

# Responsible Data Science

Transparency: Online Ad Delivery

---

**Prof. Julia Stoyanovich & George Wood**

Center for Data Science  
New York University

# Transparency themes

- ▶ Explaining black-box models
  - ▶ LIME: local interpretable explanations [Ribeiro et al., KDD 2016]
  - ▶ QII: causal inference of features on outcomes [Datta et al., SPP 2016]
  - ▶ SHAP: Shapley additive explanations [Lundberg and Lee, NeurIPS 2016]
- ▶ **Online ad targeting**
  - ▶ Racially identifying names [Sweeney, CACM 2013]
  - ▶ AdFisher [Datta et al., PETS 2015]
  - ▶ Discrimination through optimization [Ali et al., CSCW 2019]
- ▶ Interpretability

# Online ad targeting

1. Racially associated names
2. AdFisher
3. Discrimination through optimization

# Gender discrimination in online job ads

## Discrimination in Online Ad Delivery

[L. Sweeney; *CACM 2013*]

# Gender discrimination in online job ads

5

## Discrimination in Online Ad Delivery

[Latanya Sweeney](#)

A Google search for a person's name, such as "Trevon Jones", may yield a personalized ad for public records about Trevon that may be neutral, such as "Looking for Trevon Jones?", or may be suggestive of an arrest record, such as "Trevon Jones, Arrested?". This writing investigates the delivery of these kinds of ads by Google AdSense using a sample of racially associated names and finds statistically significant discrimination in ad delivery based on searches of 2184 racially associated personal names across two websites. First names, assigned at birth to more black or white babies, are found predictive of race (88% black, 96% white), and those assigned primarily to black babies, such as DeShawn, Darnell and Jermaine, generated ads suggestive of an arrest in 81 to 86 percent of name searches on one website and 92 to 95 percent on the other, while those assigned at birth primarily to whites, such as Geoffrey, Jill and Emma, generated more neutral copy: the word "arrest" appeared in 23 to 29 percent of name searches on one site and 0 to 60 percent on the other. On the more ad trafficked website, a black-identifying name was 25% more likely to get an ad suggestive of an arrest record. A few names did not follow these patterns. All ads return results for actual individuals and ads appear regardless of whether the name has an arrest record in the company's database. The company maintains Google received the same ad text for groups of last names (not first names), raising questions as to whether Google's technology exposes racial bias.

# Racially associated names

[Latanya Sweeney; CACM 2013]



**Ads by Google**

[Latanya Sweeney, Arrested?](#)  
1) Enter Name and State. 2) Access F  
Checks Instantly.  
[www.instantcheckmate.com/](http://www.instantcheckmate.com/)

[Latanya Sweeney](#)  
Public Records Found For: Latanya S  
[www.publicrecords.com/](http://www.publicrecords.com/)

[La Tanya](#)

**INSTANT checkmate™**

**LATANYA SWEENEY**  
1420 Centre Ave  
Pittsburgh, PA 15219  
DOB: Oct 27, 1959 (53 years old)

**CERTIFIED**

Personal	Criminal History		
Name, aliases, birthdate, phone numbers, etc.	Rate This Content:		
Location Detailed address history and related data, maps, etc.	This section contains possible citation, arrest, and criminal records for the subject of this report. While our database does contain hundreds of millions of arrest records, different counties have different rules regarding what information they will and will not release.		
We share with you as much information as we possibly can, but a clean slate here should not be interpreted as a guarantee that Latanya Sweeney has never been arrested; it simply means that we were not able to locate any matching arrest records in the data that is available to us.			
<b>Possible Matching Arrest Records</b>			
Name	County and State	Offenses	View Details
No matching arrest records were found.			

## Racism is Poisoning Online Ad Delivery, Says Harvard Professor

Google searches involving black-sounding names are more likely to serve up ads suggestive of a criminal record than white-sounding names, says computer scientist

**racially identifying names trigger ads suggestive of a criminal record**

<https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/>

# Racially associated names

[Latanya Sweeney; CACM 2013]

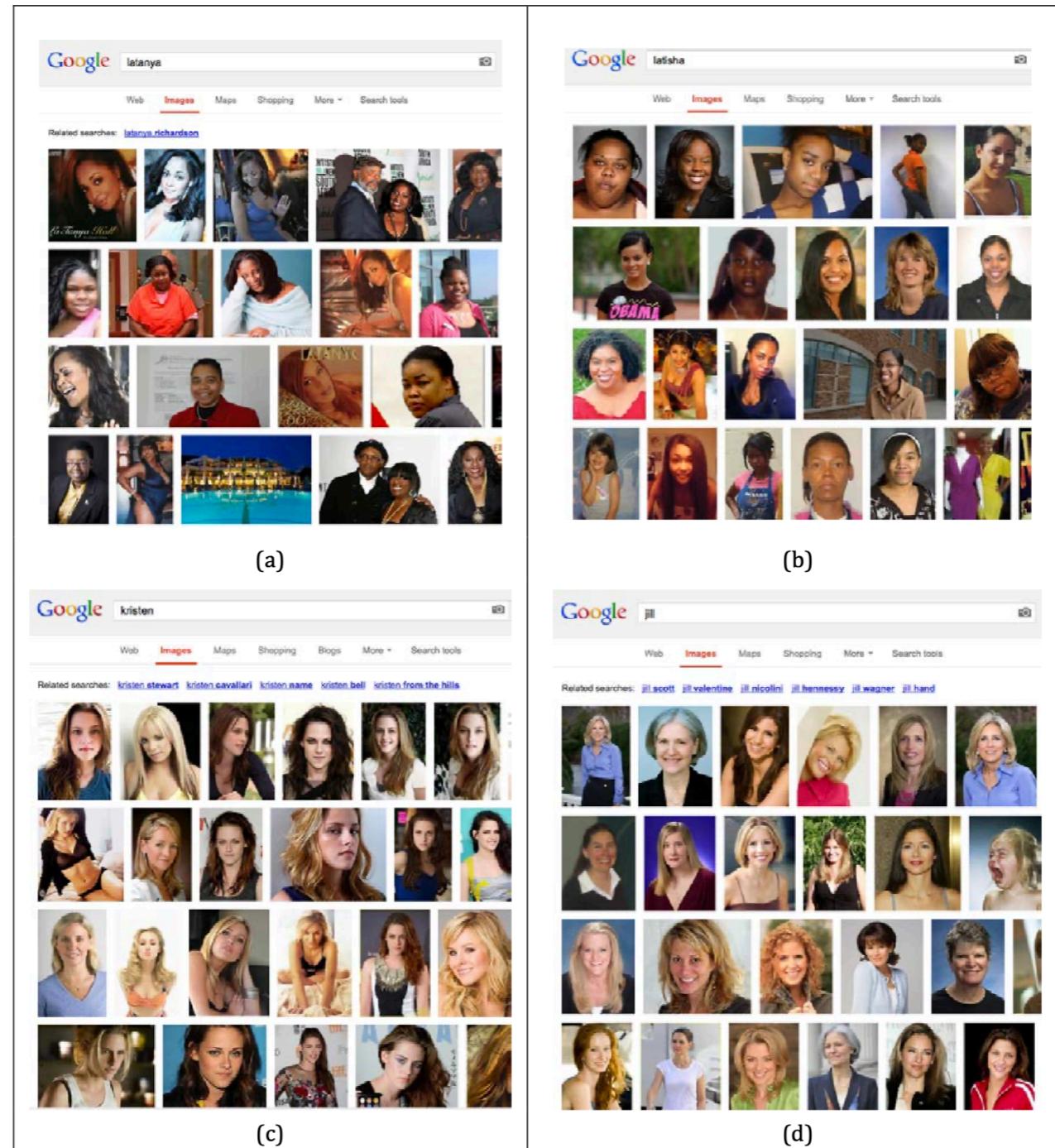


Figure 4. Sample face images on google.com retrieved for searches “latanya” (a), “latisha” (b), “kristen” (c), and “jill” (d).

