Discrimination in Online Advertising
A Multidisciplinary Inquiry

Amit Datta AMITDATTA@CMU.EDU
Anupam Datta DANUPAM@CMU.EDU
Carnegie Mellon University

Jael Makagon JAEL@BERKELEY.EDU
Deirdre K. Mulligan DMULLIGAN@BERKELEY.EDU
University of California, Berkeley

Michael Carl Tschantz MCT@ICSI.BERKELEY.EDU
International Computer Science Institute

Editors: Sorelle A. Friedler and Christo Wilson

Abstract

We explore ways in which discrimination may arise in the targeting of job-related advertising, noting the potential for multiple parties to contribute to its occurrence. We then examine the statutes and case law interpreting the prohibition on advertisements that indicate a preference based on protected class, and consider its application to online advertising. We focus on its interaction with Section 230 of the Communications Decency Act, which provides interactive computer services with immunity for providing access to information created by a third party. We argue that such services can lose that immunity if they target ads toward or away from protected classes without explicit instructions from advertisers to do so.

Keywords: Discrimination, online advertising, law.

1. Introduction

Recent studies demonstrate that computer systems can discriminate, including by gender (Datta et al., 2015; Caliskan et al., 2017; Kay et al., 2015; Lambrecht and Tucker, 2017; Bolukbasi et al., 2016), sexual orientation (Guha et al., 2010), and race (Sweeney, 2013; Angwin et al., 2016; Angwin and Parris, 2016). Although much scholarship exists on the legal consequences of discrimination, little work has explored the legal status of these concrete cases (Barocas and Selbst (2016) is the only one we are aware of). The consideration of such concrete cases, instead of abstract hypotheticals, forces us to confront the difficulties of proving a case based upon the limited evidence practically available to investigators. Such careful consideration can show what empirical evidence could aid the crafting of a case, which suggests new studies, and how laws might not be enforceable in practice. Furthermore, they have the potential to show that liability can lie with an advertising platform, not just in theory, but in practice. Such a finding can promote positive change and guide regulators to the interesting questions to ask.

An example of a real world difficulty is that while the existence of discrimination might be clear, the cause might not be. Computers may use factors associated with, but distinct from, protected attributes. This not only complicates the detection of discrimination, but also provides those intending to discriminate with a gloss of statistical rationality and leads fair-minded individuals to unwittingly discriminate via models that redundantly encode gender, race, or other protected attributes.

This paper provides a legal analysis of a real case, which found that simulated users selecting a gender in Google's Ad Settings produces employment-related advertisements differing rates along gender lines despite identical web browsing patterns (Section 3) (Datta et al., 2015). We then explore the operation of Google's advertising network to understand the various decision points that could contribute to the gender-skewed placement of such ads (Section 4). In doing so, we find that advertisers can use Google's advertising platform to target and serve employment ads based on gender. While we explore possible reasons that could have contributed to the discriminatory placement of ads, these explorations are not exhaustive. Uncovering the cause behind the discriminatory placement of ads requires further visibility into the advertising ecosystem or assumptions over how the ecosystem operates, and is beyond the scope of this paper.

We then explore legal questions and policy concerns raised by these results. Focusing on employment-related ads, we consider potential liability for advertisers and ad networks under Title VII, which makes it unlawful for employers and employment agencies "to print or publish or cause to be printed or published any ... advertisement relating to employment... indicating any preference, limitation, specification, or discrimination, based on ... sex".[1]

Due to the limited covered of Title VII we conclude that a generic advertising platform, like Google's, is unlikely to incur liability under Title VII's prohibitions regardless of any contributions they make to the illegality of an advertisement. Advertisements that run afoul of the Fair Housing Act's (FHA's) prohibition on indicating a preference however could create liability as unlike Title VII the FHA provision is of general applicability. In a case under the FHA, a court would need to consider how the advertising prohibition interacts with Section 230 of the Communications Decency Act (CDA),[2] which provides interactive computer services with immunity for providing access to information created or developed by a third party. Thus, we focus on the interaction between the prohibition on discriminatory advertising in the FHA and Section 230. We argue that despite the broad immunity generally afforded by Section 230, interactive computer services can lose that immunity if they target ads toward or away from protected classes. The loss of immunity is based on the act of targeting itself rather than any content that is contained within the four corners of the advertisement. We focus our analysis on Google, its system, documentation, consumer and advertising interfaces, and empirical research looking at it to provide useful details for our legal analysis. However, throughout, we generalize our analysis to generic machine learning systems where appropriate.

Our main contribution to the existing scholarship examining discrimination in automated decision-making is the analysis of the application of the discriminatory advertising prohibition in Title VII and the FHA in the light of Section 230. Our main novelty is drawing on the relevant regulations and case law under the parallel, but broader, provision in the Fair Housing Act, which has been more aggressively and creatively used.

We show the potential for ad platforms to face liability for algorithmic targeting in some circumstances under the FHA despite Section 230. Given the limited scope of Title VII we conclude that Google is unlikely to face liability on the facts presented by Datta et al. Thus, the advertising prohibition of Title VII, like the prohibitions on discriminatory employment practices, is ill equipped to advance the aims of equal treatment in a world where algorithms play an increasing role in decision making.

2. Related Work

We are not the first to consider possible causes of discrimination in behavioral advertising. Datta et al. (2015) themselves consider the question. Todd (2015) interviewed the parties involved looking for, but not finding, definitive answers. Lambrecht and Tucker (2017) conduct a study similar to Datta et al., but with more control, to analyze possible causes. Sweeney (2013) considers possible causes of discrimination in contextual advertising. We further discuss these works when we consider the causes they find likely.

