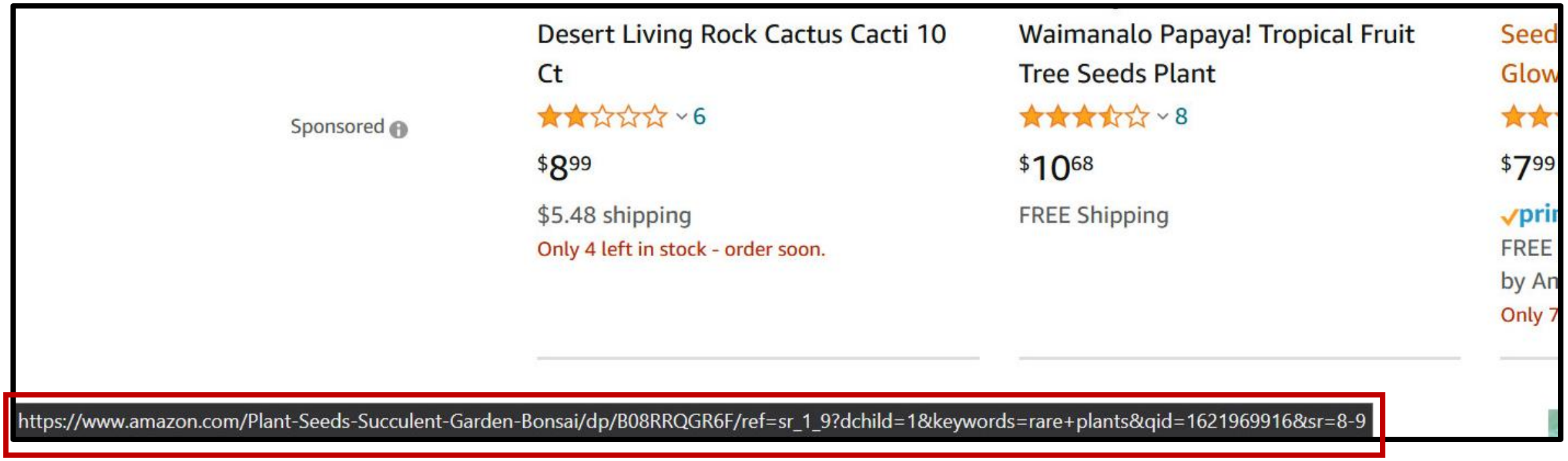
- First party = the site you are visiting (address is in the URL bar)
- Third party = other sites (i.e., origins) contacted as a result of your visit to the first party
- First-party tracking (on search engines, shopping sites)
- Third-party tracking (ads on lots of sites)

Mechanics of First-Party Online Tracking

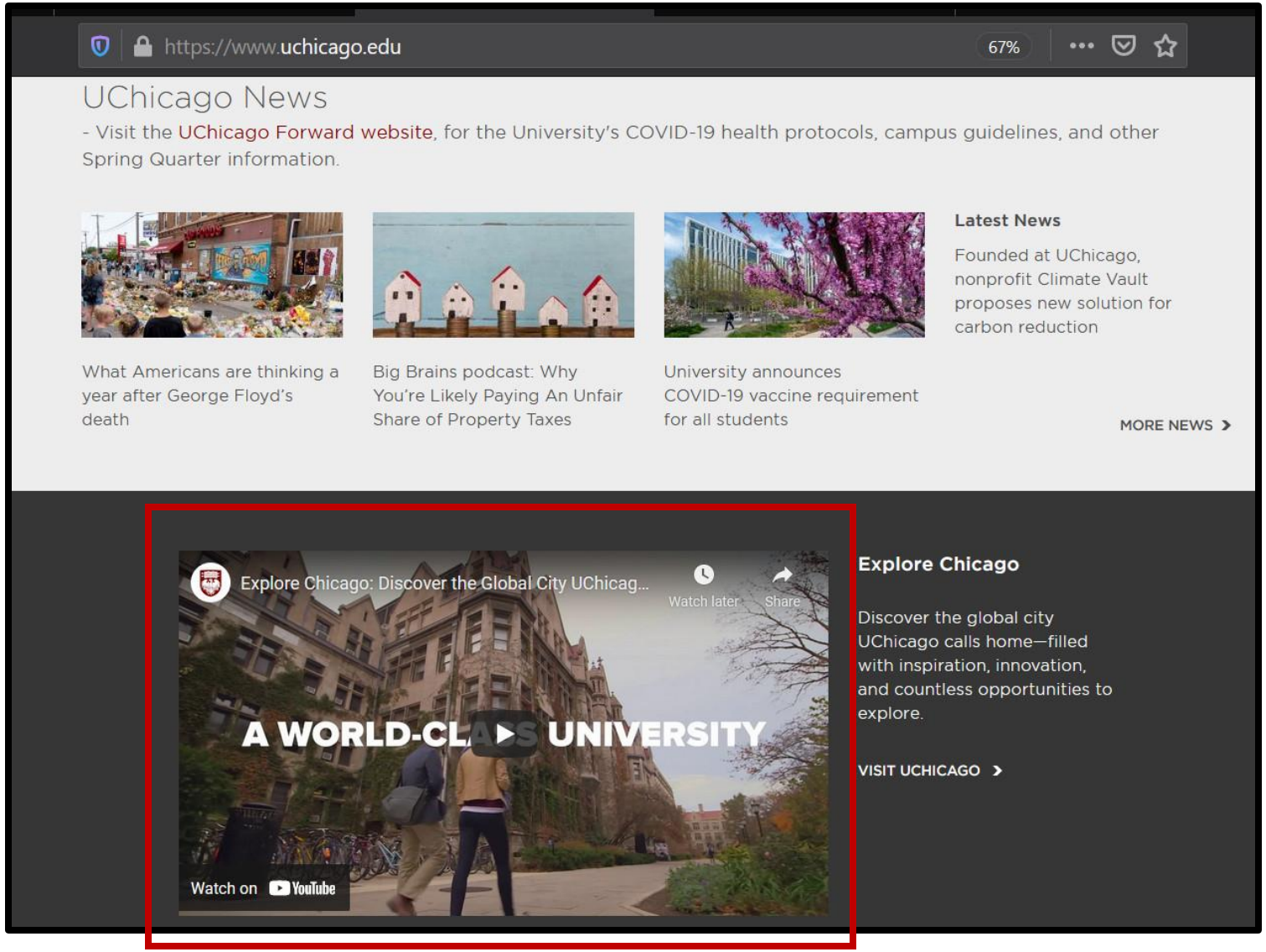
- Use cookies, JavaScript, URL parameters to track



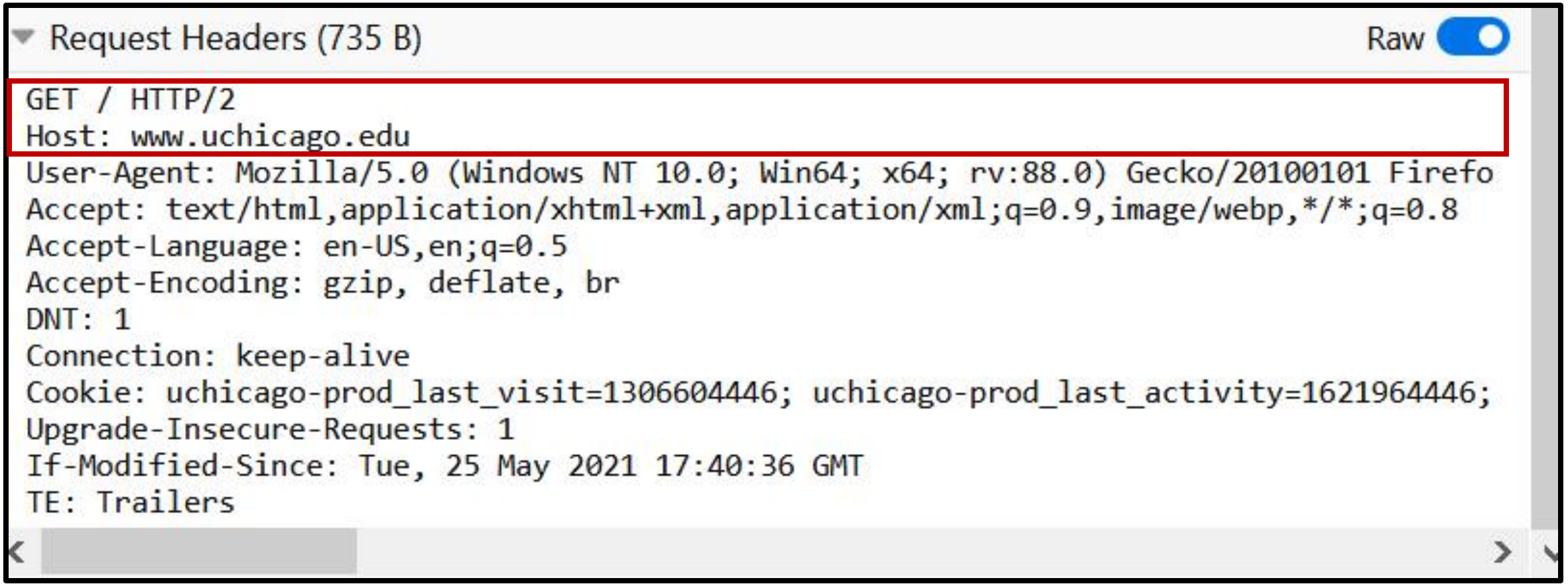
Mechanics of First-Party Online Tracking



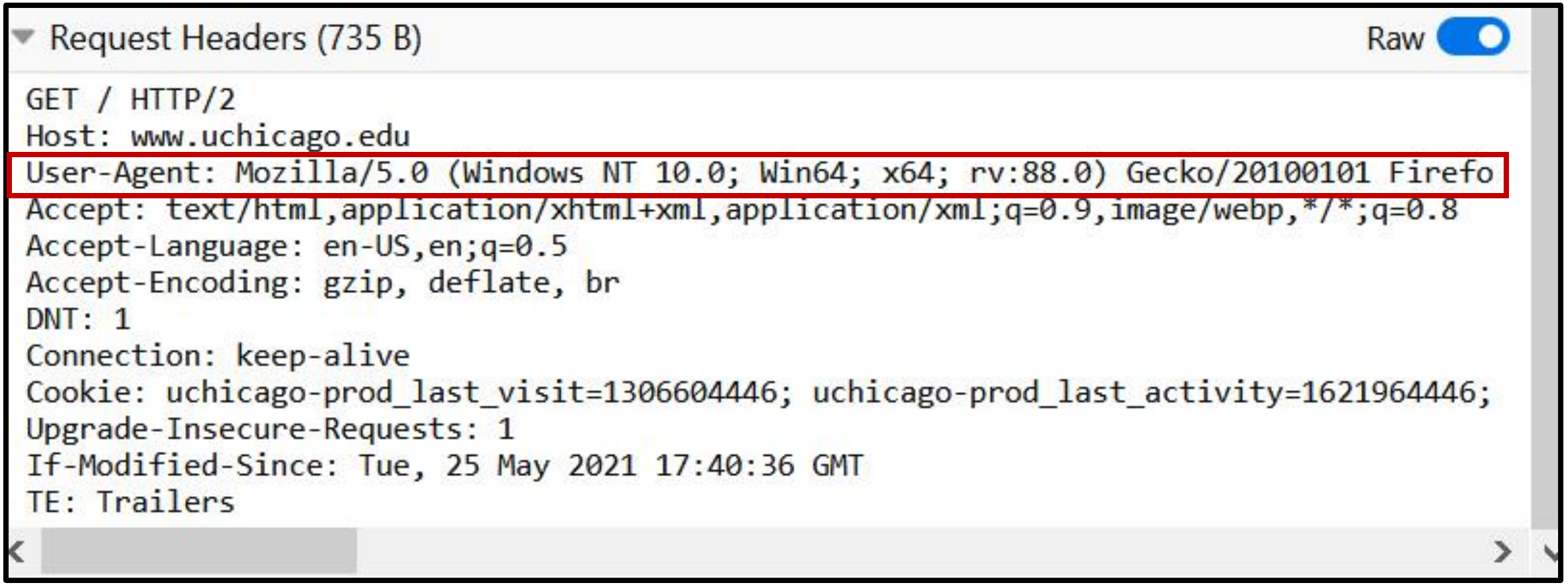
Mechanics of Third-Party Online Tracking