# Racially associated names

[Latanya Sweeney; CACM 2013]

- Ads suggestive of criminal record, linking to Instant Checkmate, appear on google.com and reuters.com in response to searches “Latanya Sweeney”, “Latanya Farrell”, and “Latanya Locket”\*
- No Instant Checkmate ads when searching for “Kristen Haring”, “Kristen Sparrow”\*, and “Kristen Lindquist”\*
- \* Name associated with an actual arrest record

# Racially associated names: details

[Latanya Sweeney; CACM 2013]

On Reuters.com, which hosts Google AdSense ads, **a black-identifying name was 25% more likely to generate an ad suggestive of an arrest record.**

More than 1,100 Instant Checkmate ads appeared on Reuters.com, with 488 having black-identifying first names; of these, 60% used arrest in the ad text. Of the 638 ads displayed with white-identifying names, 48% used arrest. This difference is statistically significant,  $\chi^2(1) = 14.32, p < 0.001$ .

**The EEOC's and U.S. Department of Labor's adverse impact test for measuring discrimination is 77 in this case, so if this were an employment situation, a charge of discrimination might result.** (The adverse impact test uses the ratio of neutral ads, or 100 minus the percentages given, to compute disparity:  $100-60=40$  and  $100-48=52$ ; dividing 40 by 52 equals 77.)

# Why is this happening?

[Latanya Sweeney; *CACM 2013*]

Possible explanations (from Latanya Sweeney):

- ▶ Does Instant Checkmate serve ads specifically for black-identifying names?
- ▶ Is Google AdSense explicitly biased in this way?
- ▶ Does Google AdSense learn racial bias from click-through rates?

# Gender discrimination in online job ads

<https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/>

In response to this blog post, a **Google** spokesperson sends the following comment:

**“AdWords does not conduct any racial profiling.** We also have an “anti” and violence policy which states that we will not allow ads that advocate against an organisation, person or group of people. It is up to individual advertisers to decide which keywords they want to choose to trigger their ads.”

**Instantcheckmate.com** sends the following statement:

“As a point of fact, Instant Checkmate would like to state unequivocally that it has never engaged in racial profiling in Google AdWords. **We have absolutely no technology in place to even connect a name with a race** and have never made any attempt to do so. The very idea is contrary to our company’s most deeply held principles and values.”

# Who is responsible?

- ▶ Who benefits?
  - ▶ Google earned approximately \$40 billion in U.S. advertising revenue in 2020
- ▶ Who is harmed?
- ▶ What does the law say?
- ▶ Who is in a position to mitigate harms?

Transparency, responsibility, **trust**

# The socio-legal landscape

L. Sweeney, <https://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery>

- The U.S. Equal Employment Opportunity Commission (EEOC) is the federal agency charged with enforcing Title VII of the Civil Rights Act of 1964, a law that applies to most employers, prohibiting employment discrimination based on race, color, religion, sex, or national origin, and extended to those having criminal records
- Title VII does not prohibit employers from obtaining criminal background information, but a blanket policy of excluding applicants based solely upon information indicating an arrest record can result in a charge of discrimination
- What about online ads suggesting someone with your name has an arrest record? **Title VII only applies if you have an arrest record and can provide the employer inappropriately used the ads**

# The socio-legal landscape

L. Sweeney, <https://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery>

- Are the ads commercial free speech?
- The First Amendment of the U.S. Constitution protects advertising, but the U.S. Supreme Court set out a test for assessing restrictions on commercial speech, **which begins by determining whether the speech is misleading.**
- Are online ads suggesting the existence of an arrest record misleading if no one by that name has an arrest record?
- Assume the ads are free speech: what happens when these ads appear more often for one racial group than another? Not everyone is being equally affected by free speech. Is that free speech of racial discrimination?

## Automated Experiments on Ad Privacy Settings (AdFisher)

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

## Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination

[Amit Datta](#), [Michael Carl Tschantz](#), [Anupam Datta](#)

To partly address people's concerns over web tracking, Google has created the Ad Settings webpage to provide information about and some choice over the profiles Google creates on users. We present AdFisher, an automated tool that explores how user behaviors, Google's ads, and Ad Settings interact. AdFisher can run browser-based experiments and analyze data using machine learning and significance tests. Our tool uses a rigorous experimental design and statistical analysis to ensure the statistical soundness of our results. We use AdFisher to find that the Ad Settings was opaque about some features of a user's profile, that it does provide some choice on ads, and that these choices can lead to seemingly discriminatory ads. In particular, we found that visiting webpages associated with substance abuse changed the ads shown but not the settings page. We also found that setting the gender to female resulted in getting fewer instances of an ad related to high paying jobs than setting it to male. We cannot determine who caused these findings due to our limited visibility into the ad ecosystem, which includes Google, advertisers, websites, and users. Nevertheless, these results can form the starting point for deeper investigations by either the companies themselves or by regulatory bodies.

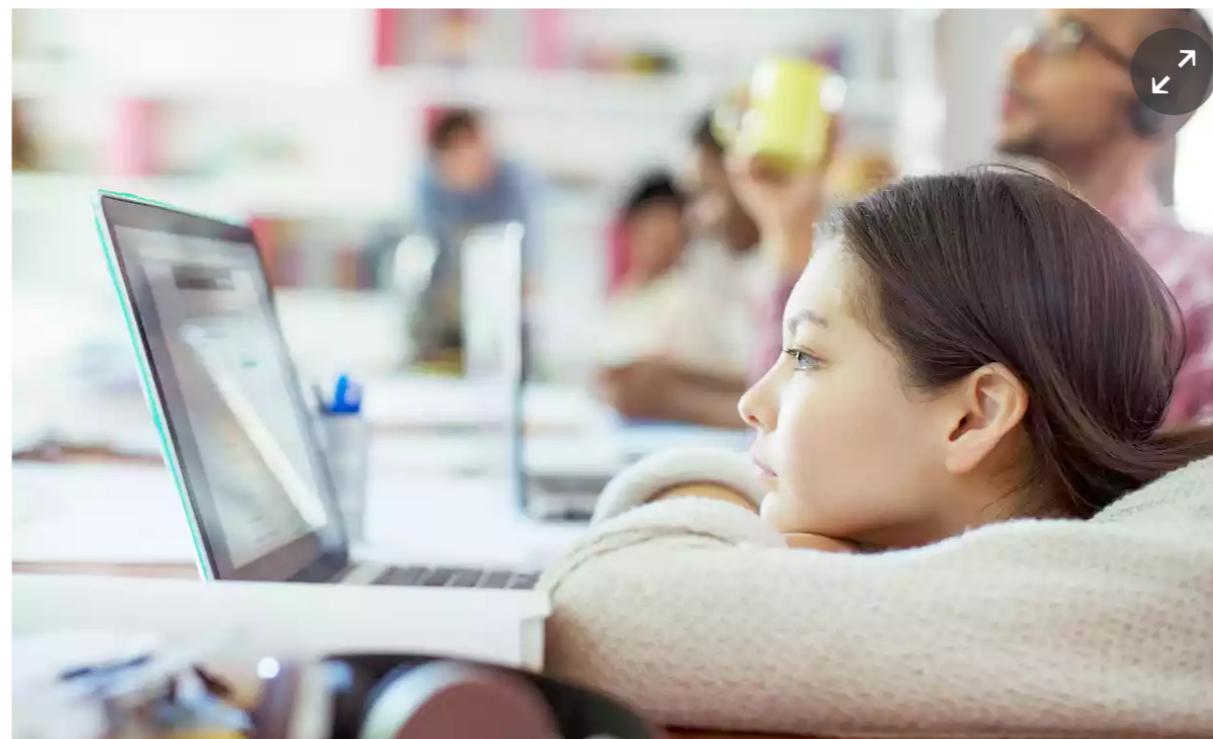


Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

## Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



i One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

<https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>

# Online job ads

18

- The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences, or demographic characteristics, except in gender
- One experiment showed that Google displayed ads for a career coaching service for “\$200k+” executive jobs **1,852 times to the male group and only 318 times to the female group**
- Another experiment, in July 2014, showed a similar discrepancy but was not statistically significant

# Ad targeting online

- **Users** browse the Web, consume content, consume ads (i.e. view, click, purchase)
- **Content providers** (or **publishers**) host online content that often includes ads. They outside ad placement to third-party ad networks
- **Advertisers** seek to place their ads on publishers' website
- **Ad networks** track users across sites to get a global view of users' behaviors. They connect advertisers and publishers

# Google ad settings

- ▶ Google ad settings aim to provide transparency and give control to users over the ads that they see

Your Google profile

The screenshot shows the Google Ad Settings interface. At the top, there are two circular icons: one for 'Gender' with a blue profile icon and one for 'Age' with the text '35-44'. Below this, a section titled 'Ads based on your interests' has a toggle switch set to 'ON'. A sub-section titled 'Improve your ad experience when you are signed in to Google sites' compares 'With Ads based on your interests ON' (green background) and 'With Ads based on your interests OFF' (grey background). The 'ON' section lists benefits like ads being delivered based on search queries and YouTube history, while the 'OFF' section lists consequences like ads being less relevant and interests being deleted.