Several law review articles have looked at the legal and policy implications of such outcomes and how policies can help prevent them. Barocas and Selbst (2016) discuss the difficulties in applying traditional antidiscrimination law as a remedy to discrimination caused by data mining (automated pattern finding). Kim (2016) explores the application of antidiscrimination norms of Title VII to computers making employment decisions and argues that this requires reassessment

1. §704(b) of Title VII of the Civil Rights Act of 1964, codified at 42 USC §2000e-3(b).
2. 47 USC §230.

of the laws. Kroll et al. (2016) explore how computational tools can ensure that automated decision making avoids unjust discrimination and conform with legal standards.

The most similar to our own work, Tremble (2017) applies Section 230 of the Communications Decency Act to content served by Facebook. While Section 230 of the Communications Decency Act frees interactive computer services like Facebook of liability for user generated content, Tremble argues that personalized content, like that on Facebook, constitutes content generated by Facebook and as such does not qualify for exclusion under Section 230.

3. A Prior Study of Google Ads

Datta et al. (2015) developed and used AdFisher, an experiment automation framework, to study how designating a consumer's gender in Google's Ad Settings profile affects Google ads. They find that indicating a male or female gender on Ad Settings produced different rates of job-related ads. Browsers set to male received more ads for a career coaching service that promoted high paying jobs than their female counterparts.

Specifically, Datta et al. carried out a randomized controlled experiment on one thousand simulated consumers (instances of the Firefox browser) using AdFisher. They randomly assign half of these consumers to configure their gender to male and the other half to female, and then have all consumers engage in identical web surfing behavior designed to signal job-hunting. Finally, they gather the advertisements displayed to each consumer on a news website. Using machine learning techniques, they identify gender-based ad serving patterns. Specifically, they train a machine learning classifier to learn differences in the served ads and to predict the corresponding gender. They then test whether the learnt patterns are statistically significant using the permutation test. This test avoids making common but questionable assumptions, such as ads being independent and identically distributed, that are unlikely to hold in highly dynamic advertising markets. They leverage the learnt classifier model to determine which ads were the strongest predictors of either gender and report them as top ads. See Datta et al. (2015) and Tschantz et al. (2015) for more details.

Using the permutation test, they find that the differences learnt by the machine learning classifier are indeed significant (p-value < 0.00005). Given the experiment's design, this result suggests with high certainty that the difference in the gender setting caused a difference in the ads served. As a consequence of using a randomized controlled experiment, the authors are able to conclude that the difference is not merely correlational but causal. The differences in the ads for the two genders is of potential concern. The top two ads for indicating a male were from a career coaching service, The Barrett Group, for "\$200k+" executive positions. Google showed the ads 1852 times to the male group but just 318 times to the female group. The top two ads for the female group were for a generic job posting service and for an auto dealer.

Thus, Datta et al. establish that indicating gender in Ad Settings affects displayed ads. Owing to the blackbox nature of their experimental setup, they are not able to explain how or why the gender setting caused the difference in ads served. In the next section, we consider some possible causes of their results.

4. Possible Causes of Discrimination

We will now consider possible ways that the results discovered by Datta et al. can manifest in an online advertising ecosystem. The advertising ecosystem is a vast, distributed, and decentralized system with several actors. There are *publishers* who host online content, *advertisers* who seek to place their ads on publishers' websites, *ad networks* who connect advertisers and publishers, and *consumers* who consume online content and ads. (The Supplementary Materials provide a more detailed description of the ad ecosystem.)

Each actor has a set of primary mechanisms through which they can introduce a difference in how men and women are treated (Factor I in Table 1). Thus, we can view the first factor as saying *who* creates the inputs that might contribute to a discriminatory outcome. In all cases, the impact of the input, and in some instances its availability, is ultimately determined by Google. Indeed, by being the central player connecting the parties, Google always plays a role. While the simulated users surely played a role in the selec-

Table 1: Possible Causes of the Datta et al. Finding Organized around Four Actors

Factor I: (Who) Possible mechanisms leading to males seeing the ads more often include:

1. (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);
2. (The advertiser)The advertiser's targeting of the ad through explicit use of demographic categories (explicit and intentional discrimination), the pretextual selection of demographic categories and/or keywords that encode gender (hidden and intentional), or through those choices without intent (unconscious selection bias), and Google respecting these targeting criteria;
3. (Other advertisers)Other advertisers' choice of demographic and keyword targeting and bidding rates, particularly those that are gender specific or divergent, that compete with the ad under question in Google's auction, influencing its presentation;
4. (Other consumers)Male and female consumers behaving differently to ads
   (a) Google learned that males are more likely to click on this ad than females,
   (b) Google learned that females are more likely to click other ads than this ad, or
   (c) Google learned that there exists ads that females are more likely to click than males are; and
5. (Multiple parties) Some combination of the above.

Factor II: (How) The mechanisms can come in multiple favors based on how the targeting was done

1. on gender directly
2. on a proxy for gender, i.e. on a known correlate of gender because it is a correlate),
3. on a known correlate of gender, but not because it is a correlate, or
4. on an unknown correlate of gender.

tion of ads by indicating their gender, this is not included in our analysis because it would suggest that, by admitting one's gender, a consumer bore some responsibility for the potentially discriminatory result. We do not believe this position to be technically accurate, nor legally defensible.

With respect to each actor we consider *how* the results may have occurred (Factor II in Table 1). Where appropriate we consider the use of gender as a targeting criteria, the intentional and unintentional use of features that correlate with gender and the impact of the bidding system.[3]

4.1. Google's Actions Alone

Google created the entire advertising platform. It designed the AdWords interface that allows advertisers to target ads based on inputs including gender. Its terms of use admonishes advertisers to comply with all applicable laws and regulations. Through examples it specifies areas where advertisers have in the past run afoul of the law.

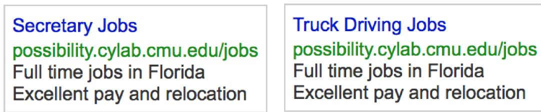
However, bans on sex-based targeting of employment, housing, and credit are not specifically addressed. Google has a set of policies for interest-based advertising that prohibit using any "sensitive information" about site or app visitors to create ads. While race, ethnicity, sexual orientation, and religion are considered "sensitive information", gender is not.