Details of HTTP **Request**



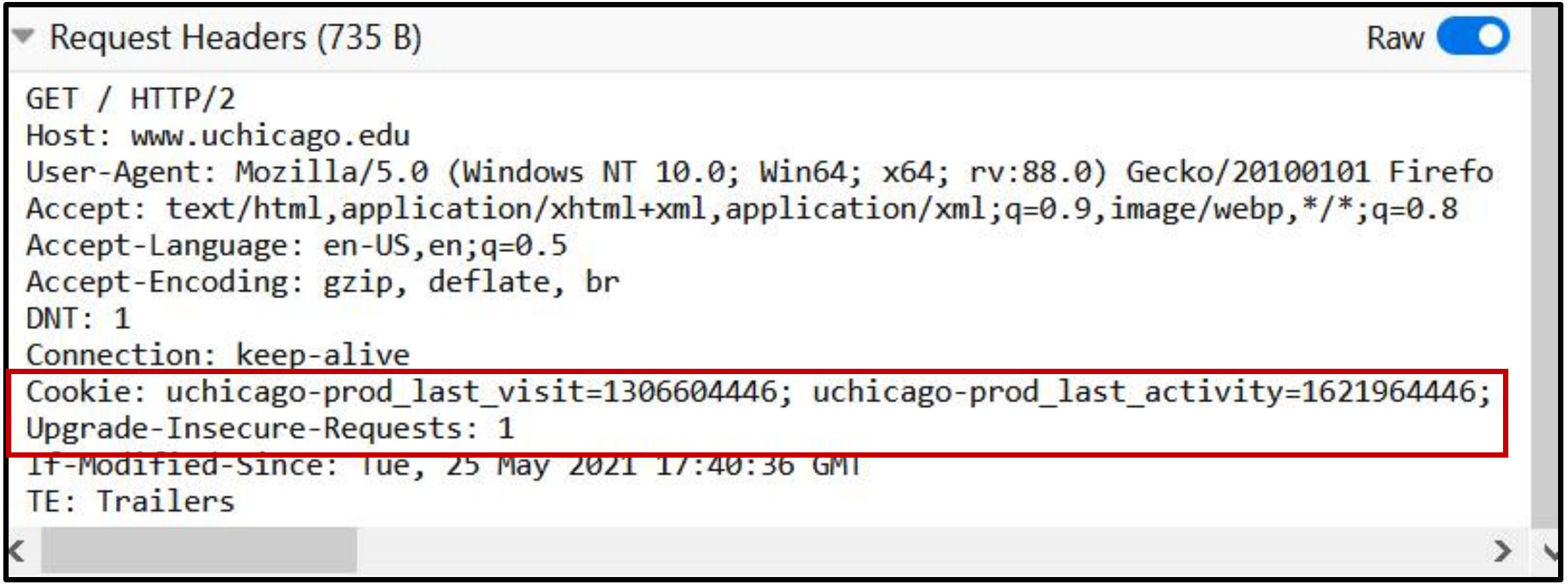
Details of HTTP **Request**



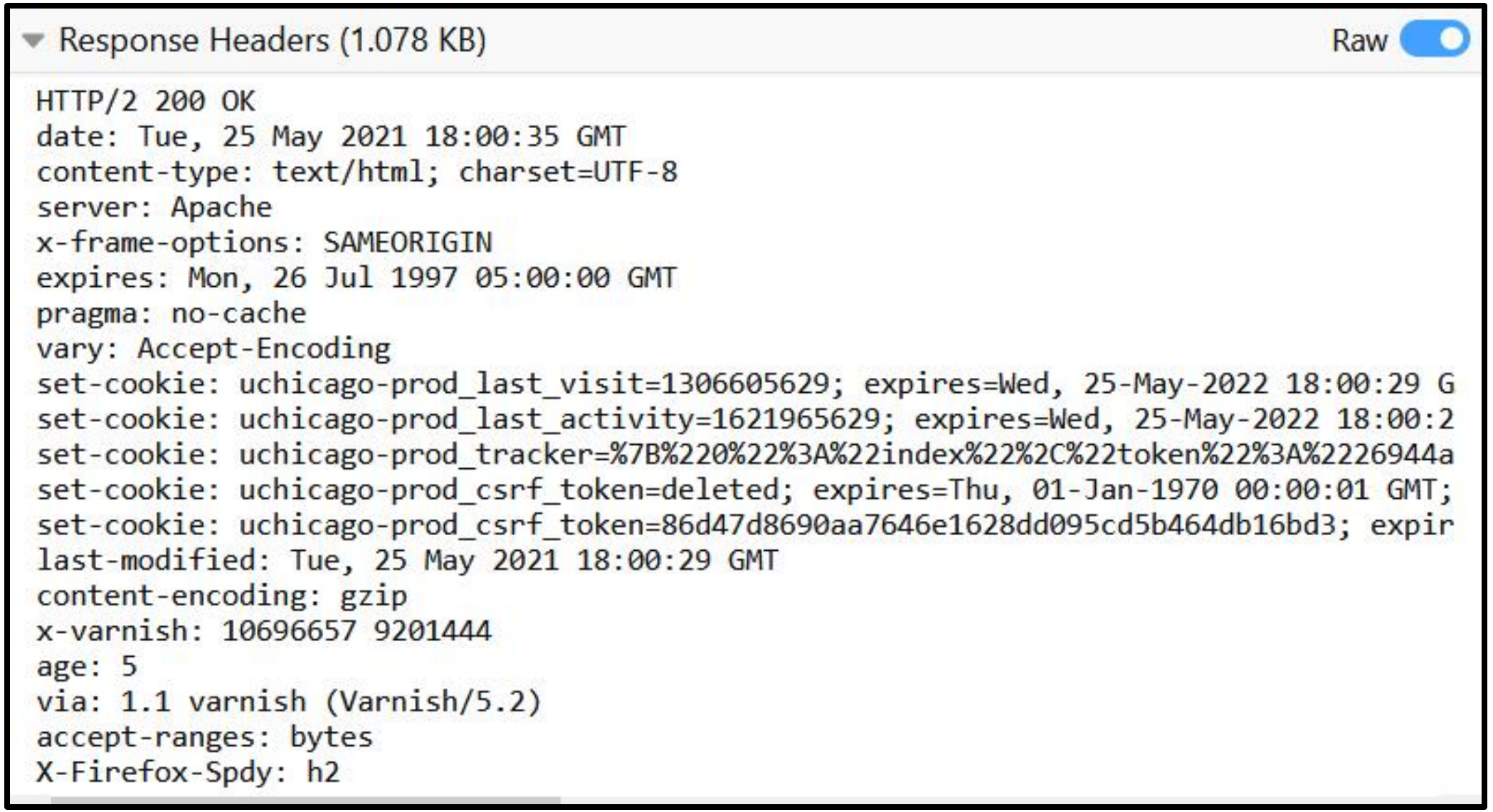
Details of HTTP Cookies

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

Details of HTTP **Request**

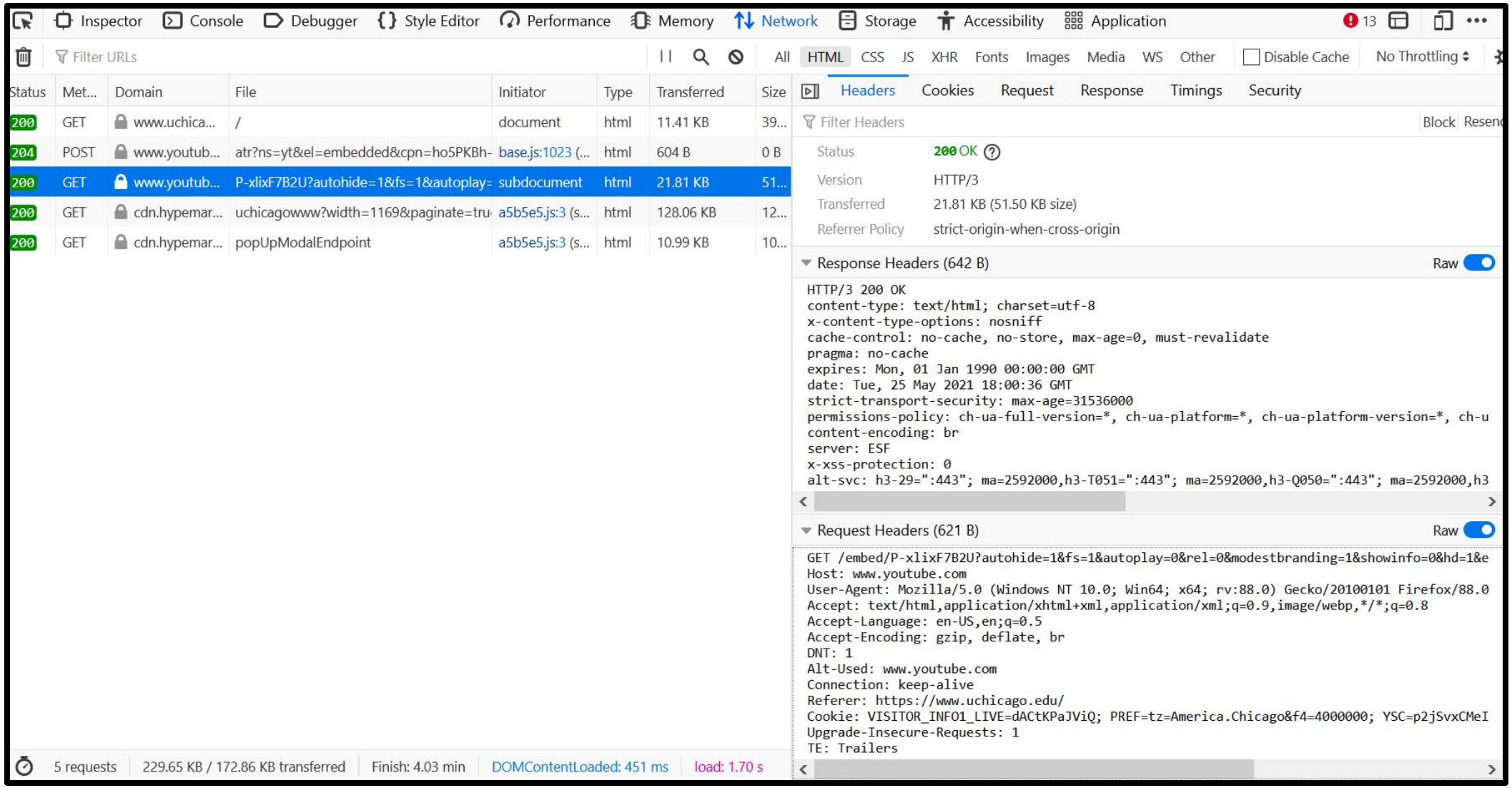


Details of HTTP **Response**

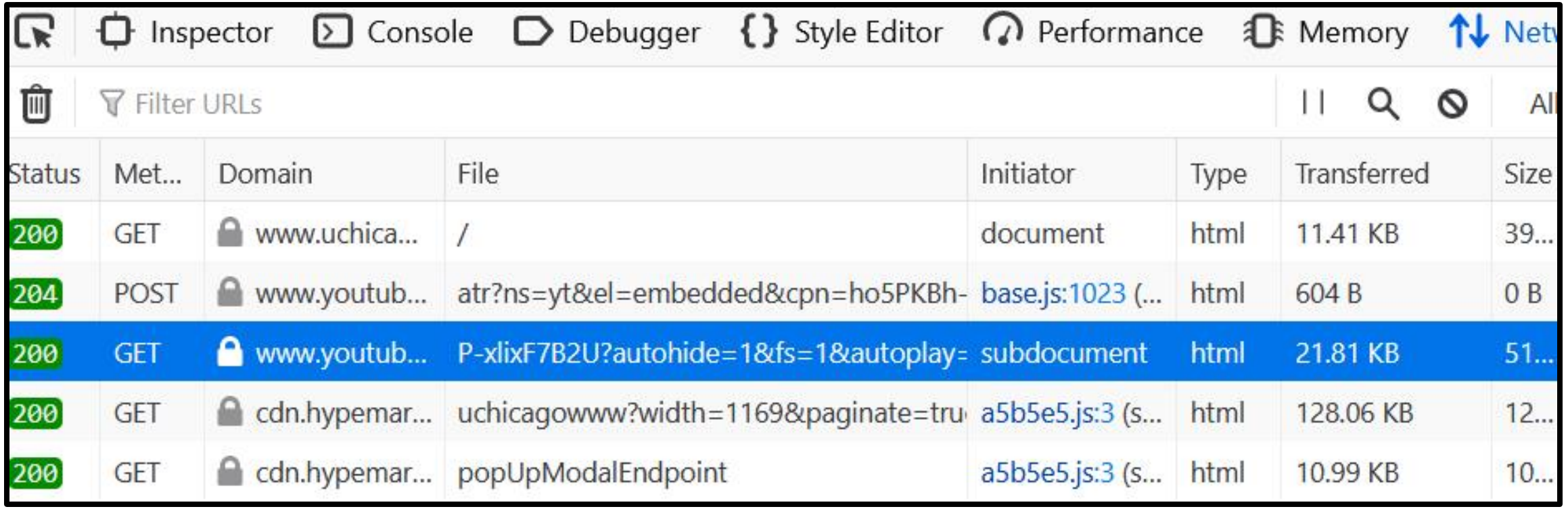


Details of HTTP **Response**

uchicago.edu → youtube.com



Headers (uchicago.edu → youtube.com)



Status	Met...	Domain	File	Initiator	Type	Transferred	Size
200	GET	www.uchica...	/	document	html	11.41 KB	39...
204	POST	www.youtub...	atr?ns=yt&el=embedded&cpn=ho5PKBh-	base.js:1023 (...	html	604 B	0 B
200	GET	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	cdn.hypemar...	popUpModalEndpoint	a5b5e5.js:3 (s...	html	10.99 KB	10...

Headers (uchicago.edu → youtube.com)

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
```

Headers (uchicago.edu → youtube.com)

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
```

Headers (uchicago.edu → youtube.com)