With Ads based on your interests ON	With Ads based on your interests OFF
<ul style="list-style-type: none"><li>The ads you see will be delivered based on your prior search queries, the videos you've watched on YouTube, as well as other information associated with your account, such as your age range or gender</li><li>On some Google sites like YouTube, you will see ads related to your interests, which you can edit at any time by visiting this page</li><li>You can block some ads that you don't want to see</li></ul>	<ul style="list-style-type: none"><li>You will still see ads and they may be based on your general location (such as city or state)</li><li>Ads will not be based on data Google has associated with your Google Account, and so may be less relevant</li><li>You will no longer be able to edit your interests</li><li>All the advertising interests associated with your Google Account will be deleted</li></ul>

<http://www.google.com/settings/ads>

# Google ad settings

- ▶ Do users truly have transparency & choice, or is this a placebo button?

**Google**

Julia

### Control your Google ads

You can control the ads that are delivered to you based on your Google Account, across devices, by editing these settings. These ads are more likely to be useful and relevant to you.

Your interests

<input checked="" type="checkbox"/> Action & Adventure Films	<input checked="" type="checkbox"/> Cats
<input checked="" type="checkbox"/> Cooking & Recipes	<input checked="" type="checkbox"/> Fitness
<input checked="" type="checkbox"/> History	<input checked="" type="checkbox"/> Hybrid & Alternative Vehicles
<input checked="" type="checkbox"/> Hygiene & Toiletries	<input checked="" type="checkbox"/> Make-Up & Cosmetics
<input checked="" type="checkbox"/> Mobile Phones	<input checked="" type="checkbox"/> Parenting
<input checked="" type="checkbox"/> Phone Service Providers	<input checked="" type="checkbox"/> Recording Industry
<input checked="" type="checkbox"/> Reggaeton	<input checked="" type="checkbox"/> Search Engine Optimization & Marketing
<input checked="" type="checkbox"/> Vehicle Brands	

[+ ADD NEW INTEREST](#)
[WHERE DID THESE COME FROM?](#)

These interests are derived from your activity on Google sites, such as the videos you've watched on YouTube. This does not include Gmail interests, which are used only for ads within Gmail. [Learn more](#)

<http://www.google.com/settings/ads>



thirdside.co

# AdFisher

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- ▶ How do user behaviors, ads, and ad settings interact?
- ▶ Individual data use transparency: ad network must share the information it uses about the user to select which ads to serve

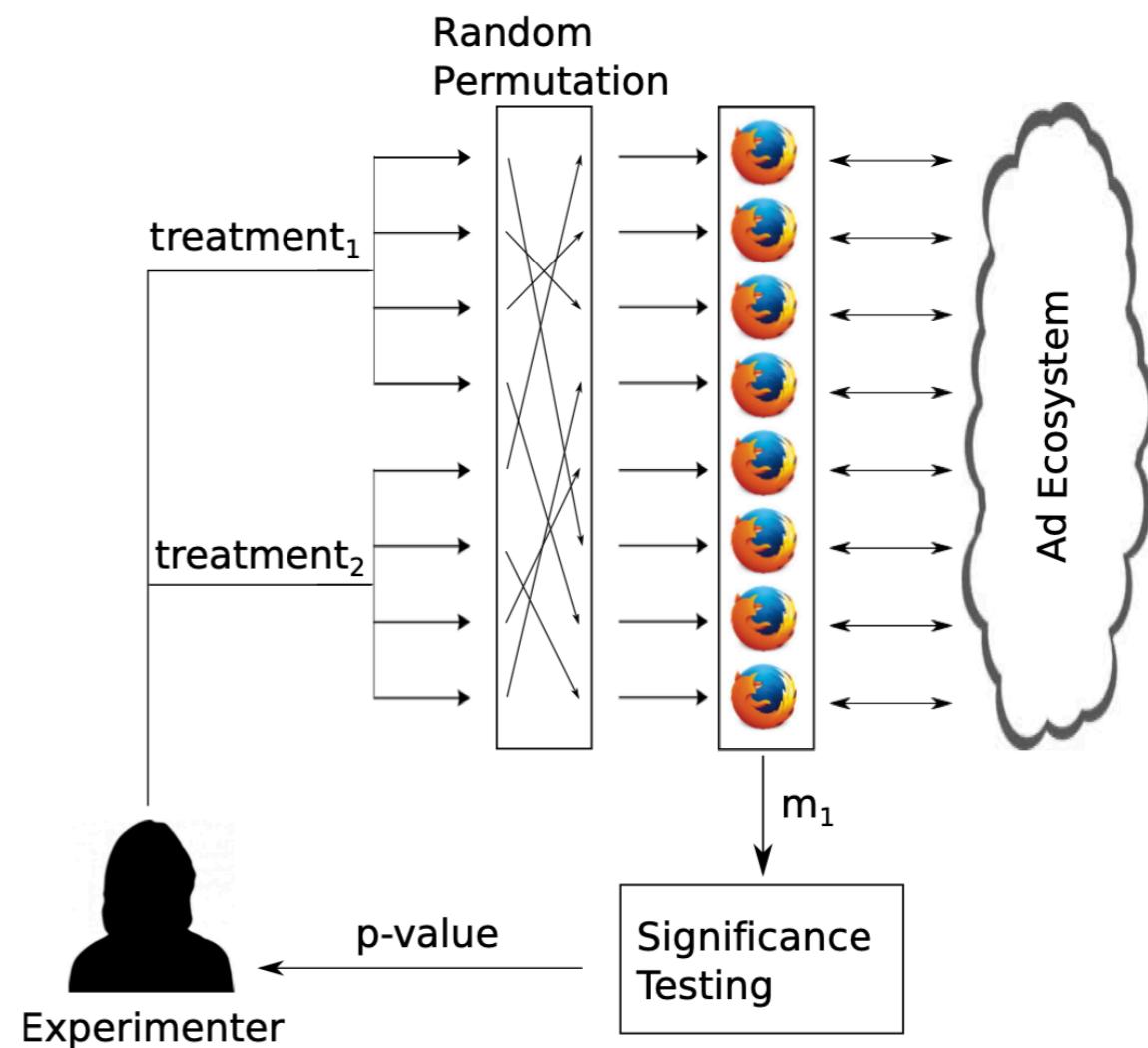


Figure 2: Experimental setup to carry out significance testing on eight browser agents comparing the effects of two treatments. Each agent is randomly assigned a treatment which specifies what actions to perform on the web. After these actions are complete, they collect measurements which are used for significance testing.

# AdFisher

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- ▶ Browser-based experiments, simulated users:

- ▶ Input:

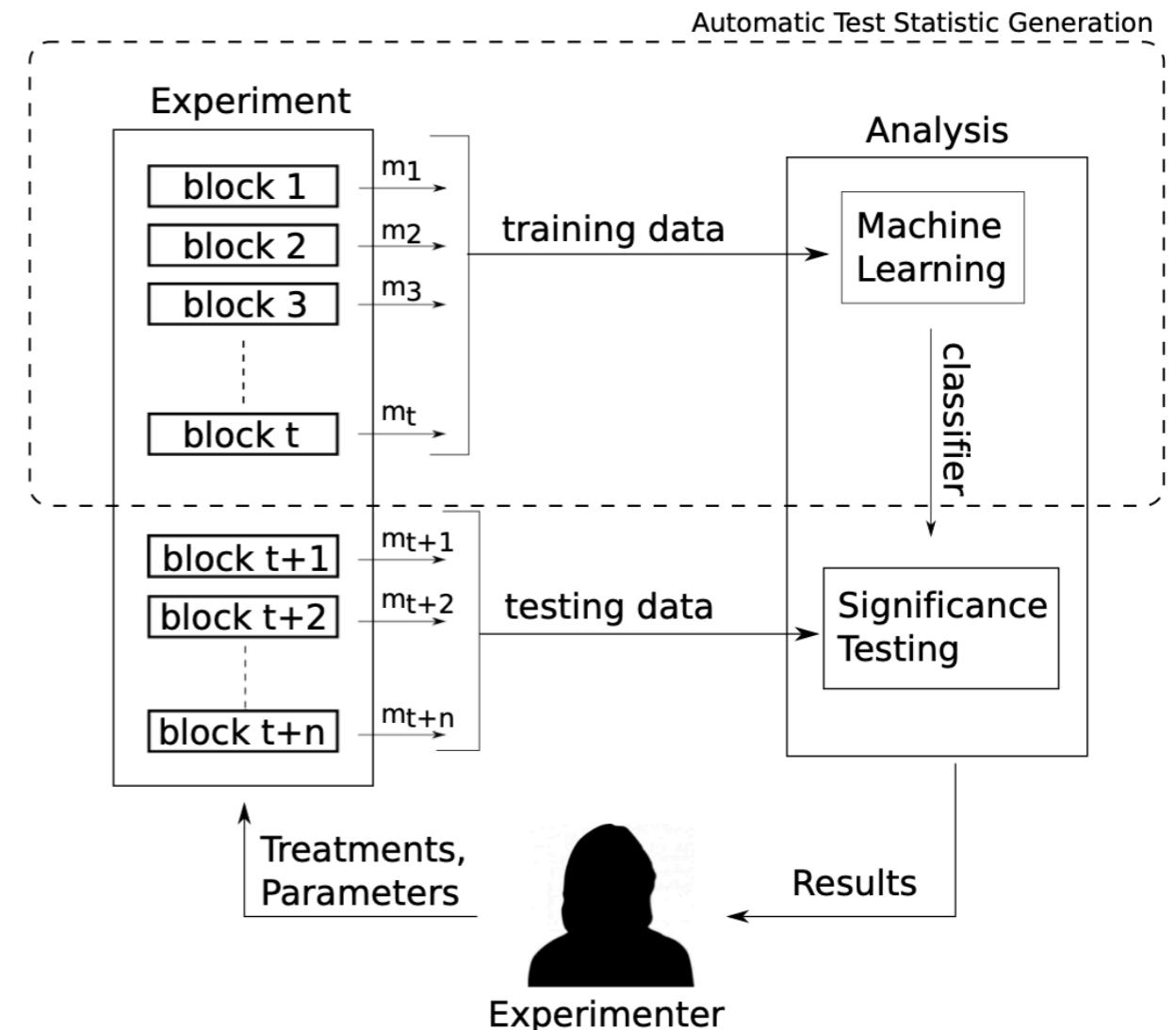
1. visits to content providing websites;
2. interactions with Google Ad Settings

- ▶ Output:

1. ads shown to users by Google;
2. change in Google Ad Settings

- ▶ Experimental design:

- ▶ Null hypothesis: inputs do not affect outputs
- ▶ Control and treatment conditions
- ▶ AdFisher can help select a test statistic



# AdFisher: gender and jobs

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- ▶ **Non-Discrimination:** Users differing only in protected attributes are treated similarly.
- ▶ **Causal test:** Does a protected attribute change ads?
- ▶ Experiment 1: **Gender and jobs** –
  - ▶ Specify gender (male/female) in Ad Settings, simulate interest in jobs by visiting employment sites, collect ads from Times of India or the Guardian
  - ▶ Result: In one experiment, males were shown ads for higher-paying jobs far more often than females

# AdFisher: gender and jobs

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Title	URL	Coefficient	appears in agents		total appearances	
			female	male	female	male
Top ads for identifying the simulated male group						
Truck Driving Jobs	<a href="http://www.bestpayingdriverjobs.com">www.bestpayingdriverjobs.com</a>	0.492	0	15	0	33
\$200k+ Jobs - Execs Only	<a href="http://careerchange.com">careerchange.com</a>	0.383	0	15	0	48
Aircraft Technician Program	<a href="http://pia.edu">pia.edu</a>	0.292	0	6	0	14
Paid CDL Training	<a href="http://pamtransport.greatcdltraining.com">pamtransport.greatcdltraining.com</a>	0.235	0	5	0	13
Unique Bridal Necklaces	<a href="http://margaretelizabeth.com/Bridal">margaretelizabeth.com/Bridal</a>	0.234	0	5	0	19
Top ads for identifying agents in the simulated female group						
Business Loans for Women	<a href="http://topsbaloans.com">topsbaloans.com</a>	-0.334	13	1	70	1
Post Your Classified Ad	<a href="http://indeed.com/Post-Jobs">indeed.com/Post-Jobs</a>	-0.267	20	16	56	24
American Laser Skincare	<a href="http://americanlaser.com">americanlaser.com</a>	-0.243	8	5	14	8
Dedicated Drivers Needed	<a href="http://warrentransport.com">warrentransport.com</a>	-0.224	3	0	14	0
Earn Your Nursing Degree	<a href="http://nursing-colleges.courseadvisor.com">nursing-colleges.courseadvisor.com</a>	-0.219	11	3	31	10

Table 5: Top URL+titles for the gender and jobs experiment (July).

# AdFisher: substance abuse

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- ▶ **Transparency:** User can view data used for ad selection
- ▶ **Causal test:** Find attribute that changes ads but not settings
- ▶ Experiment 2: **Substance abuse** —
  - ▶ Simulate interest in substance abuse in the experimental group but not in the control group, check for differences in Ad Settings, collect ads from Times of India
  - ▶ Result: No difference in Ad Settings between the groups, yet significant differences in ads served: rehab vs. stocks & driving jobs

# AdFisher: online dating

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- ▶ **Ad choice:** Removing an interest decreases the number of ads related to that interest
- ▶ **Causal test:** Find that removing an interest causes a decrease in related ads
- ▶ Experiment 3: **Online dating** –
  - ▶ Simulate interest in online dating in both groups, remove “Dating & Personals” from the interests on Ad Settings for experimental group, collect ads
  - ▶ Result: Members of experimental group do not get ads related to dating, while members of the control group do

# Recall the setup

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT\* 2018*]

- ▶ **Users** browse the Web, consume content, consume ads (see, click, purchase)
- ▶ **Content providers** (or **publishers**) host online content that often includes ads. They outsource ad placement to third-party ad networks
- ▶ **Advertisers** seek to place their ads on publishers' websites
- ▶ **Ad networks** track users across sites to get a global view of users' behaviors. They connect advertisers and publishers

**Why are males seeing ads for high-paying jobs more often?**

**What is causing gender based discrimination?**

**(1) Who is responsible and (2) how is discrimination enacted?**

# Who is responsible?

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT\* 2018*]

- ▶ **Google alone:** explicitly programming the system to show the ad less often to females, e.g. based on independent evaluation of demographic appeal of product (**explicit and intentional discrimination**)
- ▶ **The advertiser:** targeting the ad through explicit use of demographic categories (**explicit and intentional**), selection of proxies (**hidden and intentional**), or through those choices without intent (**unconscious selection bias**), and **Google** respecting these targeting criteria
- ▶ **Online advertisers:** others outbid our advertiser when targeting to females
- ▶ **Other users:** Male and female users behaving differently with respect to ads, and Google learning to predict this behavior

# How is targeting done?

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT\* 2018*]

- ▶ On gender directly
- ▶ On a proxy of gender, i.e., on a known correlate of gender because it is a correlate
- ▶ On a known correlate of gender, but not because it is a correlate
- ▶ On an unknown correlate of gender

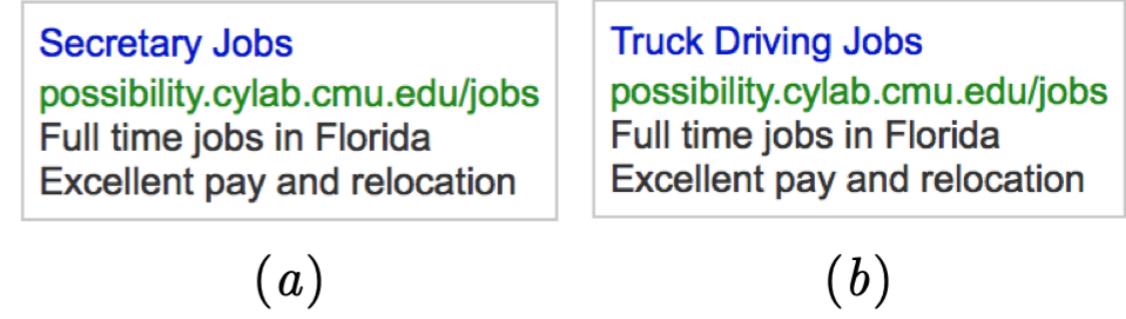


Figure 1: Ads approved by Google in 2015. The ad in the left (right) column was targeted to women (men).