Given its control over the platform there are many ways in which Google could have caused or contributed to the difference in advertisements directed to men and women observed by Datta et al. (Case 1 of Factor I). A Google employee could have manually set the ad to target by gender or a feature associated with gender. While presumably the advertising system is largely autonomously driven by programs, researchers have documented that even in highly automated systems, such as search, a sizable amount of manual curation occurs (Gillespie, 2014).

4.2. Direct Targeting of Gender by Advertisers

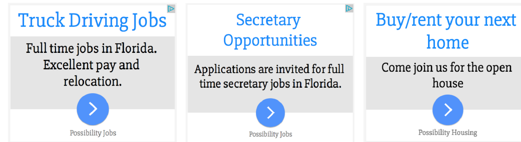
Advertisers, including The Barrett Group, which showed the ad in question, can make multiple de-

3. Since correlation is the most familiar form of statistical association, we use correlations in this paper, but all our statements may generalize to other forms of association.



(a) (b)

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



(a) (b) (c)

Figure 2: Ads approved and served by Google in 2017: truck driver jobs only to men, for secretary jobs only to women, for housing disparately.

cisions through the AdWords interface that could steer their ads toward or away from women. The simplest way gender-skewed advertising could have emerged is if the advertiser directly targeted on gender (i.e. Factor I.2+Factor II.1). AdWords offers the ability to set demographic parameters to explicitly target ads toward, or away from, a single sex. While such explicit intentional gender targeting is supported by the AdWords interface, we wanted to explore whether the Barrett Group could actually use this feature to target their advertisement. To do so we performed another study in three phases.

First, in 2015, we constructed several ad campaigns that targeted job-related ads on the basis of gender using Google's advertising platform, AdWords. Figure 1 shows two of the ads that were approved by Google. Ad 1(a) is for a secretary job targeted towards women, while ad 1(b) is for a truck-driving job targeted towards men. The other pairs of differentially targeted ads varied by pay, seniority level, and educational requirements. (We show them in the Supplementary Materials.) Our ads all had the same display and destination URLs [4]. This page has the words "Test ad. No jobs here." We also verified that Google rejects some advertisements at this stage by intentionally submitting ads with broken links or excessive exclamation points and found these were not approved.

Second, in 2017, we again tested Google's ad approval procedure and, this time, found it to be somewhat more sophisticated. While we were able to get one ad approved with the same destination URL and ad text as in Figure 1(b), the other ads were disapproved. In particular, Google AdWords reported the destination was not working and the content was misleading (shown in the Supplementary Materials). However, by changing the ad text and destination URL as well as adding more text to the destination webpage, we got the second ad approved.

Third, while these explorations make it clear that Google AdWords allows creation of discriminatory job ad campaigns, it leaves open the possibility that Google would prevent the gender-targeted employment ads from being delivered at a later point in the process. As our last step, to check whether this is the case, we enabled both the ad campaigns at the same time (differing by a few seconds) for about 12 hours in 2017. We observe that both the campaigns receive several thousands of impressions, with the truck driver campaign receiving over 70k impressions and the secretary campaign receiving over 55k impressions. The campaigns collectively cost less than \$100. The demographics of the users receiving the impressions exactly matched the targeting criteria. All the truck driver ad impressions were to men (or consumers who Google believes are men) and those for the secretary ad were all to women. This finding demonstrates that an advertiser with discriminatory intentions can use the AdWords platform to serve employment related ads disparately based on gender.

We also had ads for housing approved, targeted and served disparately (Figure 2(c)). The ad was suggestive of attending a open house for buying or renting a house. The final destination, however, had text indicating that the ad was created and served as part of a study.[5] This ad was targeted to both male and female demographics who were *American Football Fans* or *Baseball Fans*. These interests were chosen intentionally to target the male demographic more. With

4. possibility.cylab.cmu.edu/jobs

5. possibility.cylab.cmu.edu/housing

these interests, Google's AdWords estimated that the ad would be targeted to between 1B and 5B women and between 5B and 10B men. We kept the ad was kept active for about 24 hours and observed that the ad received over 23k impressions, of which 75% were to men. This again demonstrates that an advertiser can intentionally use proxies for gender to target housing ads disparately based on gender.

We look at this scenario in detail to explore whether an employer or employment agency using Google's ad network can engage in explicit, intentional discrimination. Using AdWords, the career coaching service, Barrett Group, could have intentionally targeted their ad toward males, or limited targeting to females. Our small study also suggests that Google's review process does not weed out employment-related advertisements explicitly targeted by gender. While our study shows that such direct targeting as one possible explanation for the advertising outcome, it cannot tell us whether whether the Barrett Group actually used the demographic choices to target their advertisement.

4.3. Other Possibilities for the Advertiser

In fact the Barrett Group denied targeting on gender and claimed to have sought those older than 45, receiving high pay, and of executive-level experience (Todd, 2015), pointing to another possibility: the Barrett Group could have made other choices that led indirectly to targeting on gender. For purposes of this analysis we set aside the issue of targeting based on age which is an independently prohibited act by employers and employment agencies under Title VII. This explanation points to the possibility of the advertiser choosing interests or keywords correlated with gender (i.e. Factor I.2 + Factor II.{2 or 3 or 4}). Given the targeting criteria, it seems reasonable to assume that on average ad placement would skew toward men. On average, men earn more than women. Numerous studies have documented the under representation of women in the executive suite (e.g., in 2016, only 4% of Fortune 500 CEOs were women (Zarya, 2016)). Thus, one could conclude that The Barrett Group intended to use the attributes as a proxy for gender to target males without appearing to do so—a practice called *masking* (Factor II.2). However, it may be that these were among a broader set of attributes and keywords that The Barrett Group selected, which reduces our concern with masking, that still redundantly encoded gender. Even a series of attributes that alone or in consort do not appear gender-specific may be found through statistical techniques to correlate with a gender (Factor II.4). Selecting such correlates could result in Google showing the ads more to men by attempting to satisfy The Barrett Group's request to target the correlates.