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
```

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

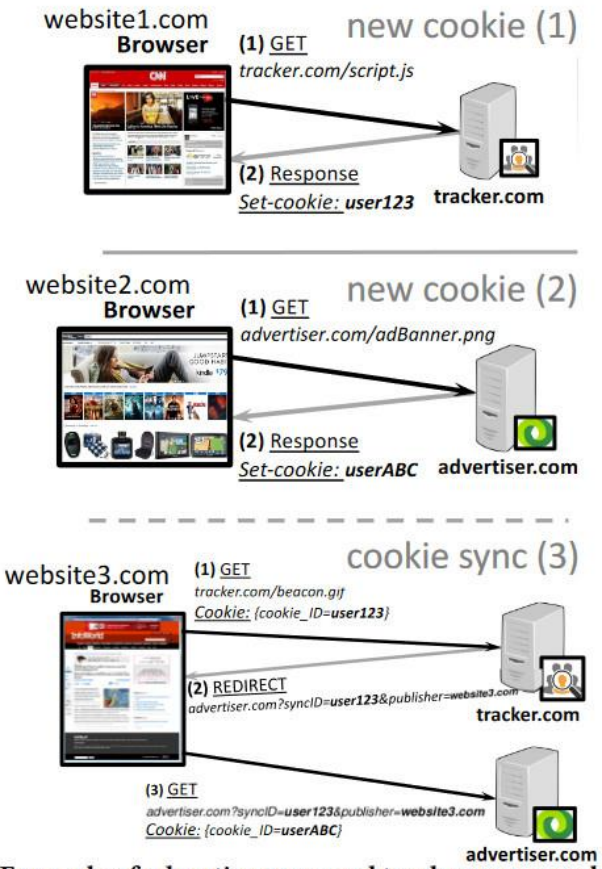


Figure 1: Example of advertiser.com and tracker.com synchronizing their cookieIDs. Interestingly, and without having any code in website3, advertiser.com learns that: (i) cookieIDs userABC==user123 and (ii) userABC has just visited the given website. Finally, both domains can conduct server-to-server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

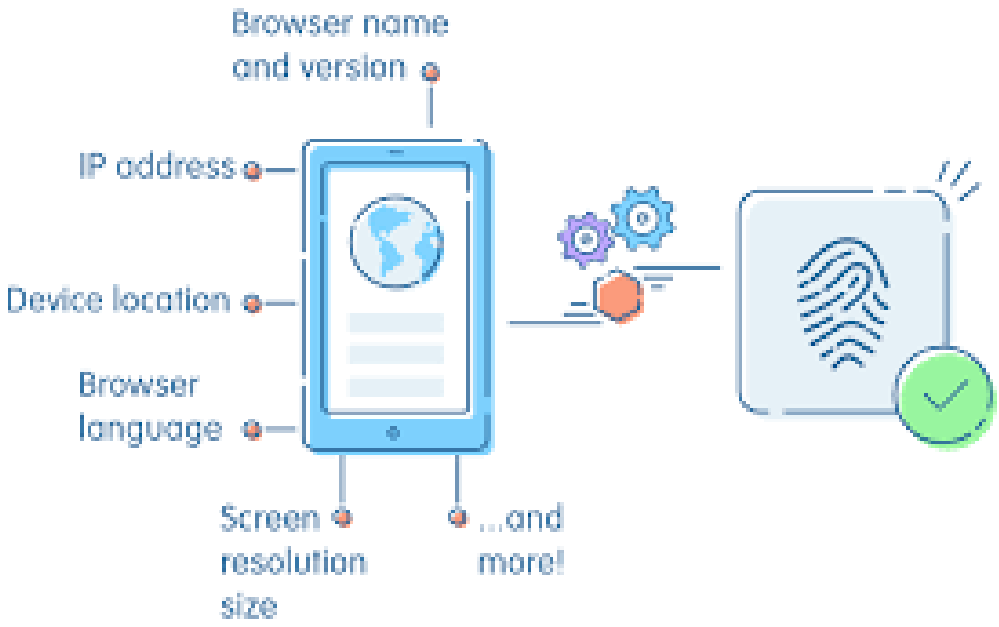
From Papadopoulos et al. “Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask,” in *Proc. WWW*, 2019.

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
- https://coveryourtracks.eff.org/

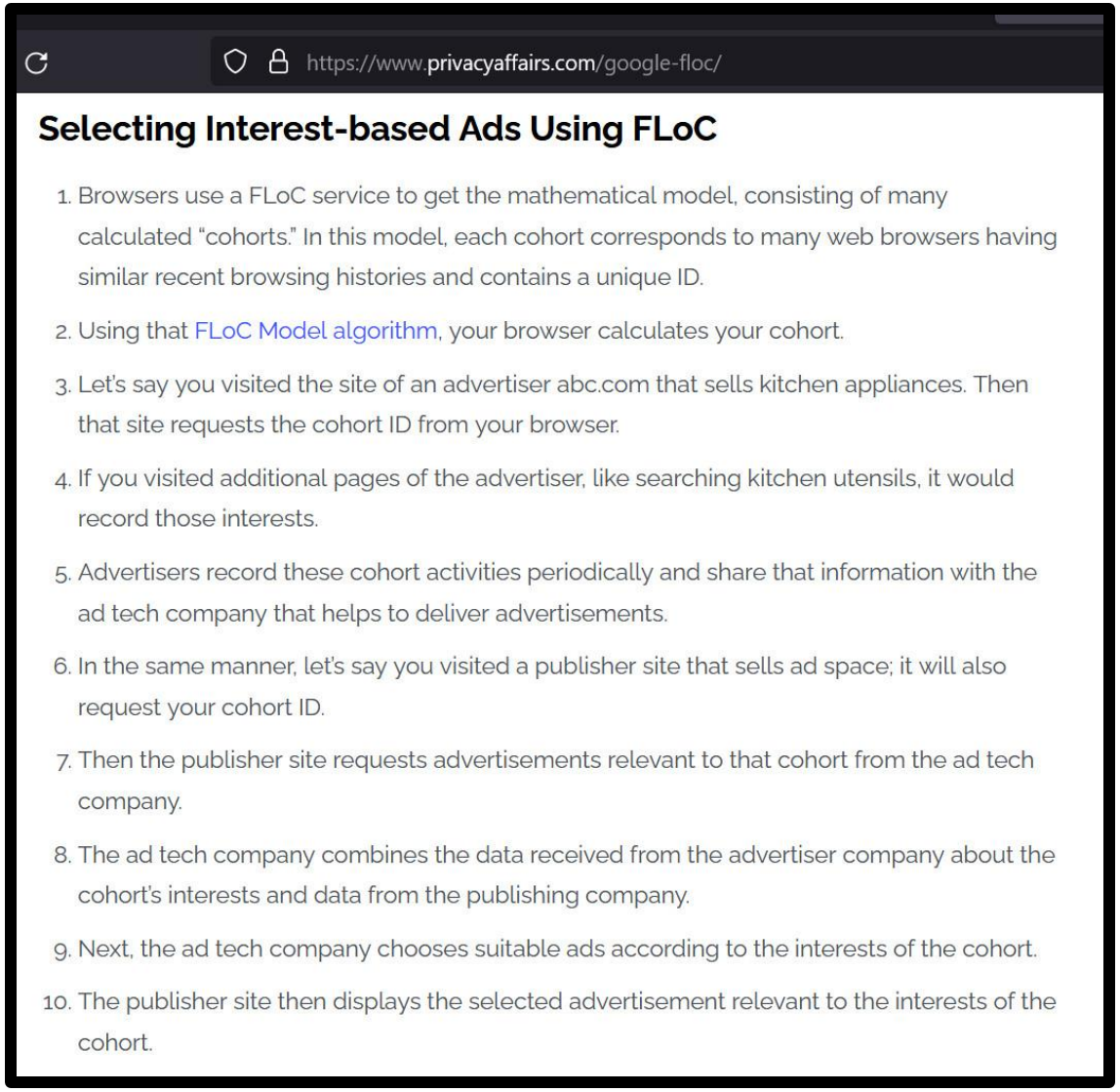


Alternatives to Cookies for Tracking / Profiling

Google’s FLoC

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

Google’s FLoC



Google’s FLoC

EFF ELECTRONIC FRONTIER FOUNDATION

About Issues Our Work Take Action Tools Donate

Google’s FLoC Is a Terrible Idea

BY BENNETT CYPHERS | MARCH 3, 2021

Update, April 9, 2021 : We've launched Am I FLoCed, 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.

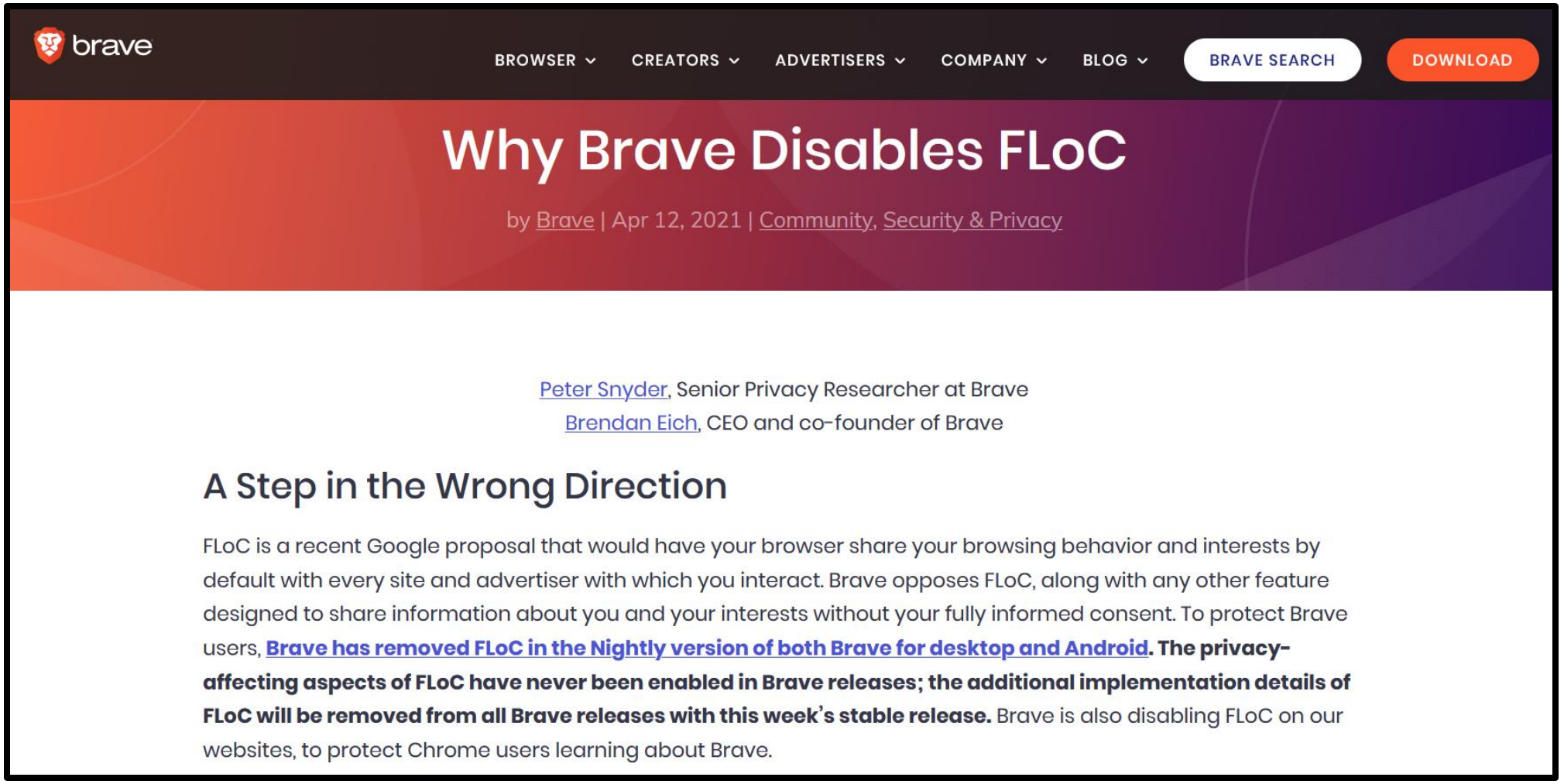