*“This finding demonstrates that an advertiser with discriminatory intentions can use the AdWords platform to serve employment related ads disparately on gender.”*

# What are the legal ramifications?

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT\* 2018*]

- ▶ Each actor in the advertising ecosystem may have contributed inputs that produced the effect
- ▶ It is impossible to know, without additional information, what the different actors — other than the consumers of the ads — did or did not do
- ▶ In particular, impossible to assess intent, which may be necessary to assess the extent of legal liability. Or it may not!
  - ▶ Title VII of the 1964 Civil Rights Act makes it unlawful to discriminate based on sex in several stages of employment. It includes an advertising prohibition (e.g. sex-specific help wanted columns in a newspaper), which does not hinge on intent
  - ▶ Title VII does not directly apply here because it is limited in scope to employers, labour organizations, employment agencies, joint labor-management committees
  - ▶ Fair Housing Act (FHA) is perhaps a better guide than Title VII, limiting both content and activities that target advertisement based on protected attributes

# Facebook ads and the Fair Housing Act

THEVERGE

## Facebook has been charged with housing discrimination by the US government

*'Facebook is discriminating against people based upon who they are and where they live,' says HUD secretary*

By Russell Brandom | Mar 28, 2019, 7:51am EDT

The Department of Housing and Urban Development has [filed charges](#) against Facebook for housing discrimination, escalating the company's ongoing fight over discrimination in its ad targeting system. The charges build on [a complaint filed in August](#), finding that there is reasonable cause to believe Facebook has served ads that violate the Fair Housing Act.

*ProPublica* first raised concerns over housing discrimination on Facebook in 2016, when reporters found that [the “ethnic affinities” tool](#) could be used to exclude black or Hispanic users from seeing specific ads. If those ads were for housing or employment opportunities, the targeting could easily violate federal law. At the time, Facebook had no internal safeguards in place to prevent such targeting.

# Facebook ads and the Fair Housing Act

THEVERGE

## Facebook has been charged with housing discrimination by the US government

*'Facebook is discriminating against people based upon who they are and where they live,' says HUD secretary*

Facebook has struggled to effectively address the possibility of discriminatory ad targeting. The company pledged to step up anti-discrimination enforcement in the wake of ProPublica's reporting, but [a follow-up report](#) in 2017 found the same problems persisted nearly a year later.

According to the HUD complaint, many of the options for targeting or excluding audiences are shockingly direct, including a map tool that explicitly echoes [redlining practices](#).

"[Facebook] has provided a [toggle button that enables advertisers to exclude men or women](#) from seeing an ad, a search-box to exclude people who do not speak a specific language from seeing an ad, and [a map tool to exclude people who live in a specified area](#) from seeing an ad by drawing a red line around that area," the complaint reads.

**"WE'RE DISAPPOINTED BY TODAY'S DEVELOPMENTS," FACEBOOK SAYS**

# Ads and the Fair Housing Act

THEVERGE

## Facebook has been charged with housing discrimination by the US government

*'Facebook is discriminating against people based upon who they are and where they live,' says HUD secretary*

By Russell Brandom | Mar 28, 2019, 7:51am EDT

This is the first federal discrimination lawsuit to deal with racial bias in targeted advertising, a milestone that lawyers at HUD said was overdue. "Even as we confront new technologies, the fair housing laws enacted over half a century ago remain clear—discrimination in housing-related advertising is against the law," said HUD General Counsel Paul Compton. "Just because a process to deliver advertising is opaque and complex doesn't mean that it's exempts Facebook and others from our scrutiny and the law of the land."

## HUD reportedly also investigating Google and Twitter in housing discrimination probe

By Adi Robertson | @thedextriarchy | Mar 28, 2019, 3:52pm EDT

<https://www.theverge.com/2019/3/28/18285899/housing-urban-development-hud-facebook-lawsuit-google-twitter>

# HUD v. Facebook



[https://www.hud.gov/press/press\\_releases\\_media\\_advisories/HUD\\_No\\_19\\_035](https://www.hud.gov/press/press_releases_media_advisories/HUD_No_19_035)

"Facebook is discriminating against people based upon who they are and where they live," said HUD Secretary Ben Carson. "Using a computer to limit a person's housing choices can be just as discriminatory as slamming a door in someone's face."

According to HUD's Charge, **Facebook enabled advertisers to exclude people whom Facebook classified as parents; non-American-born; non-Christian; interested in accessibility; interested in Hispanic culture;** or a wide variety of other interests that closely align with the Fair Housing Act's protected classes. HUD is also charging that Facebook **enabled advertisers to exclude people based upon their neighborhood by drawing a red line around those neighborhoods on a map.** Facebook also allegedly gave advertisers the option of showing ads only to men or only to women.

**The Charge further asserts that Facebook also uses the protected characteristics of people to determine who will view ads regardless of whether an advertiser wants to reach a broad or narrow audience.** HUD claims Facebook combines data it collects about user attributes and behavior with data it obtains about user behavior on other websites and in the non-digital world. Facebook then allegedly uses machine learning and other prediction techniques to classify and group users to project each user's likely response to a given ad, and in doing so, may recreate groupings defined by their protected class. **The Charge concludes that by grouping users who have similar attributes and behaviors (unrelated to housing) and presuming a shared interest or disinterest in housing-related advertisements, Facebook's mechanisms function just like an advertiser who intentionally targets or excludes users based on their protected class.**

## Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *CSCW 2019*]

# Discrimination in Facebook's ad delivery

## Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes

[Muhammad Ali](#), [Piotr Sapiezynski](#), [Miranda Bogen](#), [Aleksandra Korolova](#), [Alan Mislove](#),  
[Aaron Rieke](#)

The enormous financial success of online advertising platforms is partially due to the precise targeting features they offer. Although researchers and journalists have found many ways that advertisers can target---or exclude---particular groups of users seeing their ads, comparatively little attention has been paid to the implications of the platform's ad delivery process, comprised of the platform's choices about which users see which ads.

It has been hypothesized that this process can "skew" ad delivery in ways that the advertisers do not intend, making some users less likely than others to see particular ads based on their demographic characteristics. In this paper, we demonstrate that such skewed delivery occurs on Facebook, due to market and financial optimization effects as well as the platform's own predictions about the "relevance" of ads to different groups of users. We find that both the advertiser's budget and the content of the ad each significantly contribute to the skew of Facebook's ad delivery. Critically, we observe significant skew in delivery along gender and racial lines for "real" ads for employment and housing opportunities despite neutral targeting parameters.

Our results demonstrate previously unknown mechanisms that can lead to potentially discriminatory ad delivery, even when advertisers set their targeting parameters to be highly inclusive. This underscores the need for policymakers and platforms to carefully consider the role of the ad delivery optimization run by ad platforms themselves---and not just the targeting choices of advertisers---in preventing discrimination in digital advertising.

# Discrimination in Facebook's ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

- ▶ Follow-up work on AdFisher (Google ads, gender-based discrimination for the purposes of employment) ascertained that it was possible to target on gender for job ads
- ▶ Platforms have since taken steps to address such blatant violations

*“... Facebook currently has several policies in place to avoid discrimination for certain types of ads. Facebook also recently **built tools to automatically detect ads offering housing, employment, and credit**, and pledged to prevent the use of certain targeting categories with those ads. Additionally, Facebook relies on advertisers to self-certify that they are not in violation of Facebook’s advertising policy prohibitions against discriminatory practices. More recently, in order to settle multiple lawsuits stemming from these reports, **Facebook stated that they will soon no longer allow age, gender, or ZIP code-based targeting for housing, employment or credit ads**, and that they would also block other detailed targeting attributes that are “describing or appearing to relate to protected classes”.*

- ▶ Yet, the question still remains: **Does the ad delivery platform itself embed discriminatory outcomes?**

# Potential sources of discrimination

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

- ▶ First, platforms claim to **show users “relevant ads”**, maximizing the likelihood that a user will engage with the ad. Based on historical user engagement data, may result in skewed delivery in ways that an advertiser may not have intended
- Second, market effects and financial optimization can lead to skewed ad delivery. In a nutshell: **some populations are more “valuable” and so advertising to them costs more**. If an advertiser bids less, they won’t get to the more “valuable” population.