The conditions of Datta et al.'s study, however, allow us to probe this issue further. The simulated users lacked other attributes which could be correlated with gender. They all engaged in the same behavior. As a result, we can rule out that the difference in received advertisements resulted from differences in age, affluence, or prior work experience. If Google inferred these attributes from user behavior, all thousand users should have resulted in identical inferences. The only differentiating data in these experimental conditions was gender. If Google did use correlations with gender, it used correlations found in other populations of real consumers and applied them to Datta et al.'s synthetic population.

4.4. Decisions of Other Advertisers

Other advertisers can influence the targeting of an advertiser, such as The Barrett Group, through the selection of their ad auction bids. This possibility was raised by Google itself (Todd, 2015). If advertisers other than The Barrett Group were willing to pay more to reach women users, The Barrett Group's advertisement may have ended up predominantly being served to men (Factor I.3). If advertisers in general consider female consumers to be a more valuable demographic, they would set higher bids to advertise to them. As a result, if an advertiser, like The Barrett Group, sets equal bids for men and women, it could end up only reaching men if it is outbid by other ads for female users. The real-time auction makes it difficult for advertisers to figure out how to treat protected classes similarly.

In their study of Facebook ads, Lambrecht and Tucker (2017) suspect that the higher competition for reaching younger women was the reason behind lower impressions of job-related ads for women than men, in spite of gender-neutral targeting criteria and bids.

4.5. Behavior of Other Consumers

User behavior could also play a role in the disparate ad results found by Datta et al. Google's understanding of which users are likely to respond to an ad is built off observations of millions of users' behavior (Factor I.4). For example, Google could have found that (a) males are more likely to click this ad than females are, (b) females are more likely to click other ads than this ad, or (c) there exist other contemporary ads that females are more likely to click than males are. Google's computers may have targeted The Barrett Group's ads in response to one of the above findings to increase revenue.For example, suppose The Barrett Group pays Google per click (i.e. using the cost-per-click bidding strategy), then ad serving models that are developed over time to maximize Google's revenue may end up serving the ads to more men and fewer women.

To the extent user behavior over all expresses sex stereotypical responses to ads about job opportunities, using their behavior as an input risks building the product of sexist hiring practices, and general employment inequality, into the targeting. For example we know there are fewer women currently in the executive tier of companies so they may self-select away from The Barrett Group ad, while males who are overrepresented in the executive tier may aggressively click on it. Using these inputs to target ads constrains women's access to job advertisements based on prior patterns of discrimination and inequality reflected in the stereotypical responses of women as a whole. Sweeney (2013), who found disparate serving of ads indicating arrest records based on the race-affiliation of first names, also suggested that user inputs may have resulted in the disparity. After a conversation with the advertiser who claimed to have provided the same ad text to Google for groups of last names, she hypothesized that the bias in served ads might result from a society that "clicked ads suggestive of arrest more often for black identifying names".

The above possibilities are by no means exhaustive. In addition to variations of the above, there exist also completely different possibilities, such as the difference arising solely due to an error in Google's code or even from malicious outsiders (e.g., hackers) purposefully manipulating Google's computer systems.

We have seen that each actor in the advertising ecosystem may have contributed inputs that produced the discrepancy in ads observed by Datta et al. Without additional information it is impossible for us to know what actors—other than the users receiving the ads—did or did not do. It is also impossible to assess whether the advertisers or Google intended to target advertisements based on gender, or whether they were unaware such gendered distribution was occurring. In several instances, answers to these questions would be necessary to assess the extent, if any, of legal liability. As we discuss in Section 5 below, in two instances, however, we can draw conclusions about legal liability without assessing intent or knowledge. Liability for violating the advertising prohibition does not turn on intent; it is essentially a strict liability standard.[6] Second, Section 230 of the Communications Decency Act limits the liability interactive computer services face for content that they have not developed or created. The size of the shield against liability §230 provides to Google can be assessed without consideration of intent, and combined with the text of Section 704(b), creates an important limit on Google's exposure.

5. Title VII and Prohibitions on Discriminatory Advertising

Title VII of the 1964 Civil Rights Act makes it unlawful to discriminate on the basis of race, color, national origin, religion, or sex in several stages of employment. Title VII also prohibits employers, labor organizations, employment agencies, and joint labor-management committees from engaging in advertising that indicates a preference based on sex:

> It shall be an unlawful employment practice . . . to print or publish or cause to be printed or published any notice or advertisement relating to employment by such an employer

6. Housing Statements and §3604(c): A New Look at the Fair Housing Act's Most Intriguing Provision, 29 Fordham Urb. L.J. 187, 215-16 (2001) (describing parallel advertising prohibition as "essentially a strict liability" standard)

> [or other entity covered by the statute], or relating to any classification or referral for employment by [such entity] ... indicating any preference, limitation, specification, or discrimination based on race, color, religion, sex, or national origin.[7]

Eventually, the EEOC interpreted this as a flat prohibition on the use of sex-specific help-wanted advertising columns by covered entities.[8] (The Supplementary Materials provide details.)

While there is limited case law interpreting the advertising prohibition in Title VII, the significant case law and guidance informing the application of a similar provision of the Fair Housing Act (FHA)[9] offers guidance on its scope. Both the statutory parallels and shared objectives of Title VII and the FHA suggest that the FHA case law and the guidance documents issued by the US Department of Housing and Urban Development (HUD) interpreting the FHA's prohibition on discriminatory advertisements,[10] provide a reasonable resource for contemplating the interpretation of Section 704(b), and its potential application to online behavioral advertising.