Email Address

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)



brave

BROWSER CREATORS ADVERTISERS COMPANY BLOG BRAVE SEARCH DOWNLOAD

Why Brave Disables FLoC

by Brave | Apr 12, 2021 | Community, Security & Privacy

Peter Snyder, Senior Privacy Researcher at Brave
Brendan Eich, 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/

Google's FLoC

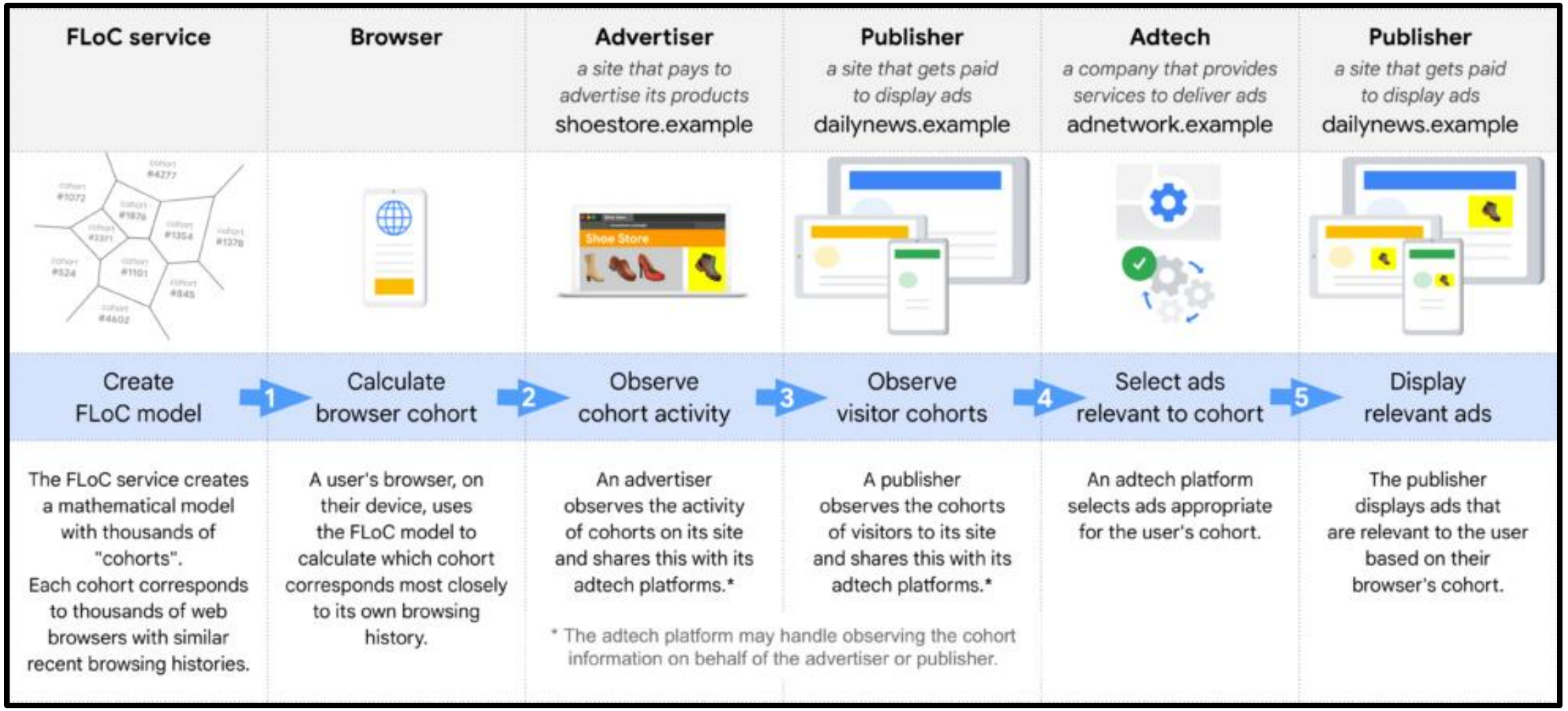
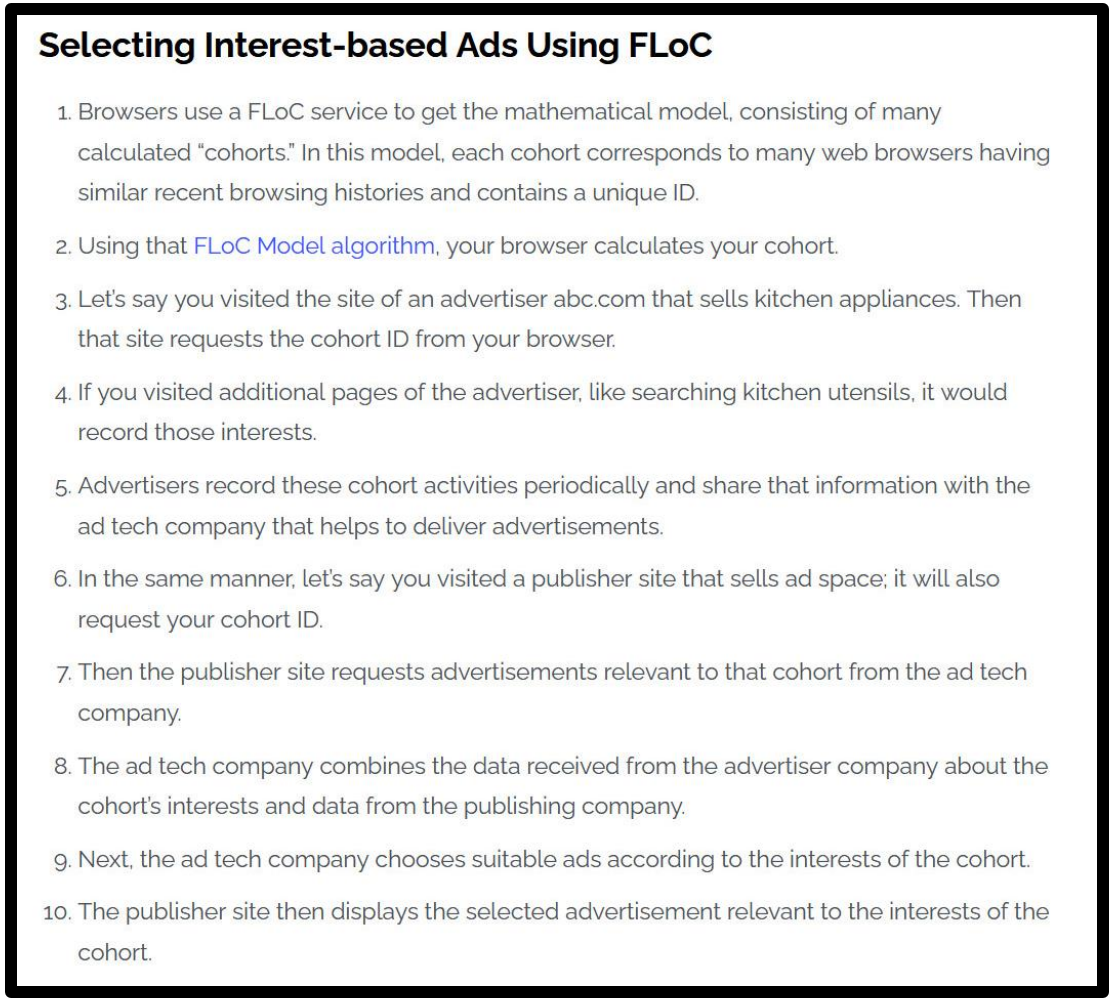


Image taken from https://arstechnica.com/gadgets/2021/04/everybody-hates-floc-googles-tracking-plan-for-chrome-ads/

Google’s FLoC

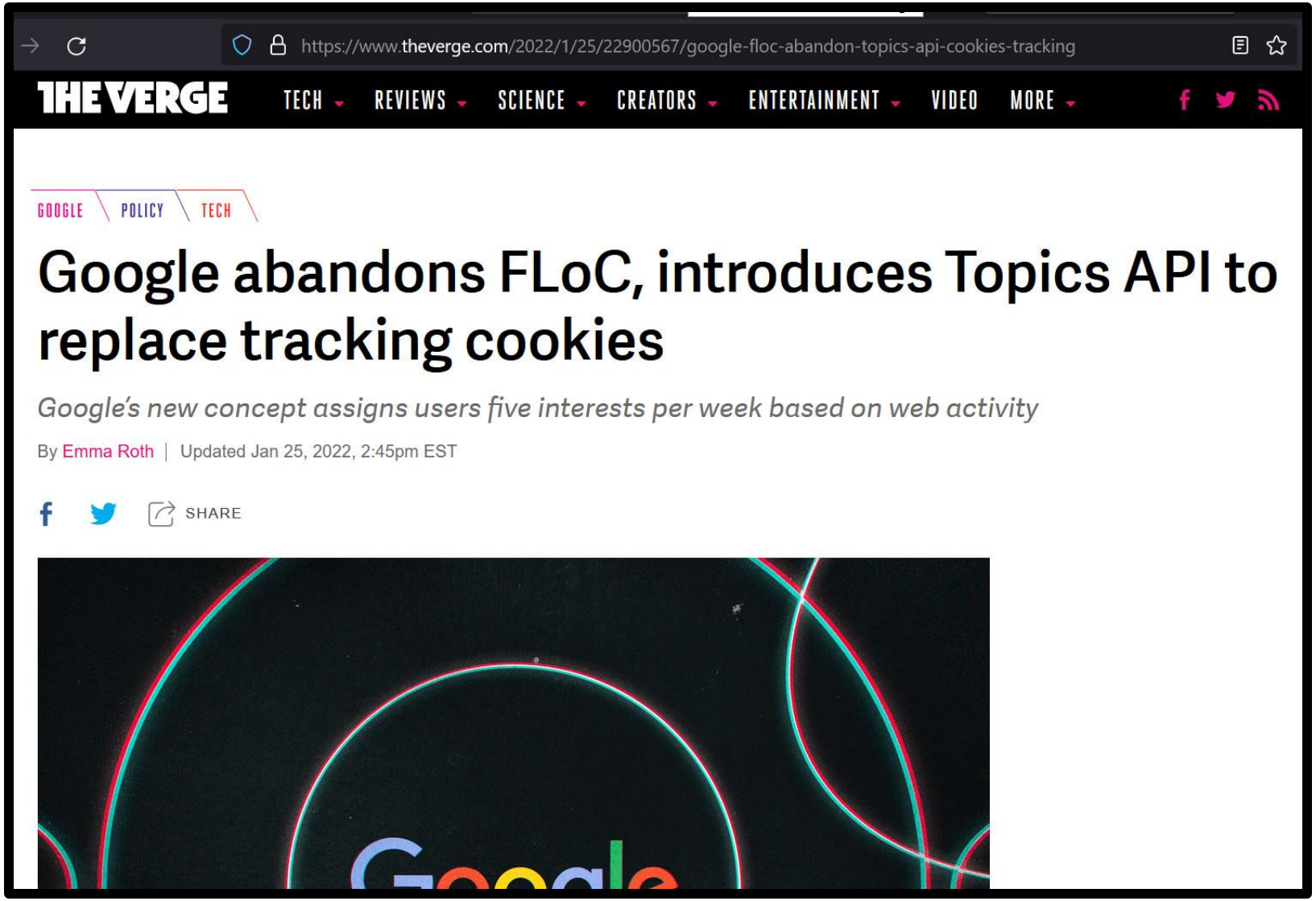


Selecting Interest-based Ads Using FLoC

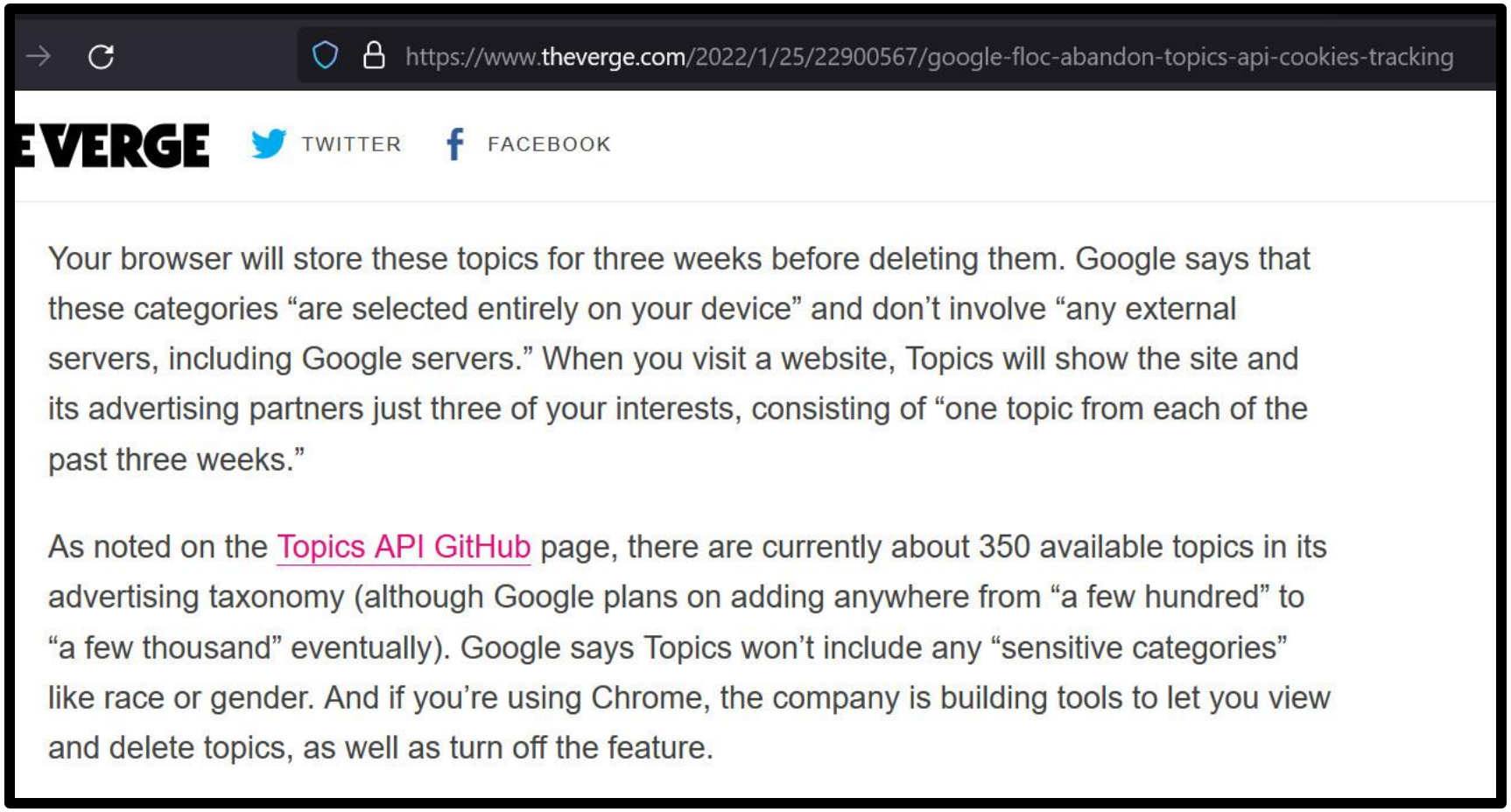
1. Browsers use a FLoC service to get the mathematical model, consisting of many calculated “cohorts.” In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.
2. Using that FLoC Model algorithm, your browser calculates your cohort.
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



Studying User Reactions to (Future) Hyper-Targeting

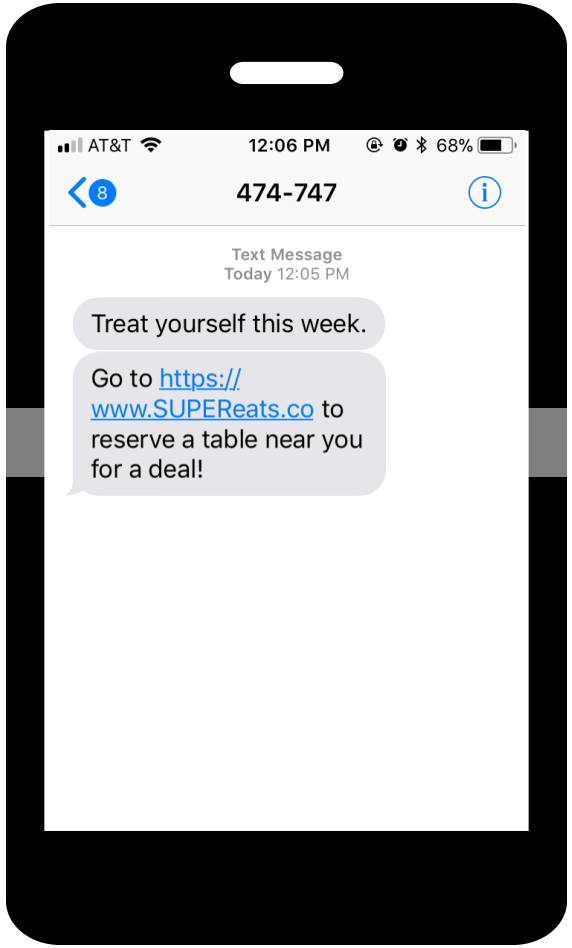
Study: Banner Ads

Generic

Personalized

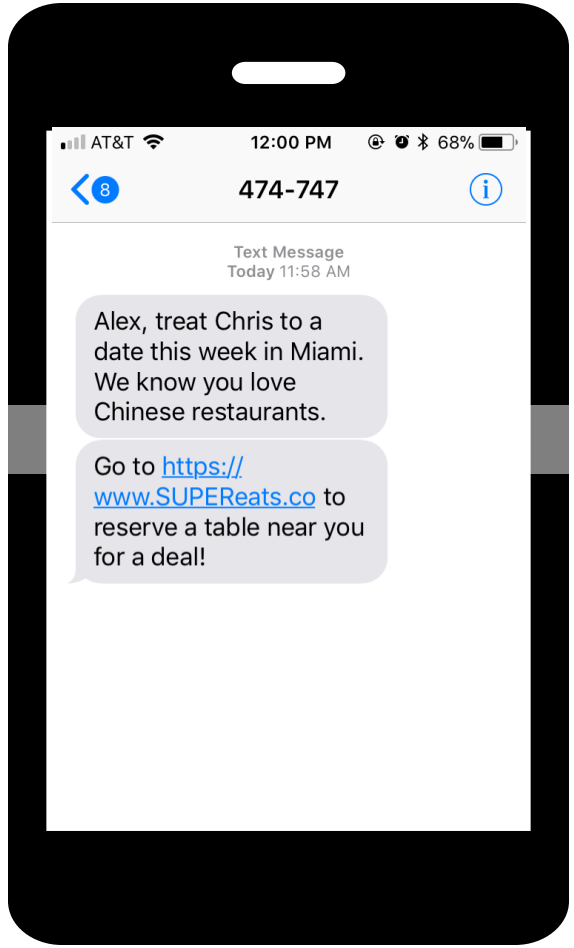
See https://www.blaseur.com/papers/robotext-full.pdf