Facebook also disputed HUD's conclusion that the system itself discriminates beyond advertisers' choices: "HUD had no evidence and finding that our AI systems discriminate against people."

<https://www.propublica.org/article/hud-sues-facebook-housing-discrimination-advertising-algorithms>

# Facebook ad delivery

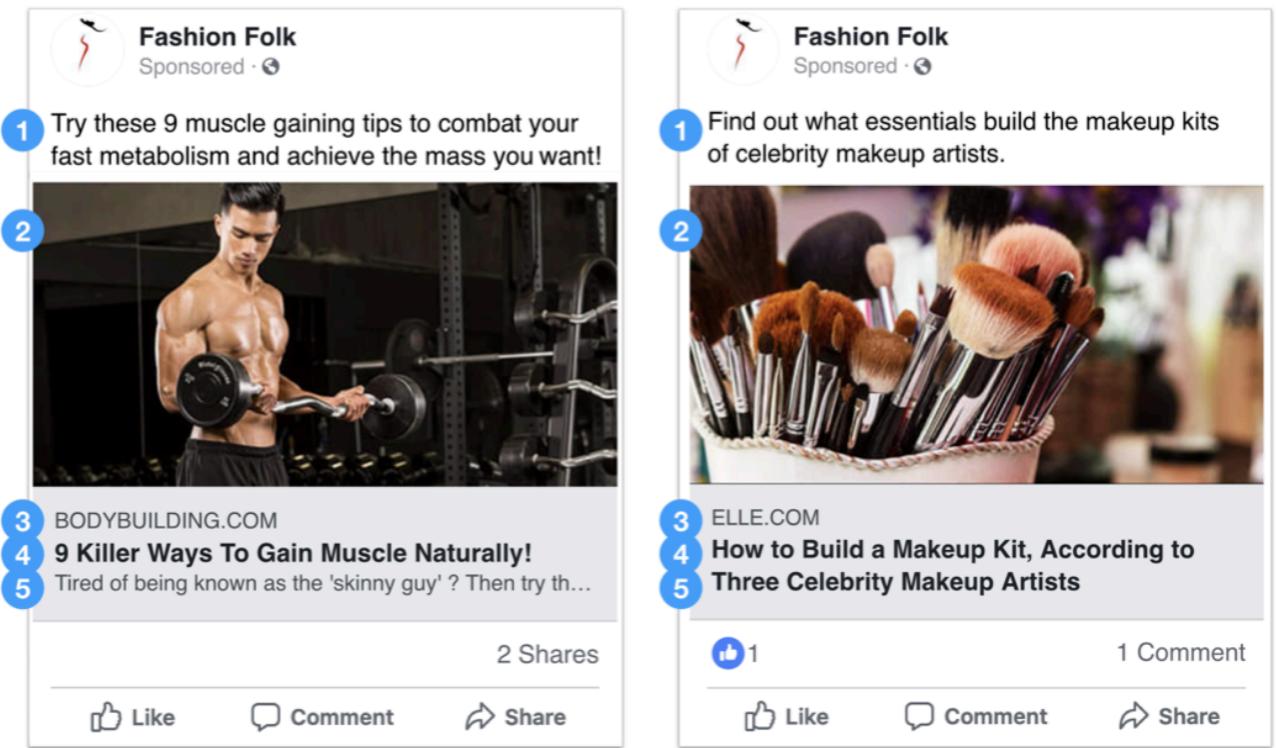
[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

## Part 1: ad creation

- ad contents
- audience selection
- bidding strategy

## Part 2: ad delivery

For every opportunity to show a user an ad (e.g., **an ad slot** is available as the user is browsing the service), the ad platform will run an **ad auction** to determine, from among all of the ads that include the current user in the audience, which ad should be shown.



**Figure 1: Each ad has five elements that the advertiser can control:** (1) the ad text, entered manually by the advertiser, (2) the images and/or videos, (3) the domain, pulled automatically from the HTML meta property og:site\_name of the destination URL, (4) the title, pulled automatically from the HTML meta property og:title of the destination URL, and (5) the description from meta property og:description of the destination URL. The title and description can be manually customized by the advertiser if they wish.

# Facebook ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

## Part 1: ad creation

- ad contents
- audience selection
- bidding strategy

## Part 2: ad delivery

For every opportunity to show a user an ad (e.g., **an ad slot** is available as the user is browsing the service), the ad platform will run an **ad auction** to determine, from among all of the ads that include the current user in the audience, which ad should be shown.

When Facebook has ad slots available, it runs an ad auction among the active advertisements bidding for that user. However, **the auction does not just use the bids placed by the advertisers**; Facebook says:

*“The ad that wins an auction and gets shown is the one with the highest **total value**. Total value isn’t how much an advertiser is willing to pay us to show their ad. It’s combination of 3 major factors: (1) Bid, (2) Estimated action rates, and (3) Ad quality and relevance.”*

*“During ad set creation, you chose a target audience ... and an optimization event ... **We show your ad to people in that target audience who are likely to get you that optimization event.**”*

# Facebook ad delivery: insights

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *CSCW 2019*]

Facebook ad delivery results can be skewed **in ways that advertisers do not intend**

Skew can arise due to:

- Financial optimization effects
- The ad delivery platform's predictions about the relevance of its ads to different user categories

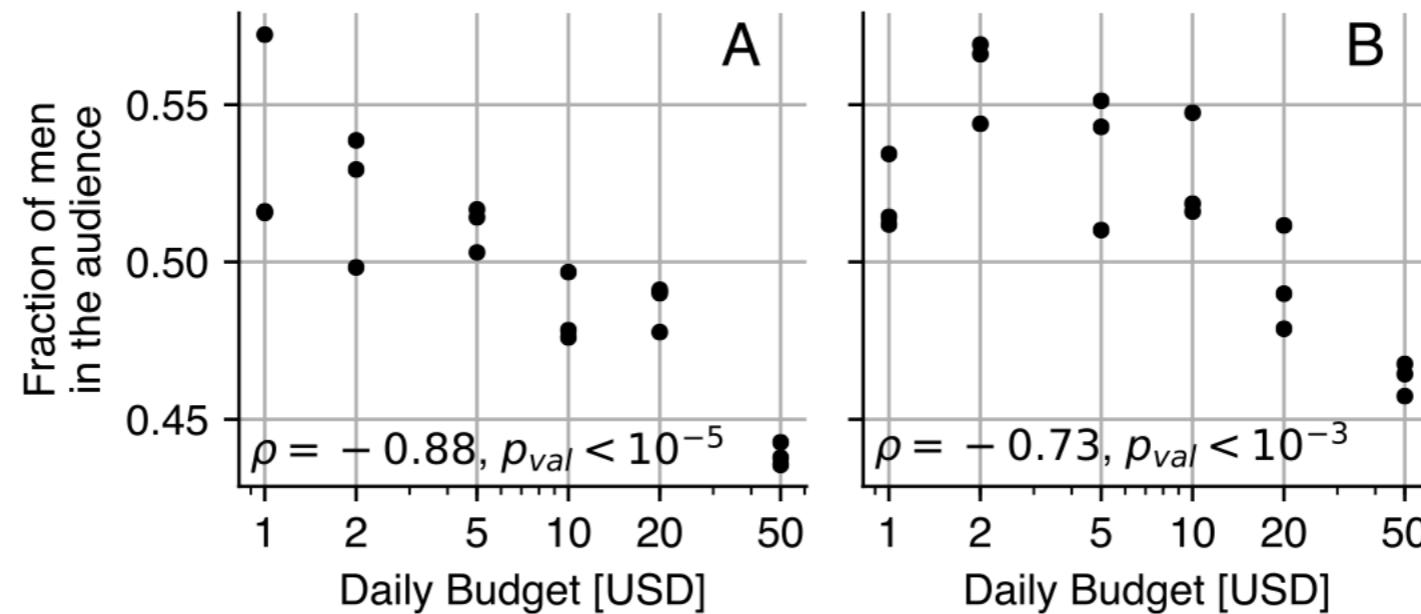
What contributes to the skew?