There are many ways to indicate improper preferences through advertising. These include not only the written or visual text of the ads, but also the ways in which advertisements are distributed or targeted. The explicit prohibition on sex-specific advertising columns in Title VII are one example of the way in which improper preferences can be revealed outside the text of the advertisement itself. In the context of the FHA, courts have found that a city's "refusal to publicize jobs outside [a] racially homogenous [white] county" was evidence of discrimination.[11] In the fair housing context,[12] regulations issued by HUD confirm that such targeting can indicate an illegal preference, stating that "selecting media or locations for advertising... which deny particular segments of the housing market information" or "refusing to publish advertising for the sale or rental of dwellings..." because of a protected class indicates a discriminatory preference.[13] Other activities that can indicate a discriminatory preference include publishing advertisements exclusively in a language other than English[14] and indicating a language preference, which could mask a preference for people of a specific national origin.[15]

5.1. Scope of Title VII

Before turning to an analysis of Section 704(b), it is important to note that the law creates a significant limitation on avenues for relief under the facts of Datta et al.'s research. Section 704(b) only prohibits certain kinds of entities from printing or publishing discriminatory advertisements: employers, labor organizations, employment agencies, and joint labor-management committees. For this prohibition to apply to the ads investigated by Datta et al., The Barrett Group, Google, or both would have to fall within the definition of one of these entities.

The Barret Group describes itself as an executive career coaching service and it does not appear to be affiliated with or promise to procure opportunities to work for particular firms, so it seems unlikely to be considered an employment agency. Additionally, there is no evidence that The Barret Group is affiliated with a joint labor-management committee. Although Google's vastly complex structure and the difficulty of knowing exactly what ads run through the platform make it difficult to be certain, we believe it is unlikely that Google would be considered an "employment agency"—an entity "regularly undertaking with or without compensation to procure employees for an employer or to procure for employees opportunities to work for an

7. Section 704(b) of Title VII of the Civil Rights Act of 1964, codified at 42 USC §2000e-3(b) (gender alone may be used where it is a bona fide occupational qualification for employment).
8. 29 C.F.R. §1604.5.
9. 42 USC §3601 et seq.
10. 42 USC §3608.
11. United States v. City of Warren, MI, 138 F.3d 1083 (6th Cir. 1998).
12. Given the statutory parallels and shared objectives of Title VII and the FHA, an examination of the FHA case law and the guidance documents issued by HUD interpreting the parallel prohibition on advertisements (42 USC §3608 (2014)), provides useful insight on how 2000e-3 could be interpreted, and its potential application in the context of online behavioral advertising.
13. 24 C.F.R. §100.75
14. Hous. Rights Ctr. v. Sterling, 404 F. Supp. 2d 1179, 1193-94 (C.D. Cal. 2004) (notices and banners in Korean would suggest to the ordinary reader a racial preference for Korean tenants.)
15. Holmgren v. Little Village Community Rptr., 342 F. Supp. 512, 513 (N.D. Ill. 1971)

employer"[16]—given the availability of specialized platforms such as Craigslist's online classifieds, LinkedIn's professional networking platform, and Monster.com's job boards. We therefore set aside this question, but we note that in addition to Federal civil rights law, laws in several states including California, New York and Pennsylvania, prohibit any person from aiding, abetting, inciting, compelling, or coercing discriminatory employment practices. These laws create potential liability for Google if its services are used by covered entities to target ads based on gender or other protected class.[17]

Despite our conclusion that Title VII is unlikely to reach The Barrett Group or Google under the facts in Datta et al.'s research, we believe it is useful to consider whether the law could create liability for advertising platforms under the similar but broadly applicable provision in the FHA, which we discuss in more detail below. This analysis requires exploring the various ways in which an illegal preference could be communicated to the public and how those variations interact with the prohibition in Section 230 on holding Interactive Computer Services, such as ad platforms, liable for content.

5.2. Ad Content and Ad Targeting

Courts analyze advertisements based on whether an ordinary reader or listener (or viewer) would interpret the advertisement to convey a preference based on a protected class.[18] Because the statute focuses on the perspective of the recipient, the intent behind the content or targeting of the ad is not relevant to whether it violates Section 704(b).[19] This is an important factor in the online behavioral advertising environment. The FHA case law connects the ordinary reader standard to the prohibition on sex-designated advertising columns by explaining that advertisements that exclusively feature white models "may discourage black people from pursuing housing opportunities by conveying a racial message in much the same way that the sex-designated columns... furthered illegal employment discrimination."[20] While informational models used to target specific populations make the expression of preference more difficult to see in one sense—the advertisements are literally withheld from the undesired class—the sort of analysis conducted by Datta et al. reveals that such targeting communicates a preference more effectively than "subtle methods of indicating racial preferences"[21] already barred by courts.

In sum, the ban on sex-specific advertising columns, case law, and guidance provided under Title VII and the FHA aim to limit both *content* and *activities* that target advertisements based on protected attributes. The regulations and case law limit activities that direct information about employment and housing opportunities to or away from individuals based on membership in a protected class. Advertisements can run afoul of Title VII both substantively through content choices, and procedurally through publishing decisions that affect the literal availability of advertisements to different recipients, or otherwise indicate an illegal preference.

Our focus in this analysis is not on the content of the ads identified by Datta et al. These ads appear neutral using phrases such as "$200k+ Jobs—Execs Only" or "Goodwill—Hiring".