Study: Robotext Ads



Generic

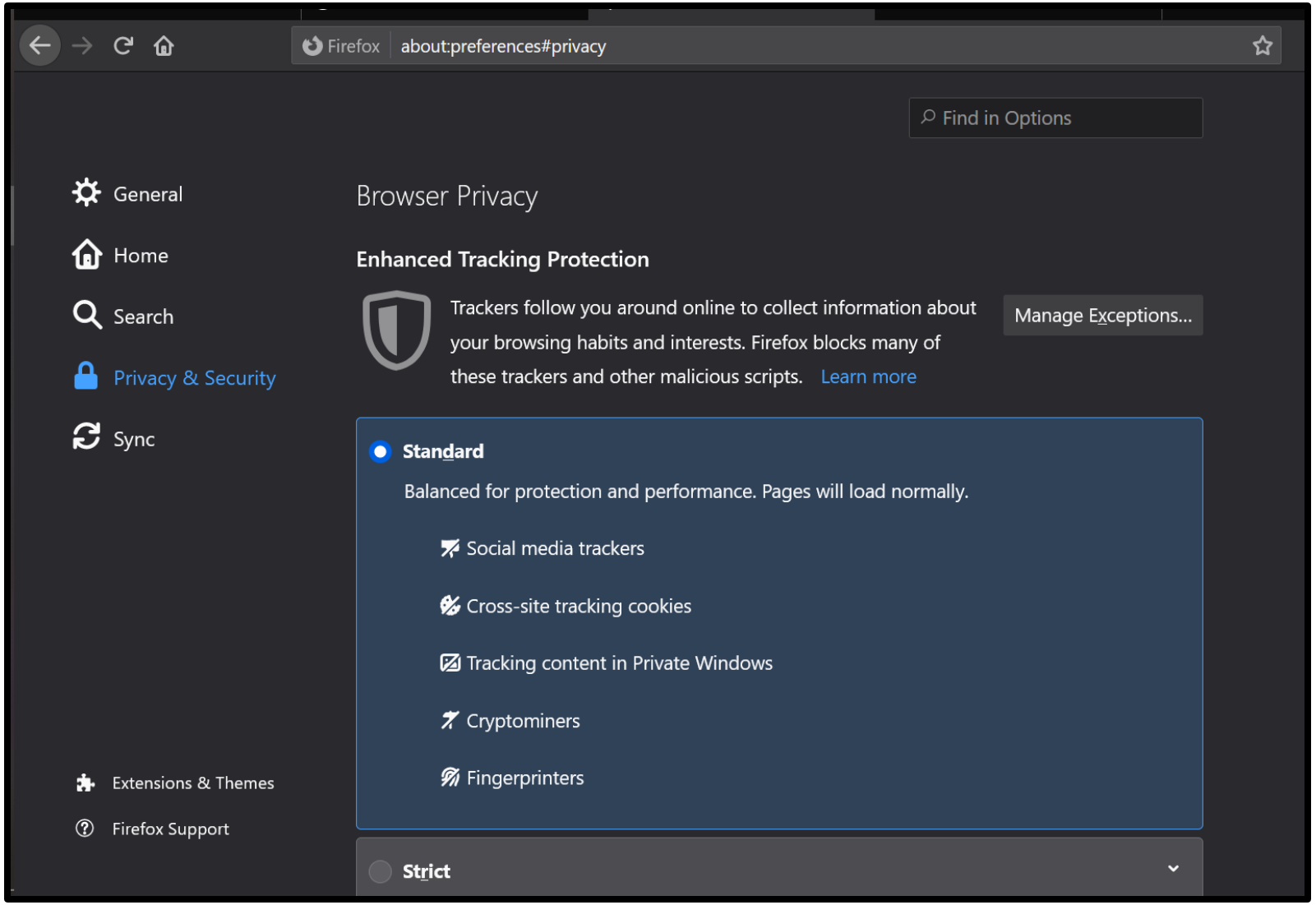
Personalized



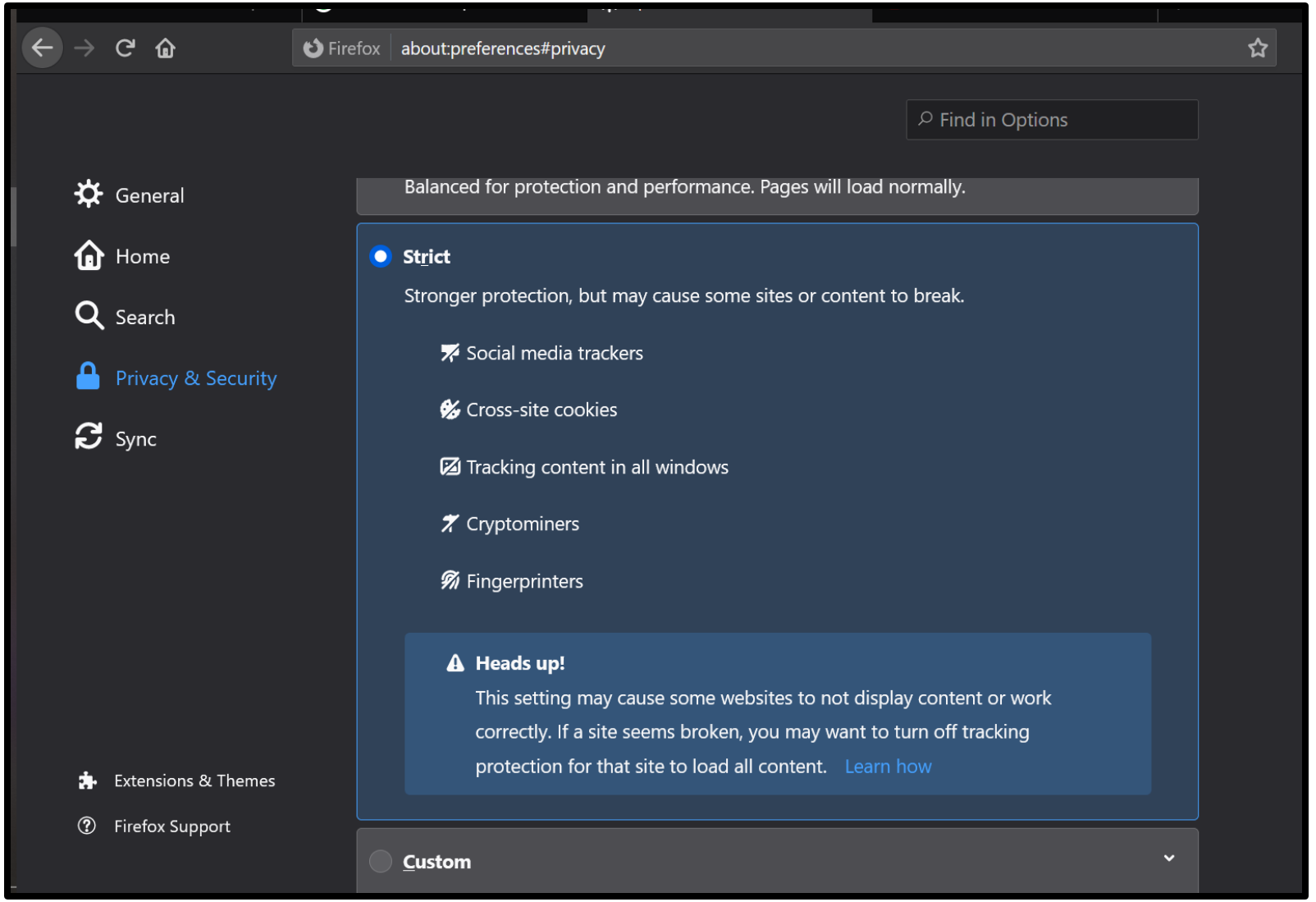
See https://www.blaseur.com/papers/robotext-full.pdf

Anti-Tracking Tools

Firefox Browser Default Settings

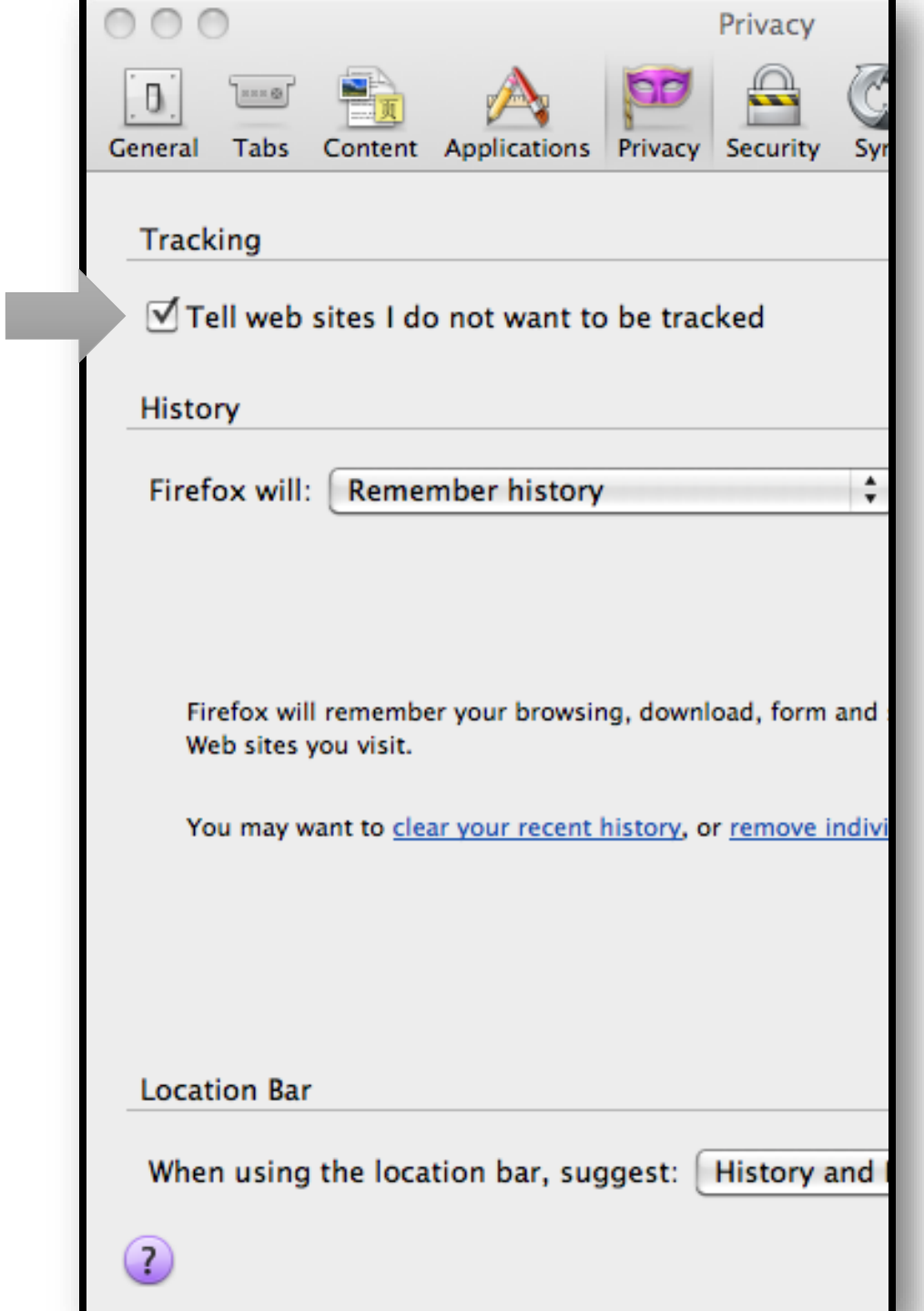


Firefox Browser Strict Settings



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



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

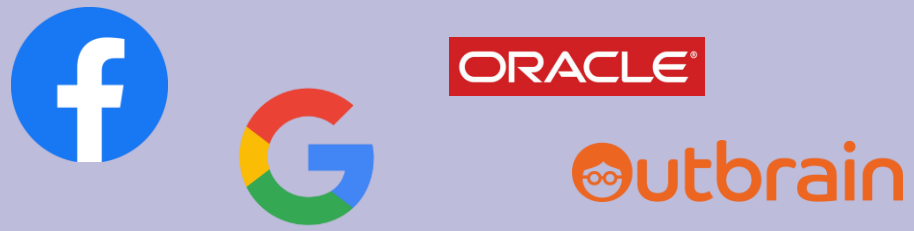
Transparency Tools

Current Transparency Tools

Privacy-Focused Browser Extensions

Browsers' Built-in Tools

Companies' Privacy Dashboards



Oh, the Places You've Been!

User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing

By

Ben Weinshel
Miranda Wei
Mainack Mondal
Euirim Choi
Shawn Shan
Claire Dolin
Michelle L. Mazurek
Blase Ur

ACM CCS 2019
November 11-15, 2019, London, UK

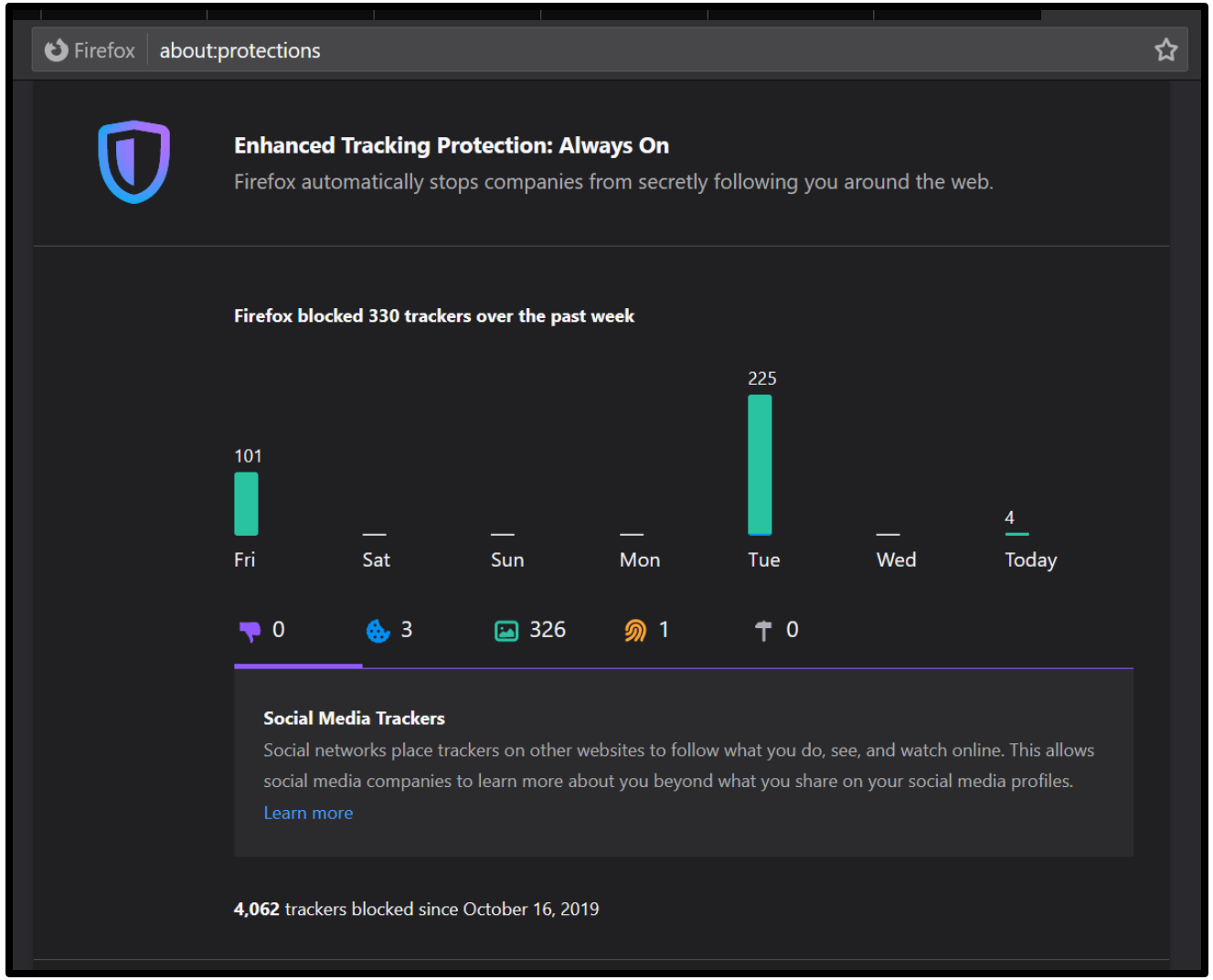
THE UNIVERSITY OF CHICAGO

SUPER GROUP — Security, Usability, & Privacy Education & Research

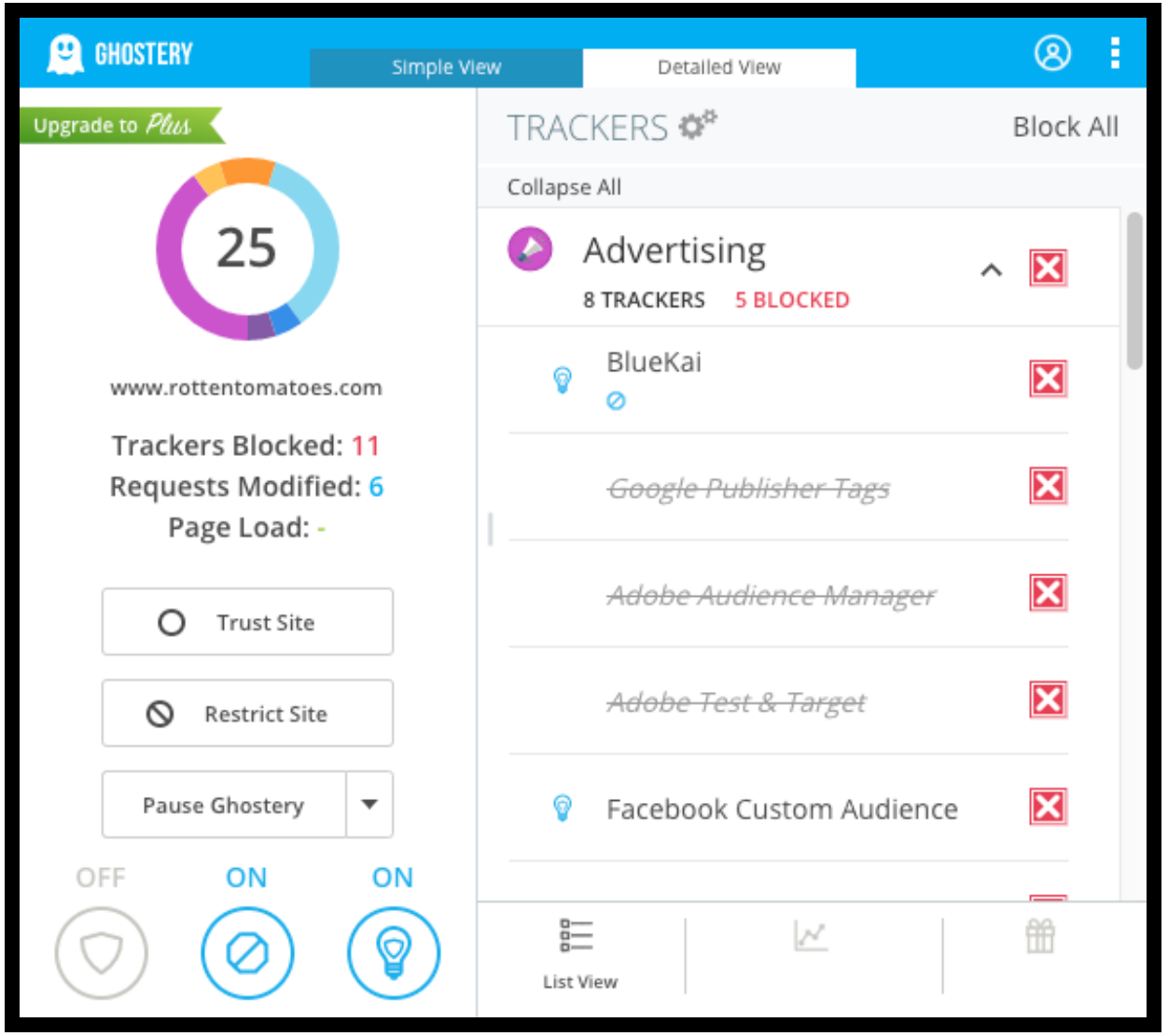
UNIVERSITY OF MARYLAND

SP² SECURITY. PRIVACY. PEOPLE

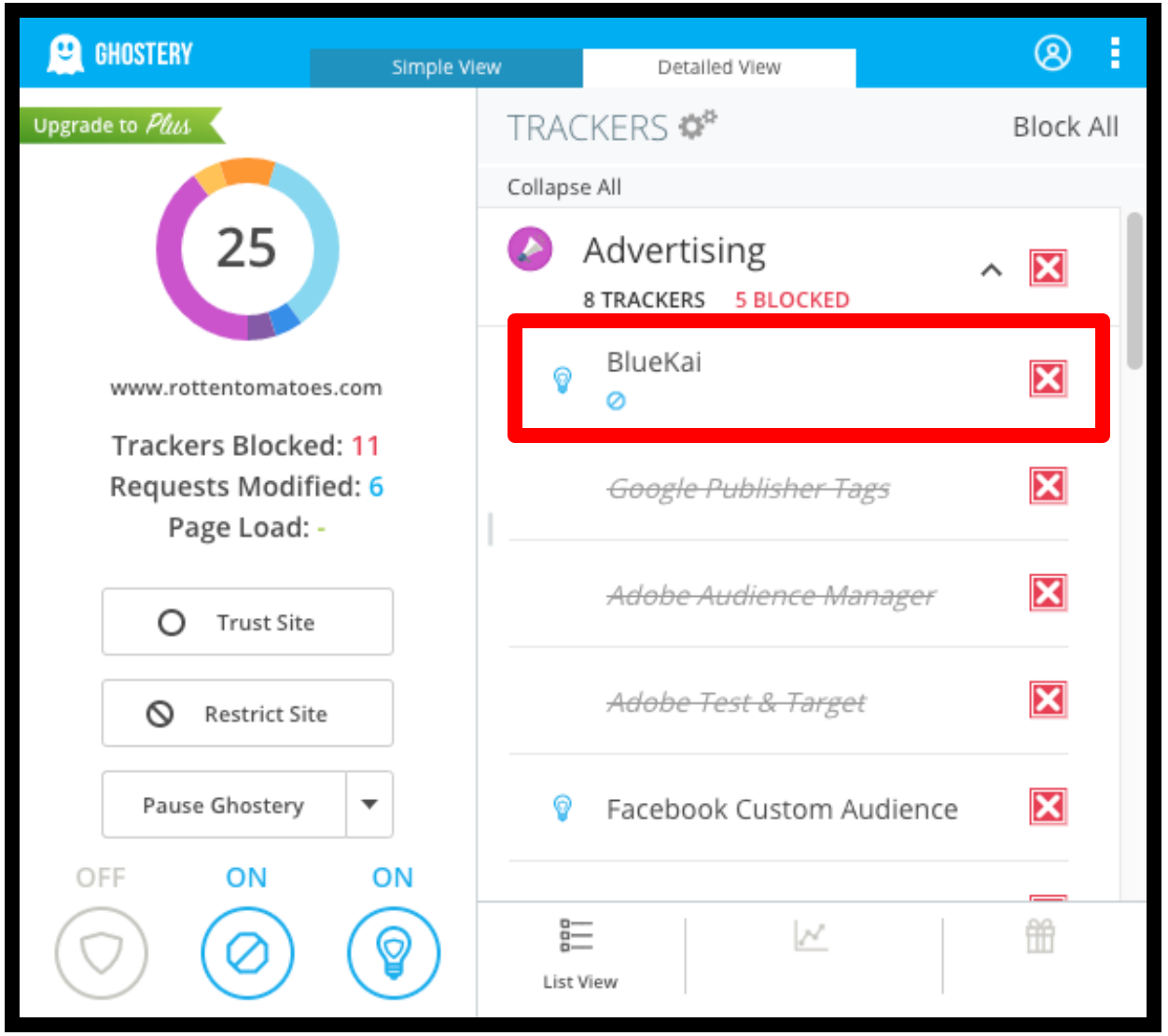
Current Tools Provide High-Level Information



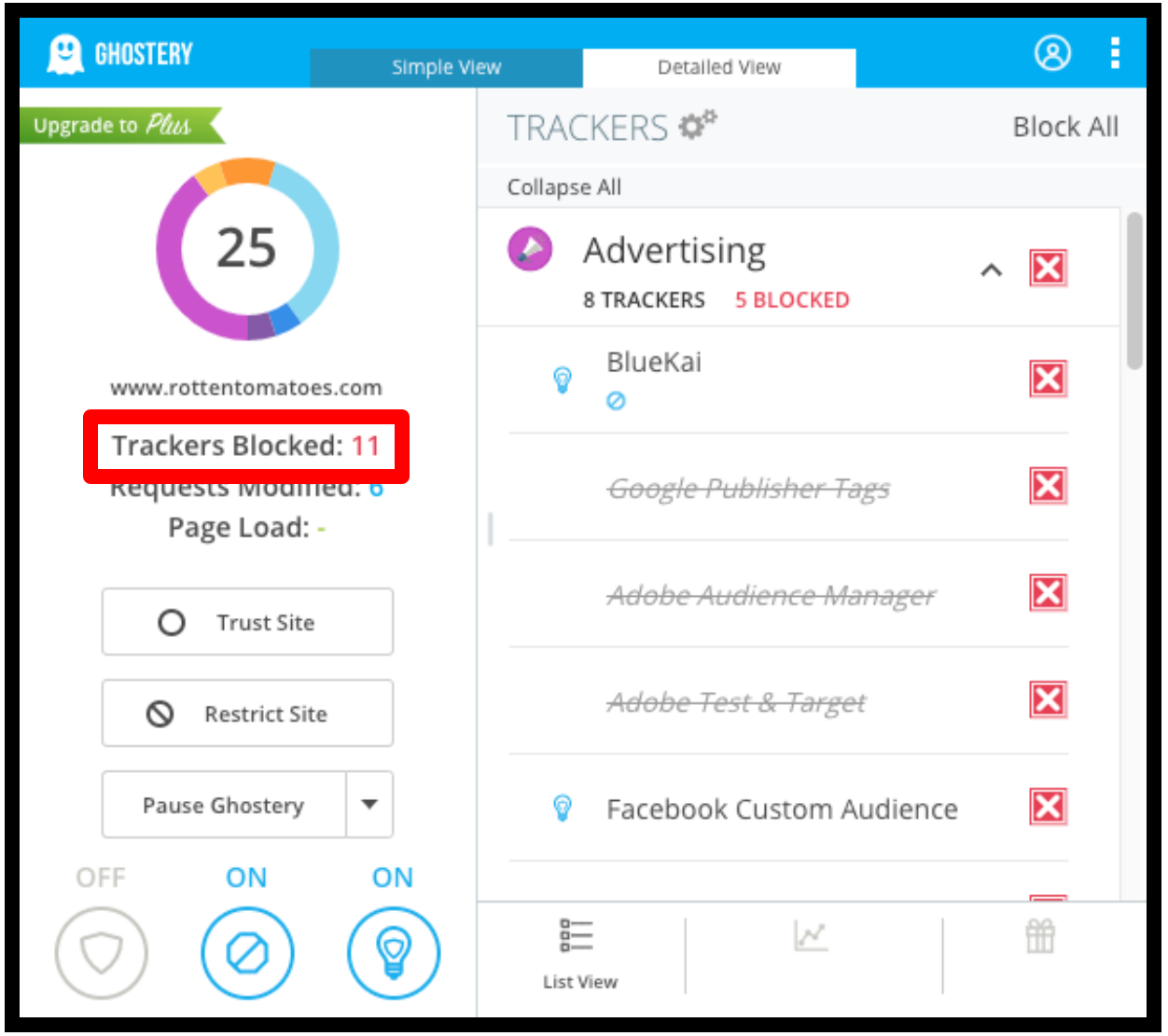
Current Tools Provide High-Level Information