- Ad content; both text and images, which are likely automatically analyzed by Facebook
- Advertiser budget

**Skew was observed along gender and racial lines, in ads for employment and housing opportunities**

# Budget impacts demographics

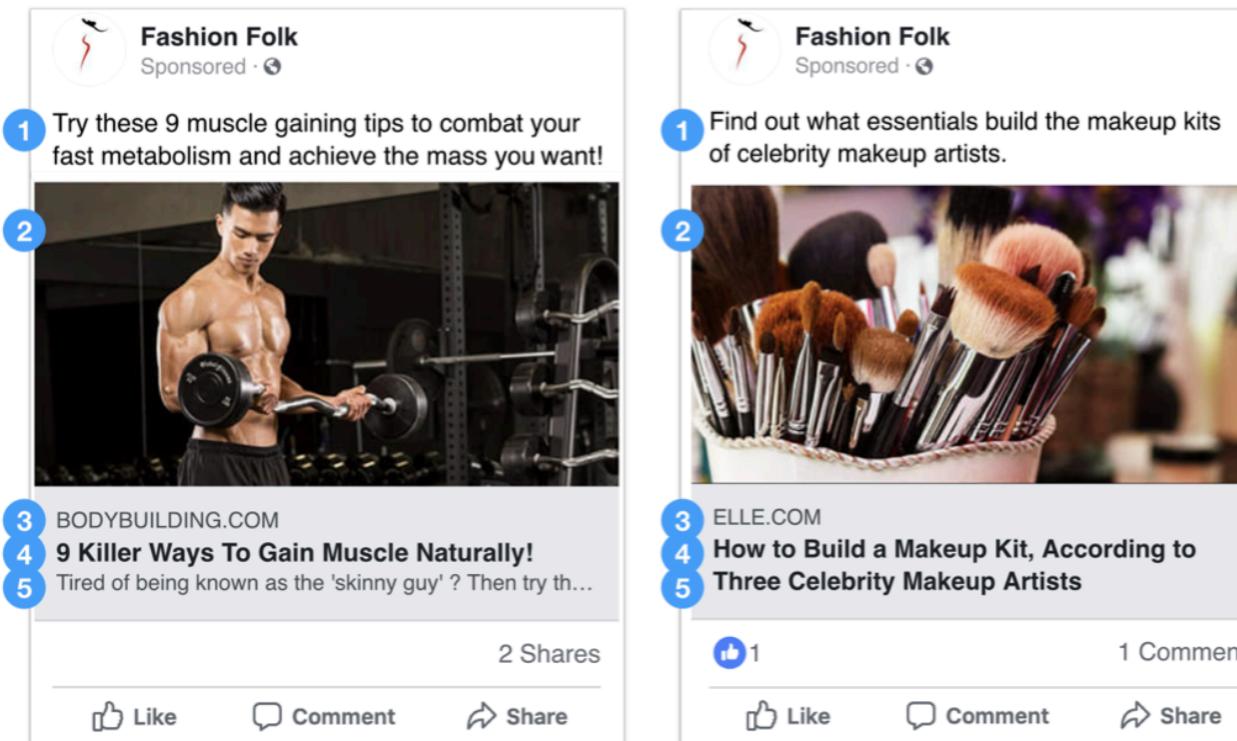
[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]



**Figure 2:** Gender distributions of the audience depend on the daily budget of an ad, with higher budgets leading to a higher fraction of women. The left graph shows an experiment where we target all users located in the U.S.; the right graph shows an experiment where we target our random phone number custom audiences.

# Ad creative impacts ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]



**Figure 1: Each ad has five elements that the advertiser can control:** (1) the ad text, entered manually by the advertiser, (2) the images and/or videos, (3) the domain, pulled automatically from the HTML meta property og:site\_name of the destination URL, (4) the title, pulled automatically from the HTML meta property og:title of the destination URL, and (5) the description from meta property og:description of the destination URL. The title and description can be manually customized by the advertiser if they wish.

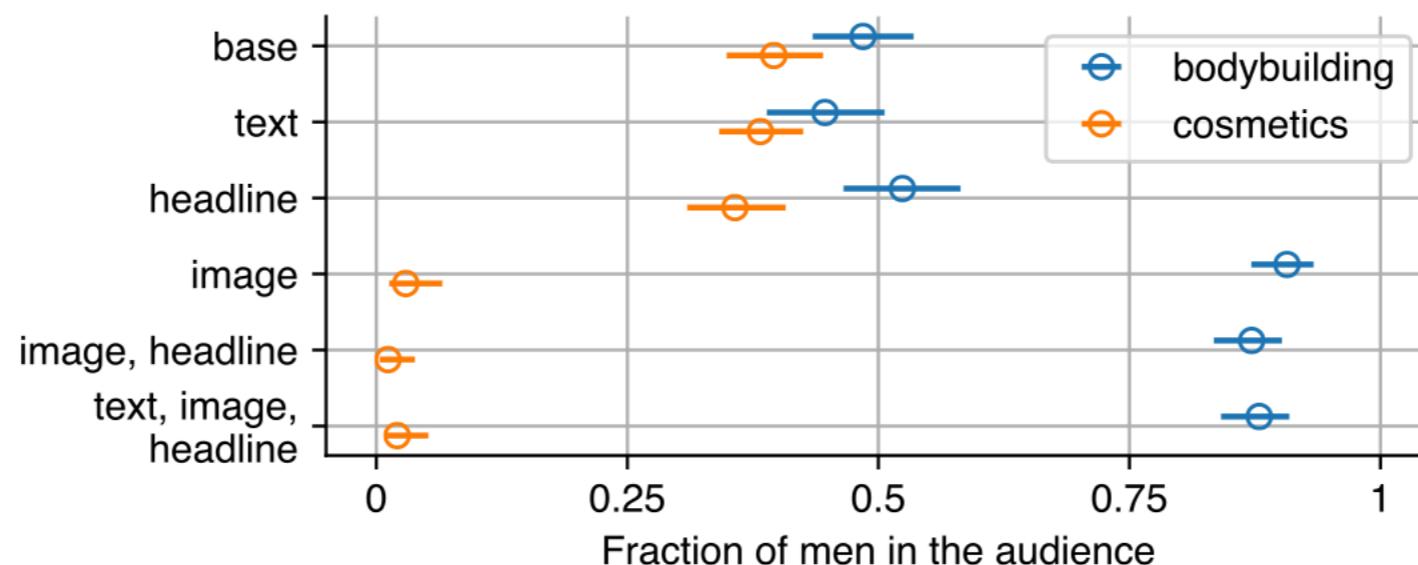
**Same bidding strategy** for bodybuilding and cosmetics, without explicitly mentioning gender

**Strong gender skew in delivery:** bodybuilding delivered to over 75% men on average, cosmetics delivered to over 90% women on average

# Ad creative impacts ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

Which component of the ad creative impacts delivery most?