Instead we explore how advertising platforms create new risks that access to information about job or housing opportunities will vary based on protected status, regardless of the intent of advertisers, and consider how such targeting would be dealt with differently under two key civil rights laws. Such targeted advertising—

16. 42 USC §2000e(c).
17. Nat'l Org. for Women v. State Div. of Human Rights, 314 N.E.2d 867, 870–71 (Ct. App. N.Y. 1974); Pittsburgh Press Co. v. Pittsburgh Comm'n on Human Relations, 287 A.2d 161, 169 (Pa. Cmmwlth. 1972); Alch v. Superior Court, 122 Cal. App. 4th 339, 389 n.48 (2004).
18. Rodriguez v. Vill. Green Realty, Inc., 788 F.3d 31, 53 (2d Cir. 2015) ("What matters is whether the challenged statements convey a prohibited preference or discrimination to the ordinary listener."); Ragin v. New York Times Co., 923 F.2d 995, 999-1000 (2d Cir. 1991) (explaining that readers can "infer a racial message from advertisements that are more subtle than the hypothetical swastika or burning cross").
19. Capaci v. Katz & Besthoff, Inc., 711 F.2d 647, 660 (5th Cir. 1983) (finding an ad violated the act although the "practices in composing and placing ads were not to carry out any policy of discrimination against women, but to achieve the best results from the ads in light of her experience as to the gender which would be more interested in the job vacancy being advertised")
20. Ragin v. New York Times Co., 923 F.2d 995, 1003 (emphasis added).
21. Ragin v. New York Times Co., 923 F.2d 995, 1000.

whether it involves placing neutral ads in sex-segregated columns[22] or advertising only to a certain demographic—can be just as damaging to equal opportunity as an employment ad that says "Women Need Not Apply." Thus, our concern is with dissemination choices that convey unlawful preferences regardless of an ad's content.

5.3. Online Ads and Civil Rights

The world of *Pittsburgh Press Co.* and similar cases, where prohibited preferences and discrimination were painfully obvious in the form of sex-segregated help-wanted columns, is long gone. Increasingly, advertisements are moving online and are being handled by large advertising platforms such as Google and Facebook.[23] These companies are generally considered to be "interactive computer services" and protected from liability as a publisher or speaker of content created and developed by others under Section 230 of the Communications Decency Act (CDA).[24]

The involvement of interactive computer services in distributing ads raises the question of the relationship between activities civil rights law prohibits—dissemination choices, venue selection, and/or steering (rather than textual indications of preference)—and the prohibition in Section 230 on holding people liable as publishers of content they did not create or develop.[25] In particular, like other recent cases involving civil rights statutes and Section 230, it raises a question of whether any of the functionality offered to third parties to indicate preferences, or independent activities conducted by advertising platforms that determine who sees advertisements, rise to co-development of the advertisement.

To answer this question, we must examine the connection between publishing and advertising. Section 230(c)(1) protects interactive computer services from publisher liability even where those services might be engaging in activity traditionally associated with publishers, such as editing or removing content. Interpretation of the term "publish" in the Fair Housing context suggests that targeting advertisements is publishing activity, and can independently indicate an illegal preference. For example, in *Mayers v. Ridley*, "publishing" of a discriminatory statement was found where a Recorder of Deeds collected restrictive covenants "in a manner that facilitates access to them by prospective buyers."[26] More broadly, the court noted that "the proscription against 'publication' should therefore be read. . . to bar all devices for making public racial preferences in the sale of real estate, whether or not they involve the printing process."[27]

As Datta et al. demonstrate, The Barrett Groups ads disproportionately targeted men. But, did that targeting indicate a preference for men? Unlike the gendered help wanted columns, the classifier used to target ads was not revealed in written text to the recipients. It is possible that a recipient of a Barrett Group advertisement might have noted an "about this ad" symbol next to it, clicked on it, and received some information about why they received the advertisement. Another possibility is a user might have looked

22. Pittsburgh Press Co. v. Pittsburgh Comm'n on Human Relations, 413 U.S. 376 (1973) (Supreme Court affirmed that sex-segregated columns for employment ads in a newspaper were in and of themselves discriminatory, even if the specific text of the ads was sex-neutral).
23. Between 2000 and 2015, print newspaper advertising revenue fell 65% (from around $60 billion to around $20 billion). Derek Thompson, The Print Apocalypse and How to Survive It, The Atlantic (Nov. 3, 2016).
24. 47 USC §230(c)(1).
25. There is also a critical question of whether Section 704(b) would apply to Google at all, given that it probably does not fall within any of the categories listed in the statute (employer, labor organization, employment agency, or joint labor-management committee). For purposes of this analysis we will set this question aside. Courts and the EEOC eventually concluded that newspapers were prohibited from displaying sex-segregated ads, despite the fact that newspapers are not included as a category in Section 704(b). Most corresponding state statutes prohibit aiding and abetting discriminatory ads, which provides another potential avenue for arguing that Google is subject to civil rights obligations.
26. Mayers v. Ridley, 465 F.2d 630, 633 (D.C. Cir. 1972).
27. Mayers v. Ridley, 465 F.2d 630, 633 (D.C. Cir. 1972). See also United States v. City of Warren, MI, 138 F.3d 1083 (6th Cir. 1998) (city violated Title VII by purchasing recruitment ads in publications with disproportionately white readers); Hous. Rights Ctr. v. Sterling, 404 F. Supp. 2d 1179, 1193-94 (C.D. Cal. 2004) (notices and banners in Korean would suggest to the ordinary reader a racial preference for Korean tenants); Holmgren v. Little Village Community Rptr., 342 F. Supp. 512, 513 (N.D. Ill. 1971) (defendant newspapers violated FHA prohibition on discriminatory advertising by publishing ads indicating a preference for buyers or tenants that spoke a particular language).

at their ad preferences and noted that they were identified as male and assumed that The Barrett Group advertisement was being targeted to them based on that criteria.

As discussed above, indications of illegal preference can be conveyed to users in more subtle and less literal ways. Targeting that has the effect of limiting an audience in a discriminatory way, even though it does not convey a preference within the advertisement itself, is addressed by both regulations and case law. The Barrett Group claims to have targeted their advertisement to those older than 45[28], receiving high pay, and of executive-level experience (Todd, 2015). It is less clear whether they indicated a preference for men over women. As discussed above, the compliance manual states that "employers are prohibited from structuring their job advertisements in such a way as to indicate that a group or groups of people would be excluded from consideration for employment."[29]

It seems that the choices The Barrett Group made could be viewed as an indication that men were preferred over women for certain jobs. We noted that HUD regulations state that "selecting media or locations for advertising which deny particular segments of the housing market information" because of a protected class indicates a discriminatory preference.[30] The input selections made by The Barrett Group denied particular segments of the market, women, information about a job-related opportunity. However, assuming The Barrett Group was truthful about the inputs, it is unclear whether the EEOC would find that in doing so they indicated a discriminatory preference.