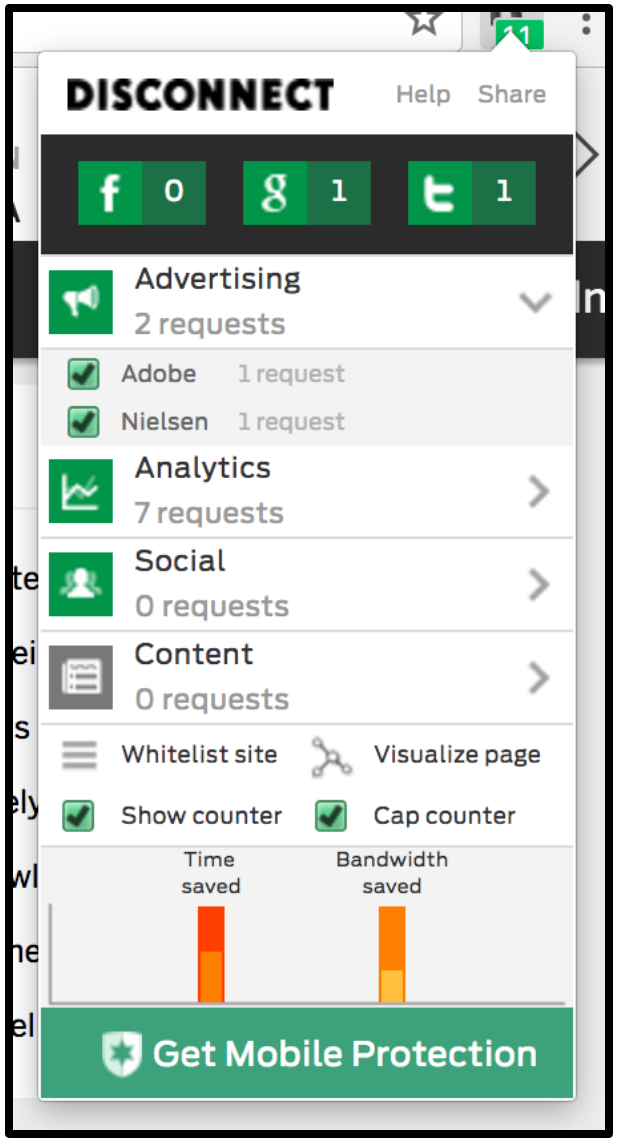
Current Tools Provide High-Level Information

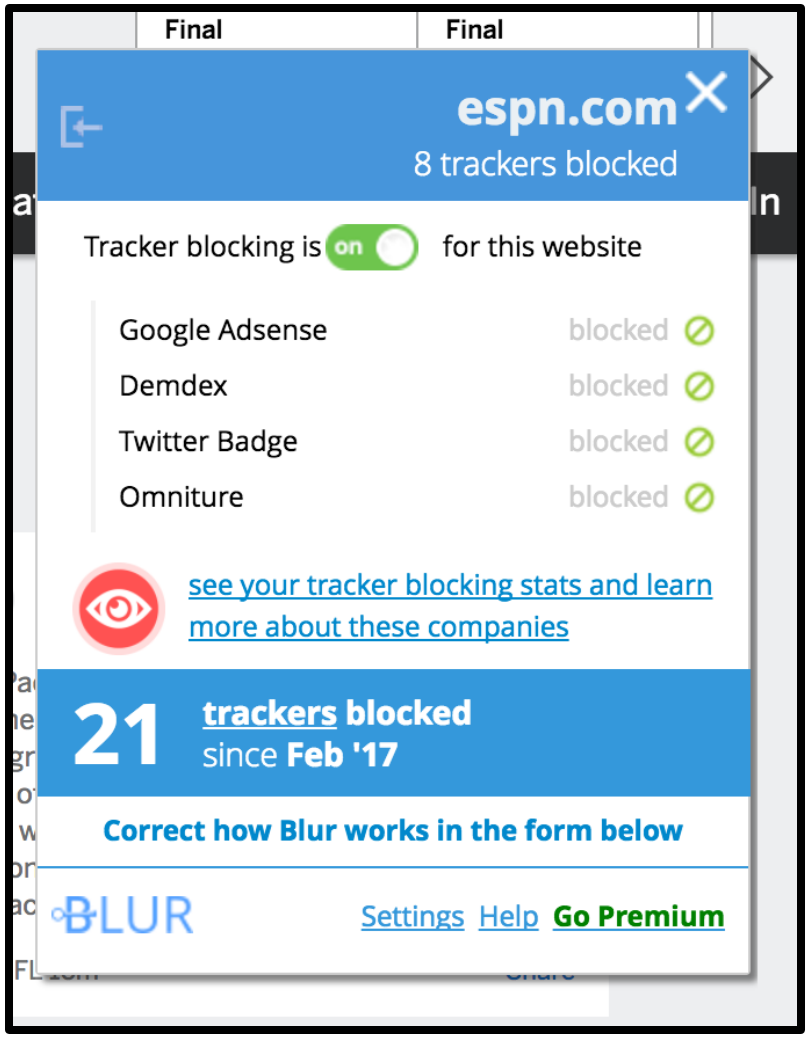


Current Tools Provide High-Level Information

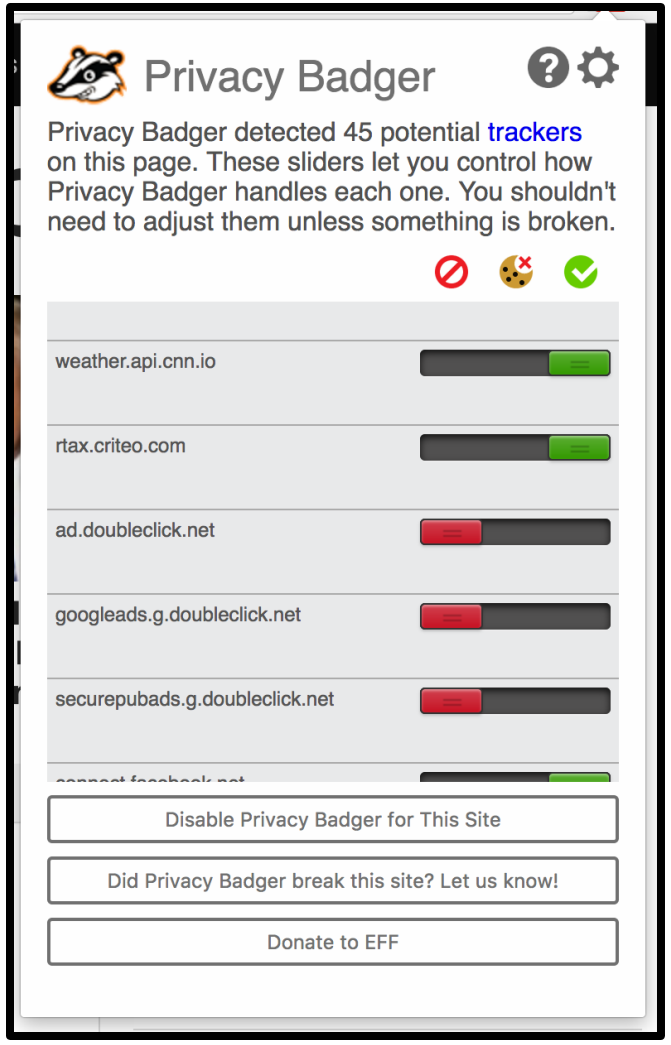


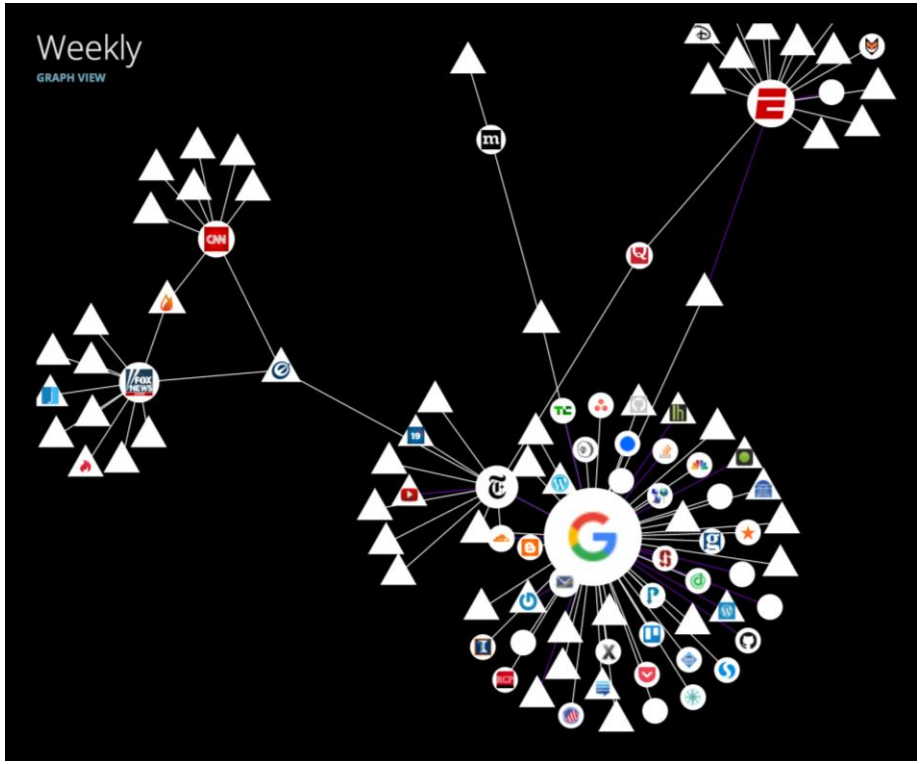
Current Tools Provide High-Level Information

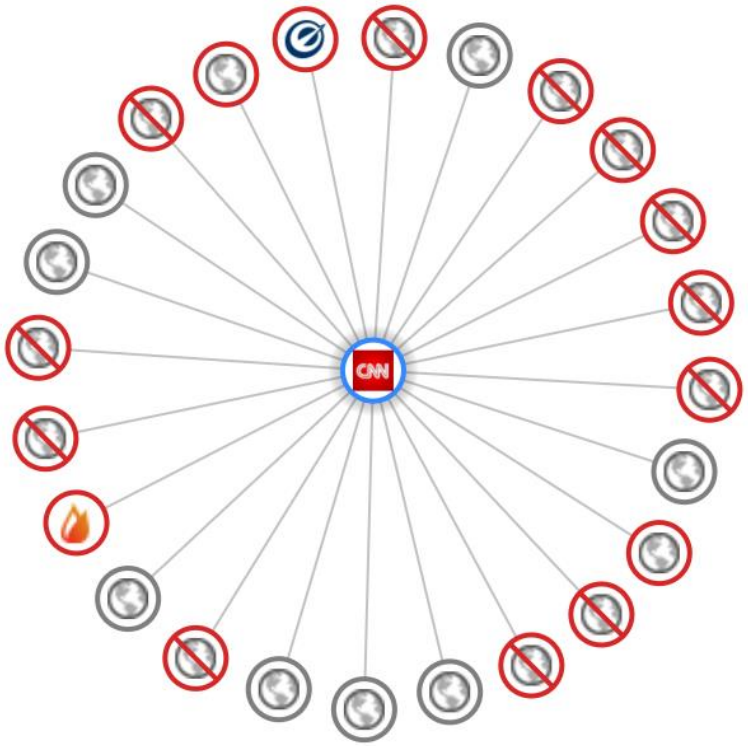




Current Tools Provide High-Level Information

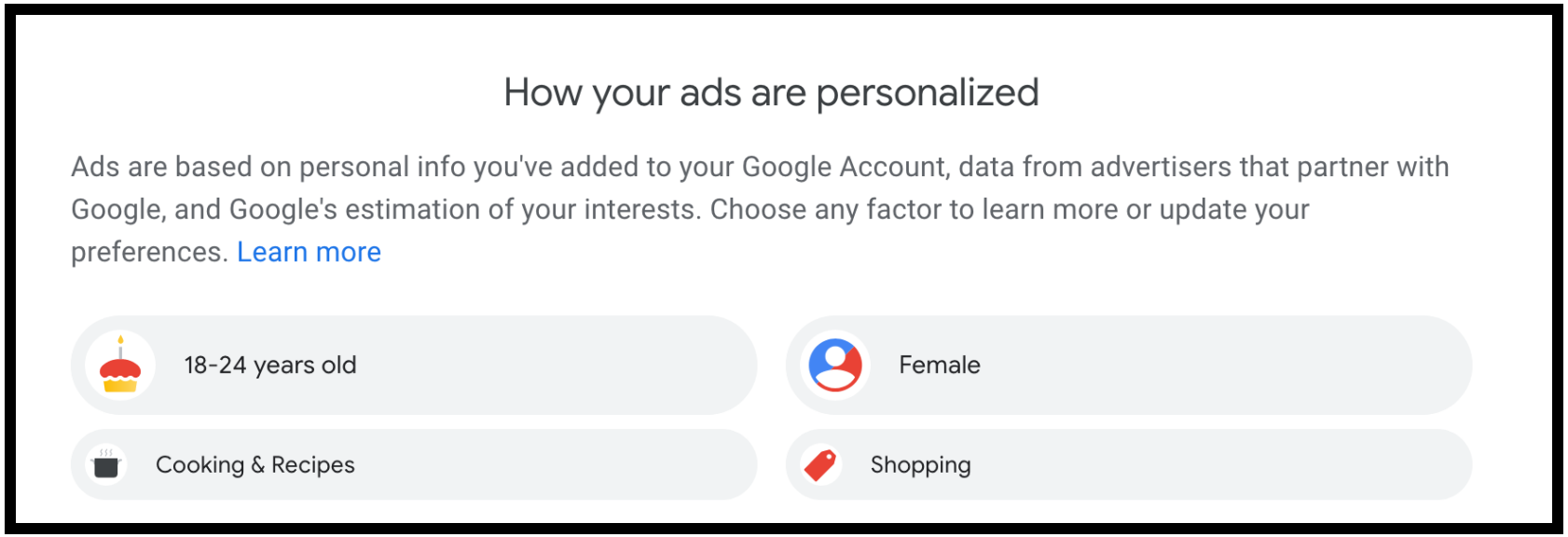






Current Tools Provide High-Level Information

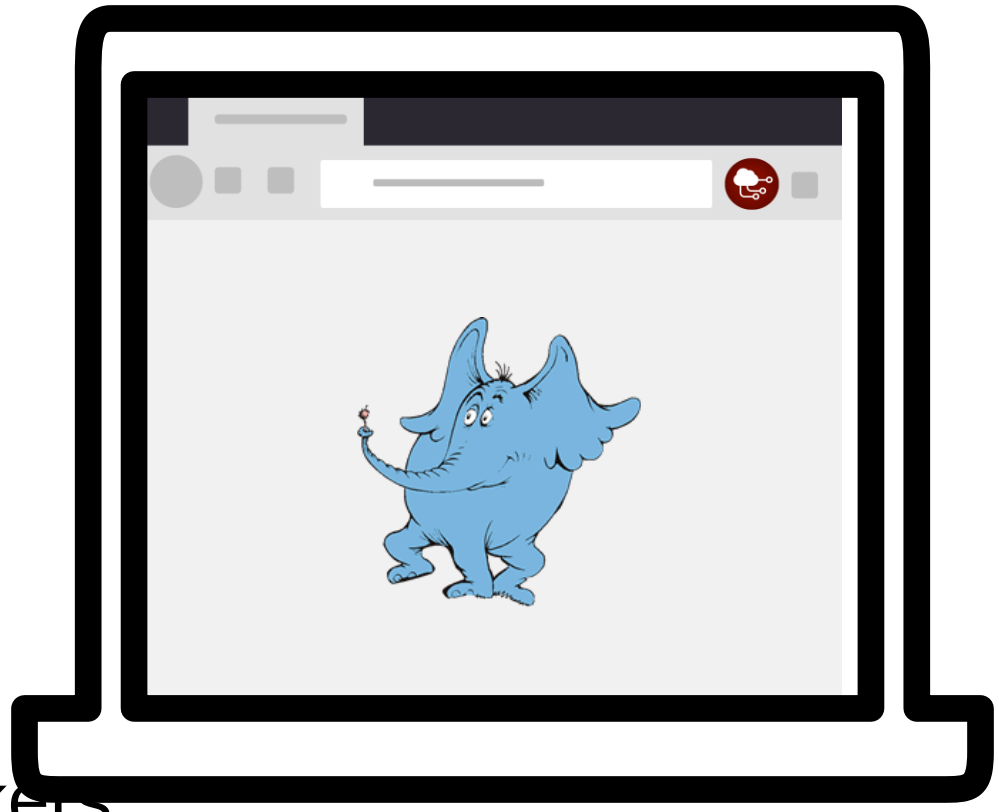
Current Tools Provide High-Level Information



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

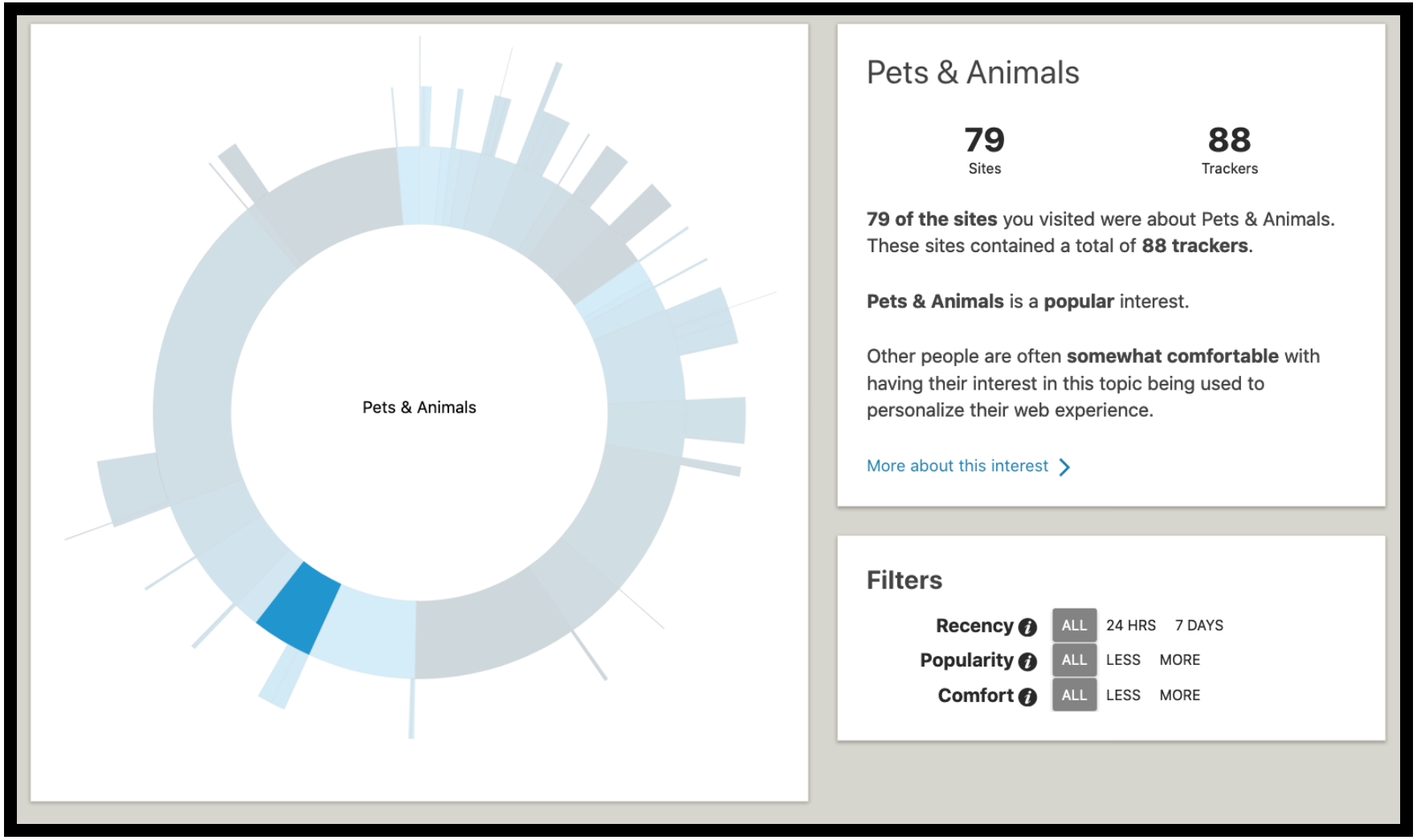
Design of Tracking Transparency (Research)



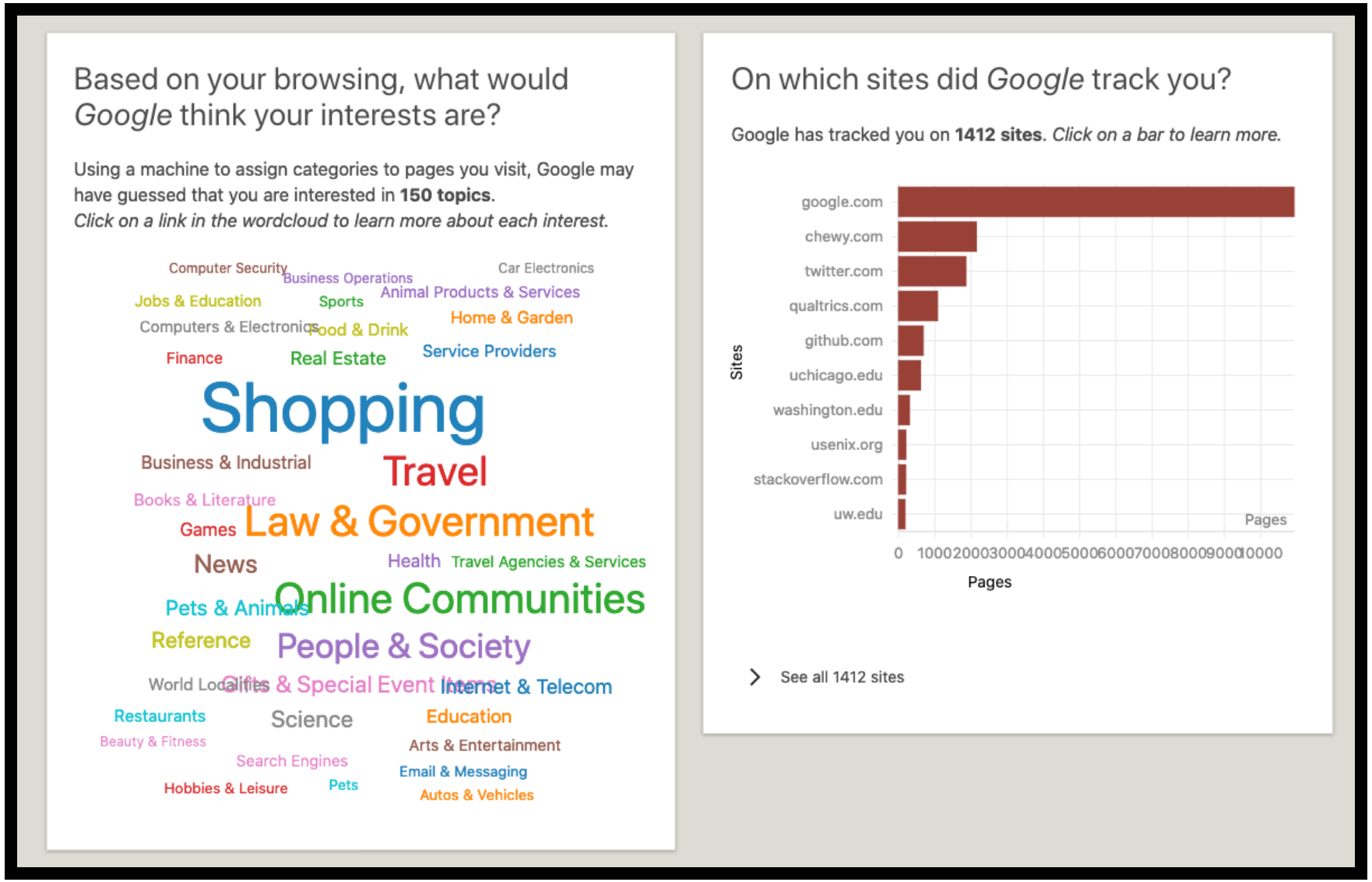
- Web browser extension
- *Locally* records tracking
  - Page metadata, page contents, trackers
  - Profiles potential interests from that page

https://git.io/trackingtransparency

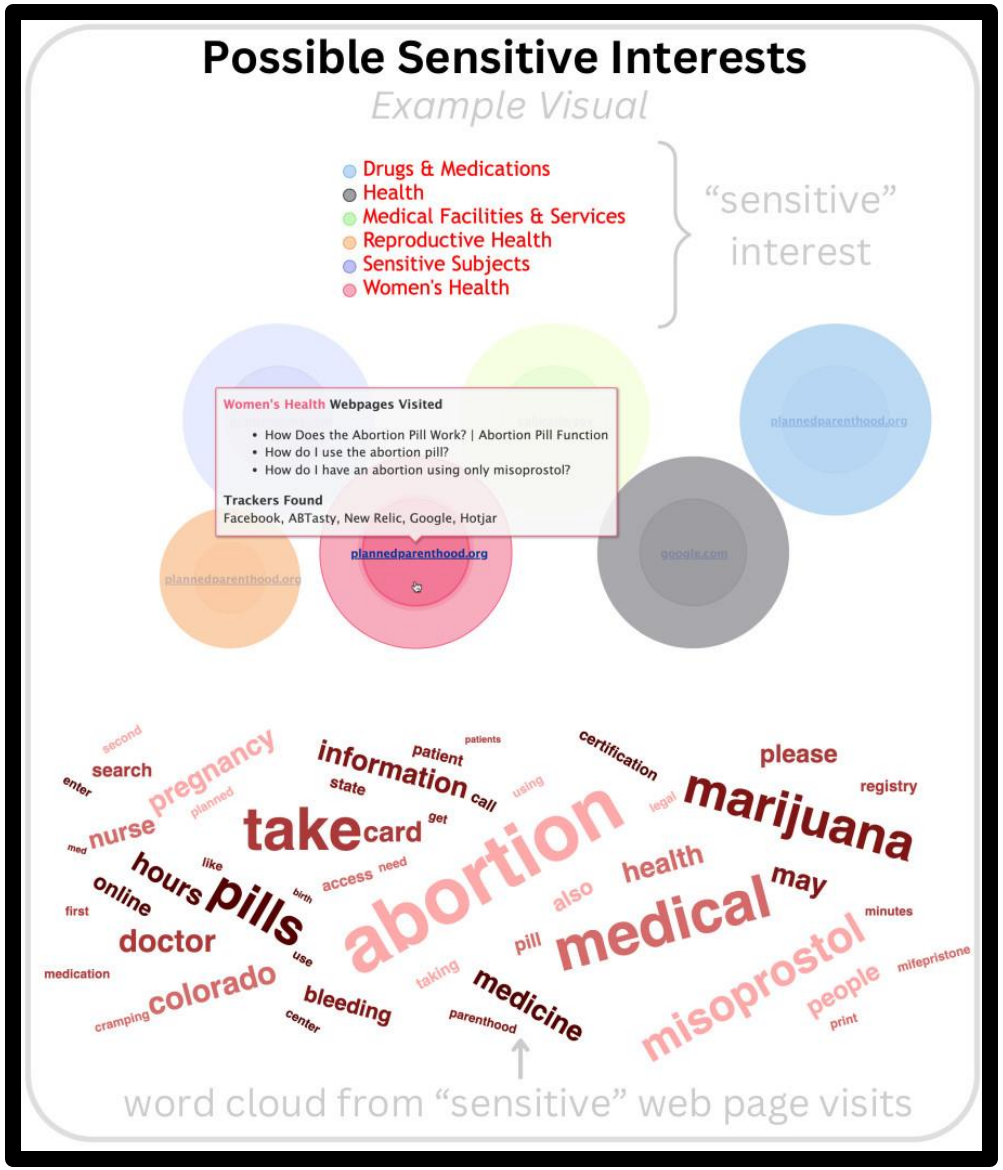
Potential Interests



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.
Computer Security
Business Operations
Car Electronics
Jobs & Education
Sports
Animal Products & Services
Computers & Electronics
Food & Drink
Home & Garden
Finance