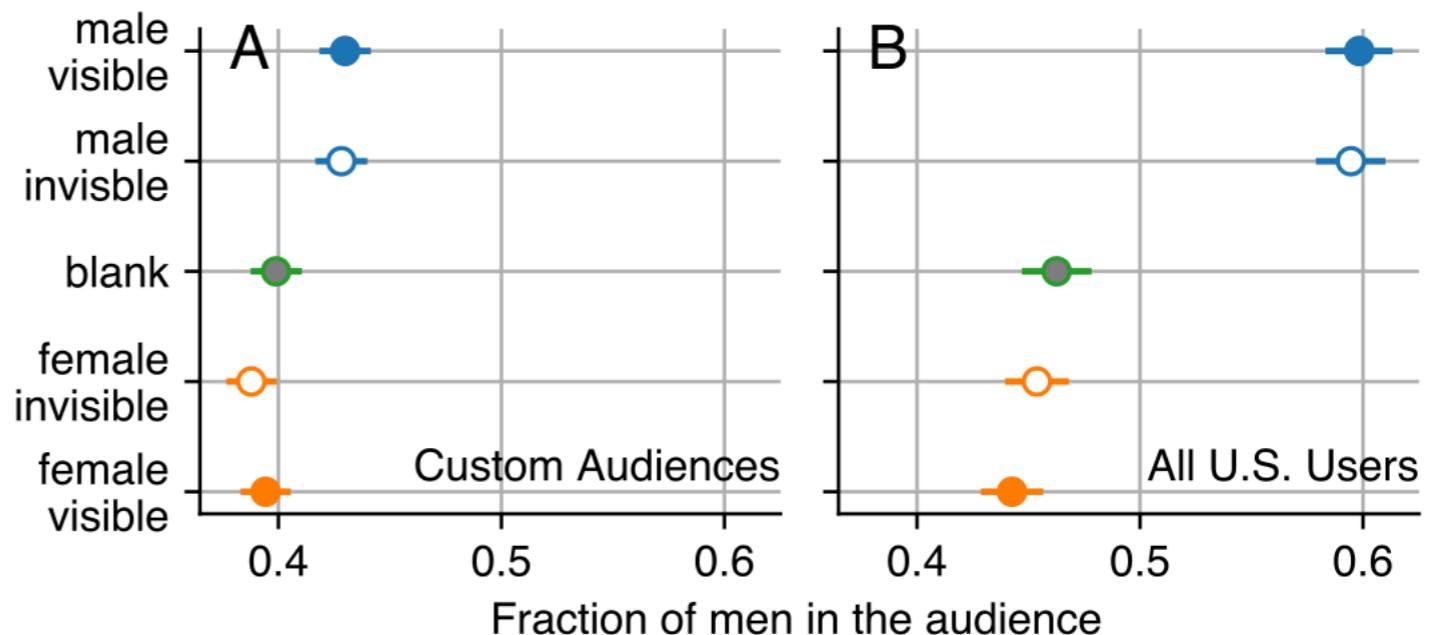


**Figure 3:** “Base” ad contains a link to a page about either bodybuilding or cosmetics, a blank image, no text, or headline. There is a small difference in the fraction of male users for the base ads, and adding the “text” only decreases it. Setting the “headline” sets the two ads apart but the audience of each is still not significantly different than that of the base version. Finally, setting the ad “image” causes drastic changes: the bodybuilding ad is shown to a 91% male audience, the cosmetics ad is shown to a 5% male audience, despite the same target audience.

# Transparent images are still targeted

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

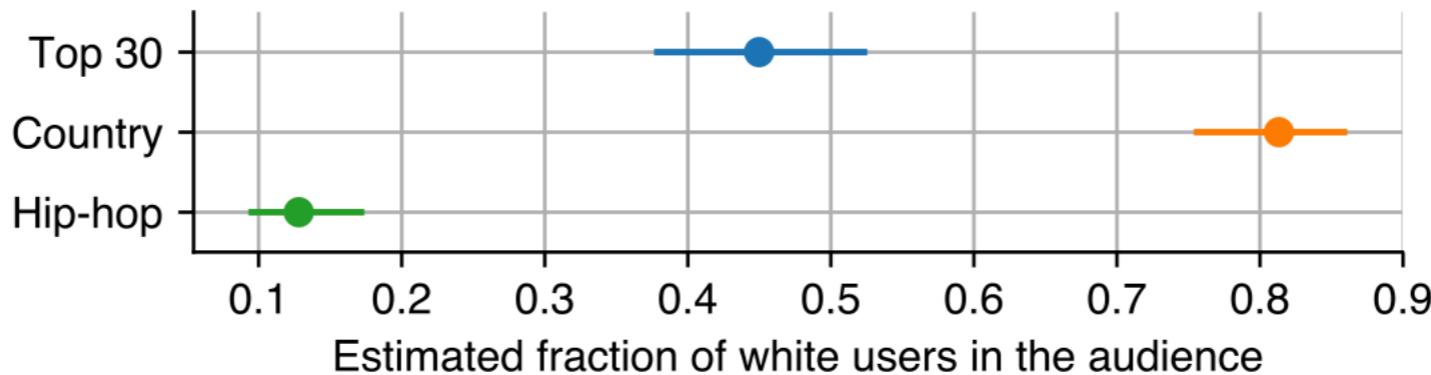
No.	Masculine		Feminine	
	Visible	Invisible	Visible	Invisible
1				
2				
3				
4				
5				



This strongly suggests that Facebook uses **an automated image classification mechanism** to steer different ads towards different subsets of the user population

# Entertainment ads targeted by race

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

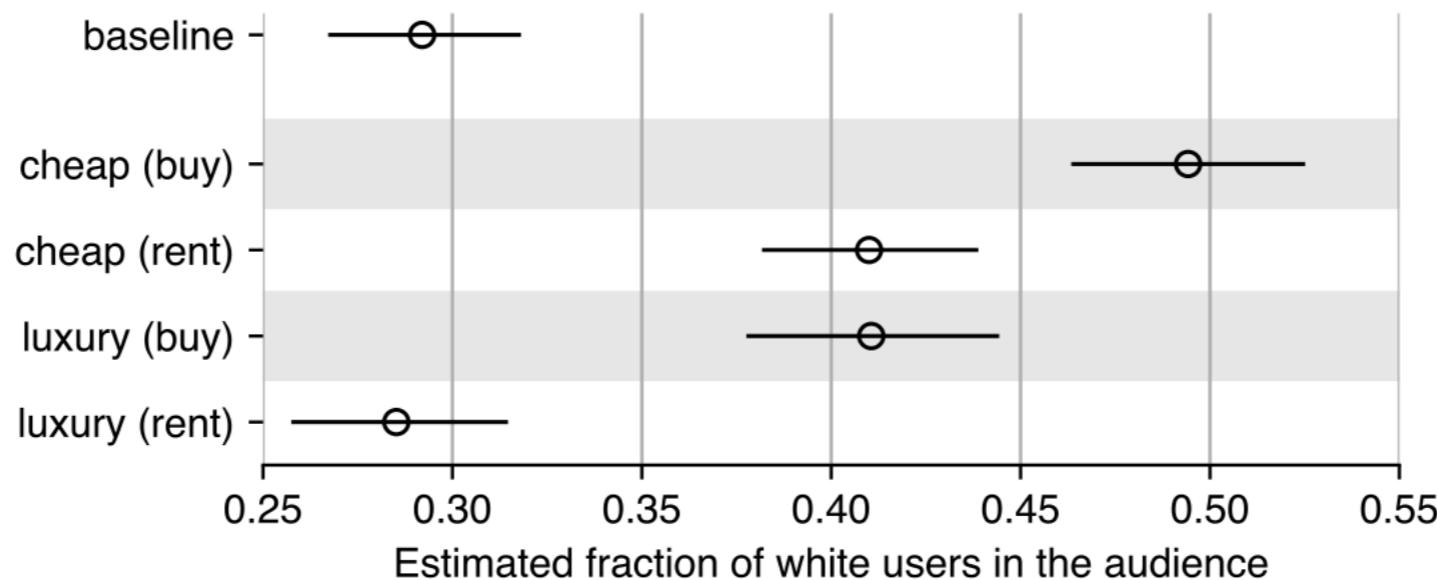


**Figure 7: We run three campaigns about the best selling albums. *Top 30* is neutral, targeting all. *Country* implicitly targets white users, and *Hip-hop* implicitly targets Black users. Facebook classification picks up on the implicit targeting and shows it to the audience we would expect.**

We hold targeting parameters fixed, run ads that are stereotypically of interest to different races. We find that **Facebook ad delivery follows the stereotypical distribution**, despite all ads being targeted in the same manner and using the same bidding strategy.

# Racial skew in housing ads

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]



**Figure 9: Results for housing ads, showing a breakdown in the ad delivery audience by race. Despite being targeted in the same manner, using the same bidding strategy, and being run at the same time, we observe significant skew in the makeup of the audience to whom the ad is delivered (ranging from estimated 27% white users for luxury rental ads to 49% for cheap house purchase ads).**

We cannot make claims about what particular properties of our ads lead to this skew, or about how housing ads in general are delivered.

# Skew vs. discrimination

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; CSCW 2019]

Throughout this paper we refer to differences in the demographics of reached audience as “skew” in delivery. **We do not claim any observed skew per se is necessarily wrong or should be mitigated.** Without making value judgements on skew in general, we do emphasize the distinct case of ads for housing and employment. In particular, the skew we observe in the delivery of ads for cosmetics or bodybuilding might be interpreted as reinforcing gender stereotypes but is unlikely to have legal implications. On the other hand, **the skew in delivery of employment and housing ads is potentially discriminatory in a legal sense.**

# Differences with traditional media

50

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *CSCW 2019*]

1. Advertiser has more control over ad placement in traditional media, while here Facebook can select a narrower audience and skew delivery in unexpected ways
2. More difficult for a user to break out of the **information bubble**, more difficult for an advertiser to reach a user if Facebook does not deem the ad “relevant” to that user
3. **Public interest scrutiny** is very-very-very...-very difficult