It would seem an odd outcome if employers prohibited from advertising in gender specific help wanted columns that signaled gender preference but were at least practically available for all readers to peruse, could engage in a similar practice only with the classifier obscured.[31]

Google's targeting might also suggest a discriminatory preference to ad recipients. For example, if gender is the sole attribute in a user's ad settings, the user might conclude that it is the feature on which job-related ads (and all others) are targeted to them. Even when an ad is delivered to everyone on the advertising platform, an ordinary user might perceive it to be targeted to their gender given the limited transparency they have into its full functioning. Admittedly, this may argue too much, but the standard focuses on the perception of the ordinary reader or listener. But again, the existing law addresses targeting that more subtly conveys a preference. Whether an online ad platform targeted on gender explicitly or on attributes that correlated to it, that targeting would skew who learned about an opportunity.

5.4. CDA §230 and Google's Ad Platform

Assuming then that holding an entity liable for targeting advertisements is holding them liable as a publisher where the entity at issue is an interactive computer service, Section 230 comes into play. Interactive computer services are protected from liability for content created by others. However, if an interactive computer service materially contributes to the development of discriminatory content they are treated like an "information content provider,"[32] and lose the protection §230 offers.[33]

Generally, entities are treated as an interactive computer service (ICS) if they provide "neutral tools" that others use to create discriminatory content. For example, Craigslist was protected against claims under the Fair Housing Act based on user-generated ads that violated the FHA because "[n]othing in the service Craigslist offers induces anyone to post any particular listing or

28. We note, but don't address, that the targeting criteria on its face expresses an age preference which is an independent violation of the Age Discrimination in Employment Act, which prohibits publishing an "advertisement indicating preference. . . based on age" (29 U.S.C.A. §623).
29. EEOC Compliance Manual Vol. 2, Sec. 632.2(a).
30. 24 C.F.R. §100.75.
31. The concept of steering under the FHA provides another way to articulate concerns with the outputs of online behavioral advertising systems. It addresses issues such as withholding information from certain groups of individuals.
32. Information content providers are defined as "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service."
33. See Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008).

express a preference for discrimination."[34] Similarly, where a defendant website provided an online questionnaire that was used to publish allegedly defamatory content, but left the selection of content exclusively to a third party, Section 230 provided immunity.[35]

In contrast, if the ICSes materially contribute to the discriminatory aspect of content, they are not protected by Section 230. Thus a website designed to match people with available housing was considered a content provider because it required each user to answer a series of questions disclosing his sex, sexual orientation and whether he would bring children to a household and compiled this information into a profile page that displayed the descriptions and preferences of users gleaned from the questions.[36] By forcing "subscribers to provide the information as a condition of accessing its service" and providing "a limited set of pre-populated answers," the website became "much more than a passive transmitter of information provided by others; it becomes the developer, at least in part, of that information." [37]

Courts have found that Google's ad serving platform meets the definition of an ICS.[38] With regard to Google's advertising platform, the question whether Google's actions go beyond those typical of an ICSand into those that would be associated with an information content provider is highly contextual. The AdWords platform is a black box mechanism that makes it difficult to identify who is responsible discriminatory outputs. Below we discuss four potential scenarios which reveal how potential legal liability shifts depending upon how targeting occurs.

5.4.1. Possible Causes of Ad Targeting

We now go through the possible causes of disparate ad targeting outlined in Section 4 and explore the legal ramifications of each of them.

(1) Targeting was fully a product of the advertiser selecting gender segmentation. In this scenario, Google is probably not creating or developing content. Instead, by allowing, but not requiring, advertisers to choose to target their ads to men or women, Google is providing a "neutral tool" that is protected by Section 230.[39] This tool allows third parties to determine who receives their ads, which is likely protected as a publisher function. Policies that allow advertisers to control who sees their ads are "precisely the sort of website policies and practices" to which "section 230(c)(1) extends."[40] In sum, in this scenario, the ads and who they target is information "provided [to Google] by another information content provider",[41] making Google not liable even if misused under a generally applicable provision like that of the FHA.[42]

(2) Targeting was fully a product of machine learning—Google alone selects gender. In this scenario, Google, and not the advertiser, is doing the targeting. Google, using programs that are part of its AdWords platform, decides who sees an ad based on Google's opinion of who is most likely to click on it. Advertisers are not part of the decision, and in fact they may be unaware that such a decision is being made.

34. Chicago Lawyers' Comm. for Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666 (7th Cir. 2008), as amended (May 2, 2008).
35. Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1124 (9th Cir. 2003).
36. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1164 (9th Cir. 2008).
37. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1166 (9th Cir. 2008).
38. Rosetta Stone Ltd. v. Google Inc., 732 F. Supp. 2d 628, 632 (E.D. Va. 2010) (claim against Google for unjust enrichment based on its practice of allowing trademarks to appear on its AdWords advertising platform was barred because in that context "Google is no more than an interactive computer service provider"); Goddard v. Google, Inc., 640 F. Supp. 2d 1193, 1198 (N.D. Cal. 2009) (allegations based on keywords in Google's AdWords advertisements were barred because Keyword Tool was a neutral tool permitted within the scope of CDA immunity).
39. See Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1121 (9th Cir. 2003); Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008).
40. Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 20 (1st Cir. 2016).
41. 47 USC §230(c)(1).
42. This conclusion is troubling in light of the increasing stratification of recipients that ad platforms, such as Google and Facebook, are able to achieve. For example, ProPublica reported in 2016 that Facebook provides advertisers with the option to exclude groups using "Ethnic Affinities," which included categories such as "African American" and "Asian American" (Angwin and Parris, 2016). Under current interpretations of Section 230, Facebook may avoid liability for providing these options if they are considered a neutral tool.