Real Estate
Service Providers
Shopping
Business & Industrial
Travel
Books & Literature
Games
Law & Government
News
Health
Travel Agencies & Services
Pets & Animals
Online Communities
Reference
People & Society
World Localities
Gifts & Special Event Items
Internet & Telecom
Restaurants
Science
Education
Beauty & Fitness
Arts & Entertainment
Search Engines
Email & Messaging
Hobbies & Leisure
Pets
Autos & Vehicles
On which sites did *Google* track you?
Google has tracked you on **1412 sites**. *Click on a bar to learn more.*
google.com
chewy.com
twitter.com
qualtrics.com
github.com
uchicago.edu
washington.edu
usenix.org
stackoverflow.com
uw.edu
Sites
Pages
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
Pages
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

THE UNIVERSITY OF CHICAGO

UNIVERSITY of WASHINGTON

UNIVERSITY OF MARYLAND

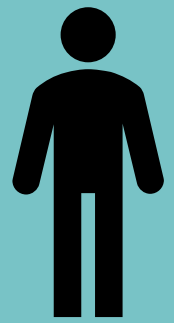
UNIVERSITY OF Southampton

29TH USENIX SECURITY SYMPOSIUM

Possible Expectation of Targeted Advertising

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

Expectation:



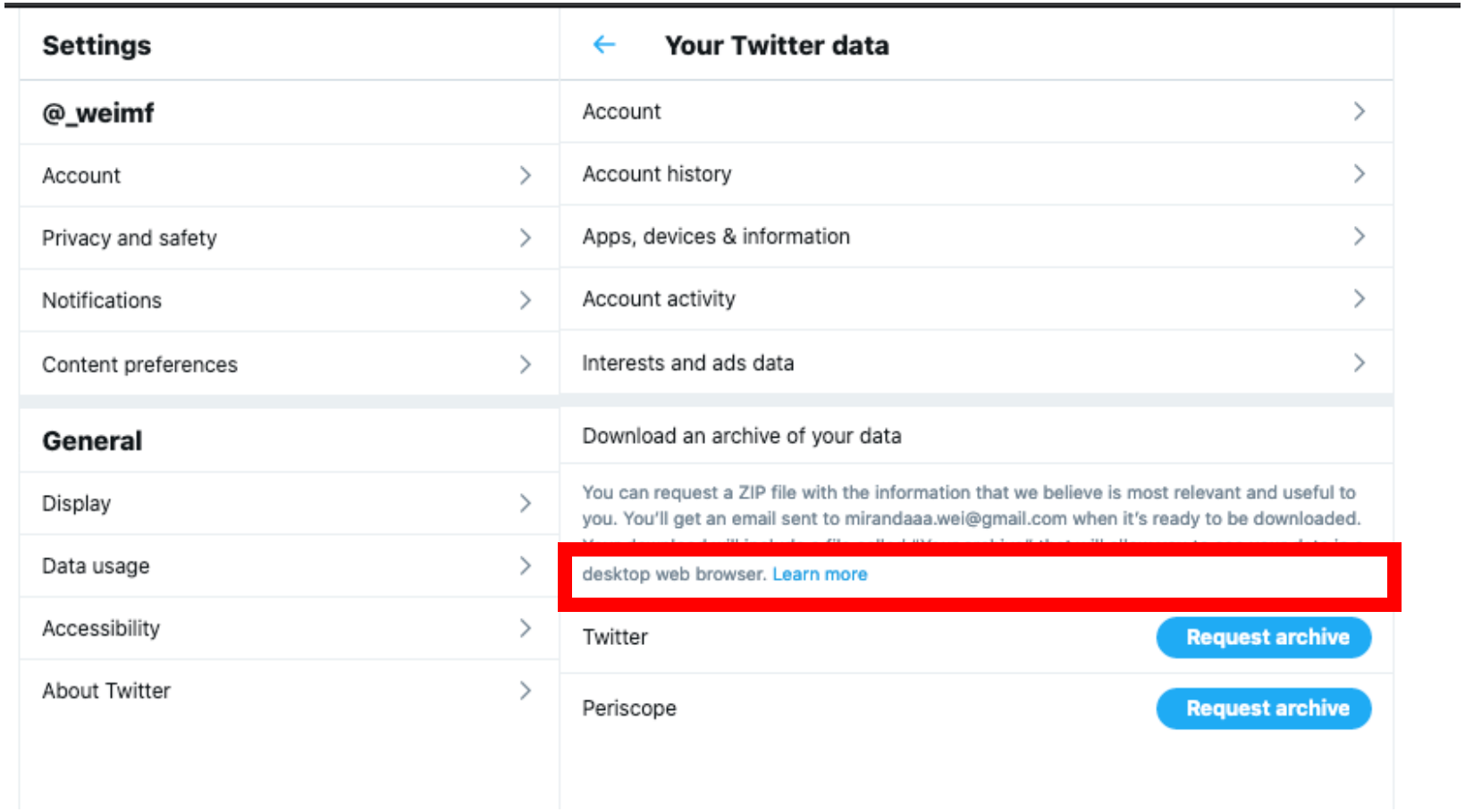
Blase Ur

gender: M
location: Chicago, IL
interests: bicycles, succulents, punk rock

Reality: “Organic Ketchup Buyer” + More

```
targetingType: Tailored audiences (lists)

targetingValue: NCS_PD_04358_Kraft_Organic and natural ketchup buyers_1_26362226

…

targetingType: Keywords

targetingValue: #parenting

…
```

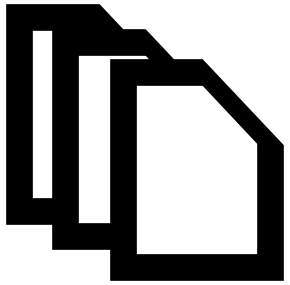
Methods Enabled by Subject Data Access Rights

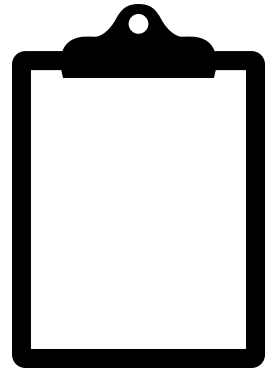


Contents of ad-impressions.js


```
[ {
  "ad" : {
    "adsUserData" : {
      "adImpressions" : {
        "impressions" : [ {
          "deviceInfo" : {
            "osType" : "Ios",
            "deviceId" : "#############",
            "deviceType" : "iPhone X"
          },
          "displayLocation" : "SearchTweets",
          "promotedTweetInfo" : {
            "tweetId" : "###################",
            "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



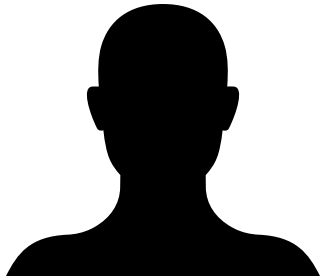


request Twitter data

upload ad-related data

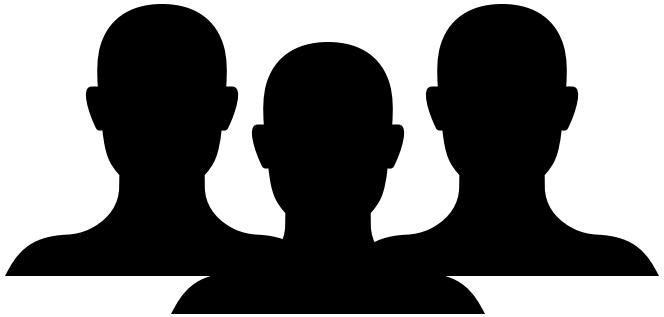
take customized survey

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

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

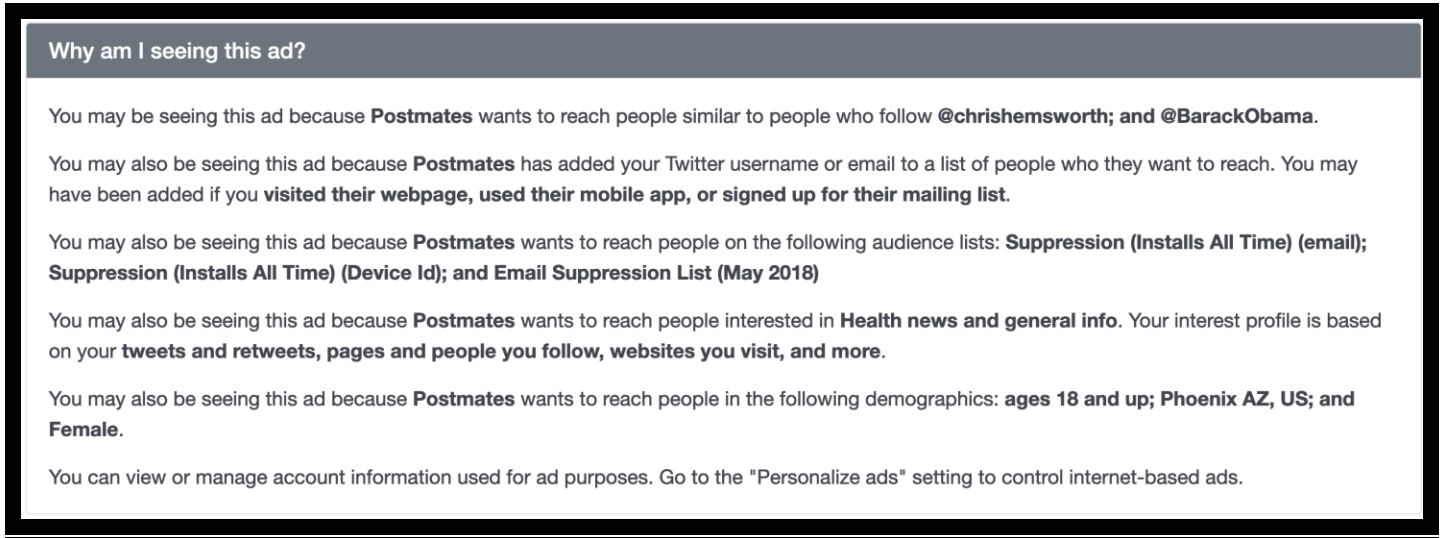
Postmates
Learn more about **Postmates**

- 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**

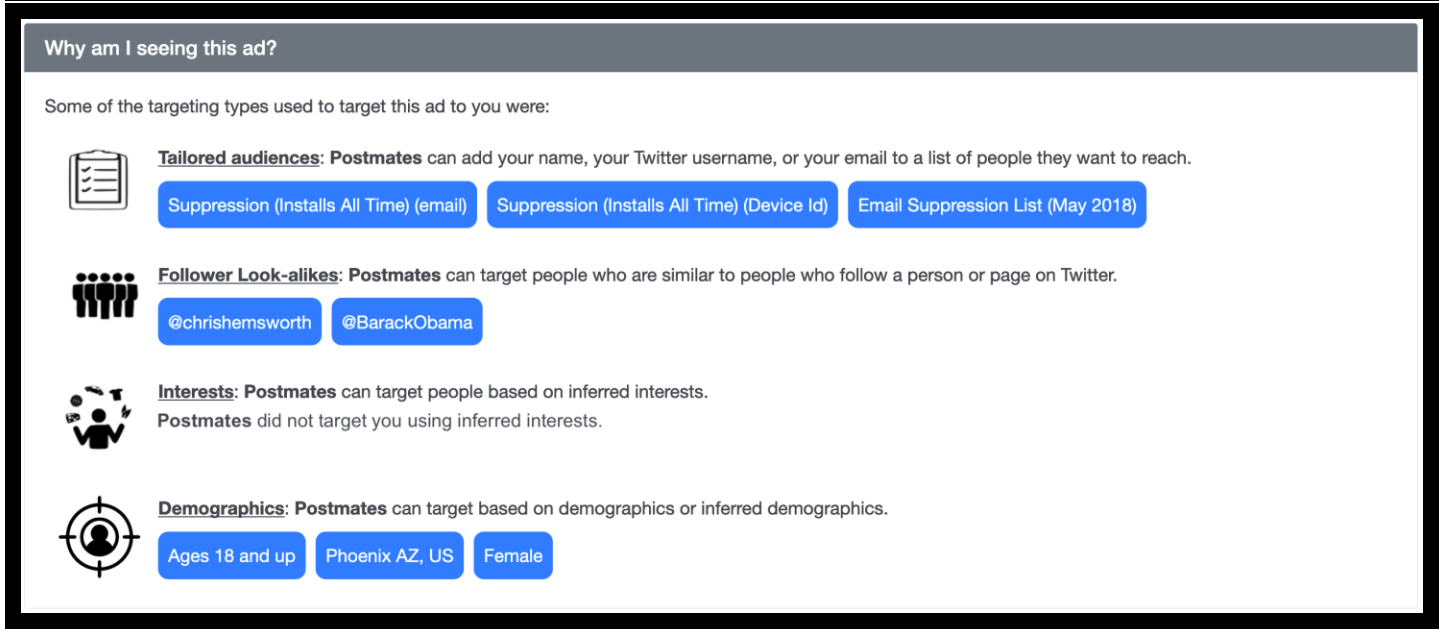
There may be other reasons you're seeing this ad, including that **Postmates** wants to reach **people who are ages 18 and up, are located in Phoenix AZ, US, and are Female**. This information is based on your Twitter profile and where you've connected to the internet.

Speculative Ad Explanations

detailed text



detailed visual



Speculative Ad Explanations

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**.

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.

control

Why am I seeing this ad?

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.

Speculative Ad Explanations: Results

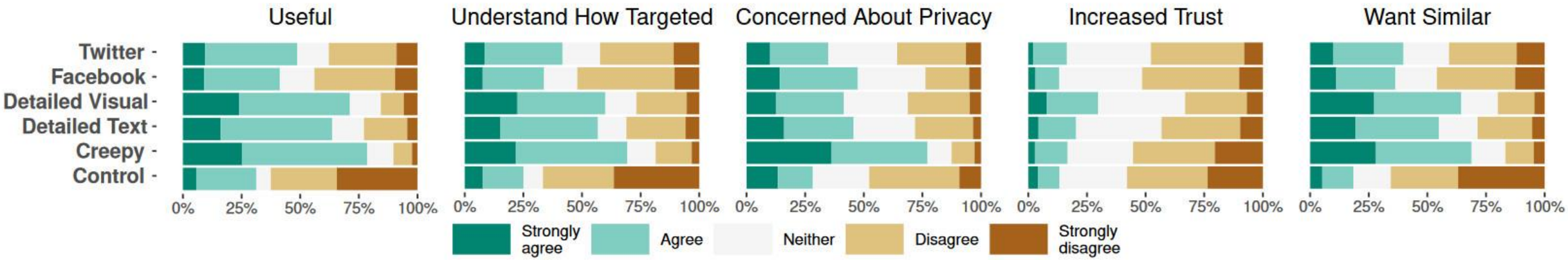
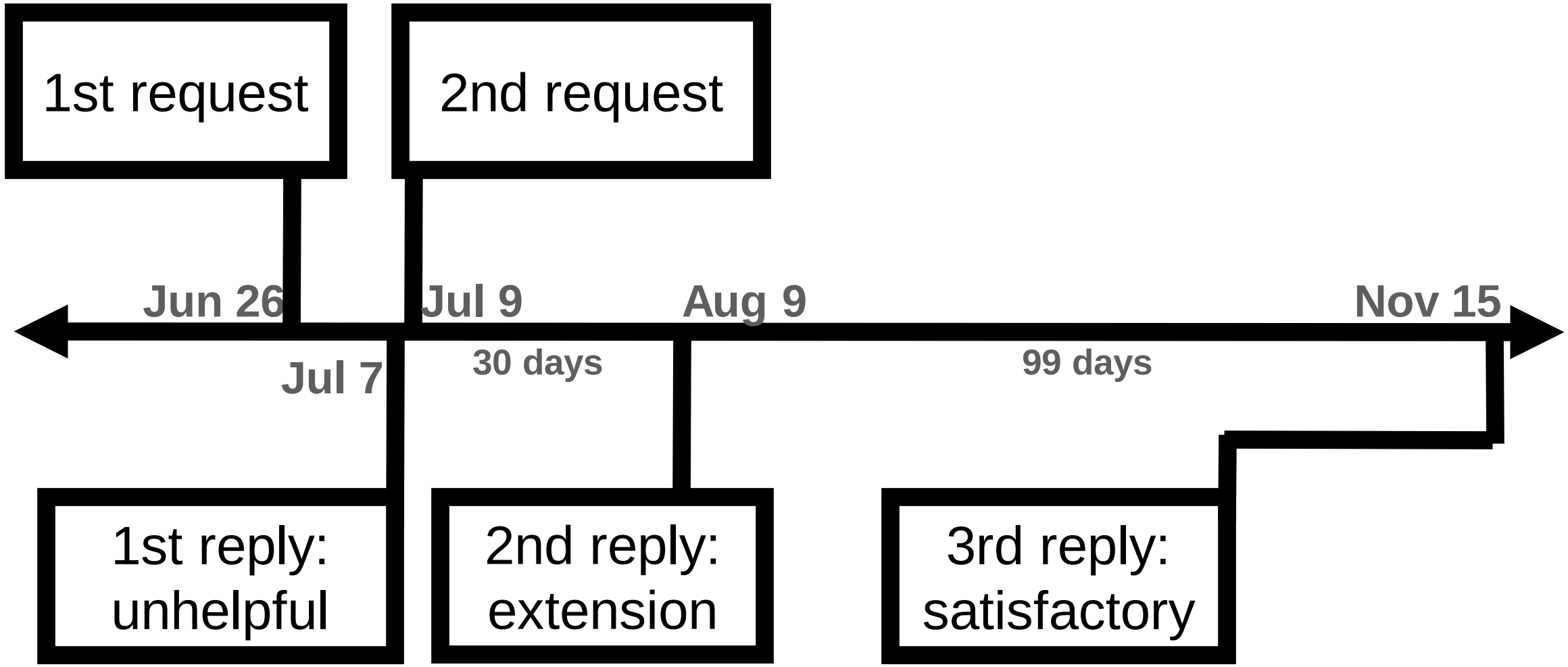


Figure 5: Participants' level of agreement to questions about ad explanations.

Clarifications From Twitter



The Politics of Transparency

The Politics of Transparency

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 our public database. (See documentation here.) 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. You can view 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.

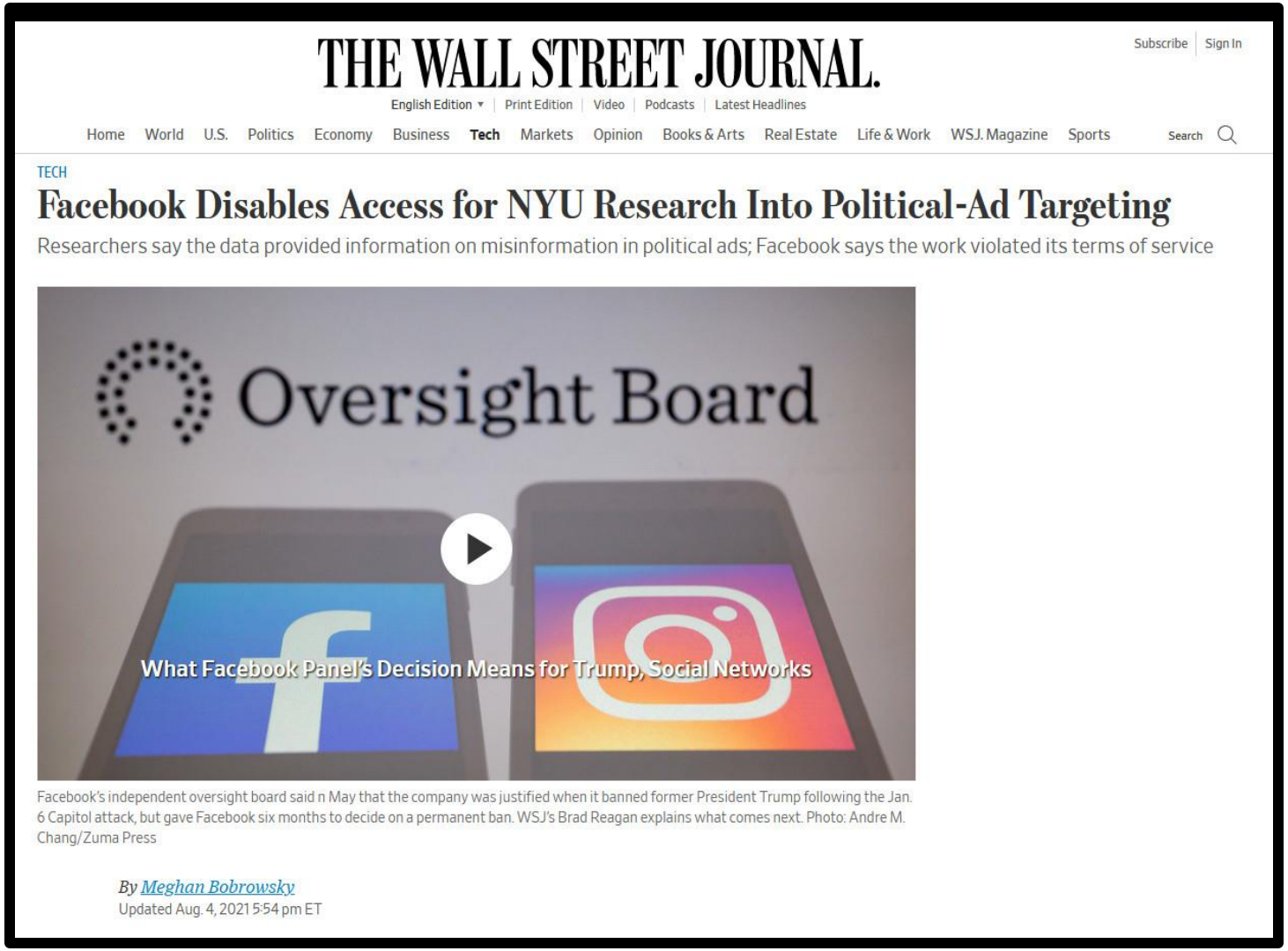
What we don't collect

Anything personally identifying, including:

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

https://adobservatory.org/

The Politics of Transparency



THE WALL STREET JOURNAL.

Subscribe | Sign In

English Edition | Print Edition | Video | Podcasts | Latest Headlines

Home World U.S. Politics Economy Business **Tech** Markets Opinion Books & Arts Real Estate Life & Work WSJ. Magazine Sports Search

TECH

Facebook Disables Access for NYU Research Into Political-Ad Targeting

Researchers say the data provided information on misinformation in political ads; Facebook says the work violated its terms of service

Facebook's independent oversight board said n May that the company was justified when it banned former President Trump following the Jan. 6 Capitol attack, but gave Facebook six months to decide on a permanent ban. WSJ's Brad Reagan explains what comes next. Photo: Andre M. Chang/Zuma Press

By Meghan Bobrowsky
Updated Aug. 4, 2021 5:54 pm ET

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