This is critical for understanding the application of Section 230 to an ad platform's activities.

Courts have adopted a "material contribution test to determine whether a website operator is 'responsible, in whole or in part, for the creation or development of"' unlawful information.[43] Under this test, a "material contribution to the alleged illegality of the content ... means being responsible for what makes the displayed content allegedly unlawful."[44] As a court in the fair housing context has put it, "[c]ausation in a statute such as § 3604(c) must refer to causing a particular statement to be made, or perhaps the discriminatory content of a statement."[45]

In this scenario, the "content" that would implicate Section 704(b) is the targeting of the advertisement. And if the advertiser has not selected for gender, Google decides to target the ads toward individuals based on whether they are men or women. This is different from providing a neutral tool such as allowing the advertiser the option to select for gender or suggesting keywords. Instead, Google alone is responsible for targeting certain employment ads toward men and away from women, and it is the targeting itself—irrespective of the content of the ad—that indicates a preference that would violate Section 704(b). As a result, Google is making a material contribution to the publishing enterprise. It is responsible, at least in part, for the creation or development of information and therefore would not be protected by Section 230. Because Google is likely outside the scope of Title VII, there is no risk of liability for the fact pattern in Datta et al's research. However, under the FHA, targeting that arose in this way would be the material basis for the illegality of an otherwise facially neutral ad.

(3) Targeting was fully a product of the advertiser selecting keywords. Courts have examined Google's keyword suggestions in the context of its Sponsored Link ad program and determined that it is a neutral tool that does not rise to the level of information creation or development.[46] In Goddard for example the court determined that the choice of keyword ultimately falls to the advertiser, and Google "does nothing more than provide options that advertisers may adopt or reject at their discretion."[47] That does not mean, however, that the "keyword tool" is per se neutral. In a situation where Google uses keywords to target ads (as opposed to serving them up in Sponsored Links), such targeting may rise to the level of a material contribution along the lines of Scenario (2) above. The specific keywords chosen and the role they play in targeting would be key in determining whether a material contribution had been made.

(4) Targeting was fully the product of the advertiser being outbid for women. Another possible situation is where a job advertisement does not reach women because other advertisers win the auctions for those ad placements. In this scenario, third parties are involved to some extent because they are selecting the price they are willing to pay for an ad placement. Nevertheless, Google would bear the same responsibility as if the bidding did not occur at all. That is because ultimately, the decision about where to place the ad is made by Google. The advertisers make a decision about how much they will pay, but they have no say over who finally sees an

43. Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 413 (6th Cir. 2014) (quoting 47 USC § 230(f)(3)).
44. Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 410 (6th Cir. 2014).
45. Chicago Lawyers' Comm. for Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666, 671 (7th Cir. 2008).
46. Goddard v. Google, Inc., 640 F. Supp. 2d 1193, 1197 (N.D. Cal. 2009); Jurin v. Google, Inc., 695 F. Supp. 2d 1117, 1119, 1123 (E.D. Cal. 2010). Rosetta Stone Ltd. v. Google Inc., 732 F. Supp. 2d 628, 630 (E.D. Va. 2010). But see 800-JR Cigar, Inc. v. GoTo.com Inc., 437 F. Supp. 2d 273 (D.N.J. 2006) (holding that a keyword tool was not entitled to Section 230 immunity "because the alleged fraud is the use of the trademark name in the bidding process, and not solely the information from third parties that appears on the search results page").
47. Goddard v. Google, Inc., 640 F. Supp. 2d 1193, 1198, 1199 (N.D. Cal. 2009). That "neutral tool" concept was used extensively in Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008) to determine whether a website has engaged in "development" that would negate §230 protection: "providing neutral tools to carry out what may be unlawful or illicit searches does not amount to development' for purposes of the immunity exception". "To be sure, the website provided neutral tools, which the anonymous dastard used to publish the libel, but the website did absolutely nothing to encourage the posting of defamatory content – indeed, the defamatory posting was contrary to the website's express policies."

ad. This is true even if an advertiser tried to target its employment ads toward women. Google is therefore still in the position of doing the targeting that makes the most material contribution to employment ads that express a preference for men by virtue of the fact that it decides to show more of these ads to men than to women.

Again, although there is no Title VII liability, under the FHA this could be a basis for liability, and existing case law suggests that Section 230 would not provide a barrier.

5.4.2. The Challenge of the Black Box

These hypothetical scenarios represent guesses at how the AdWords system might work. In fact, as discussed in depth in Section 4 above, it is likely that the results in the Datta et al. experiment arose from a combination of advertiser and platform choices. Identifying how various actors produced the results found by Datta et al. requires inside information that we do not possess. As a result, we are forced to make assumptions about whether liability would arise and how it would be apportioned.

Ultimately, given the parameters of the research in this case and the applicable statutes, it does not appear that any violations of law have occurred. Google would first have to fall under the coverage of Section 704(b), and, as currently drafted, that does not appear to be the case. However, in other circumstances we argue that the act of targeting itself could be considered a contribution to development of illegal content and under other statutes, specifically the FHA, could create a risk of liability.

6. Conclusions

Datta et al. (2015) demonstrate discriminatory outputs from Google's advertising system. Less clear is why or how it happened.

We have presented a selection of possible causes, but cannot without further access to Google's advertising system determine which is the actual cause. Analyzing potential legal liability under civil rights law requires an understanding of the entities covered by the law, as well as how discriminatory outputs arose.

Our analysis of existing case law concludes that Section 230 may not immunize advertising platforms from liability under the FHA for algorithmic targeting of advertisements that indicate a preference for or against a protected class. We argue that, in cases where an advertising platform, rather than the advertiser, makes the key decisions resulting in the ad being shown in different rates to members and non-members of a protected class, the ad platform becomes a co-developer of the ad, thereby losing its immunity. However, only some of possible targeting scenarios would satisfy this condition.

Although Section 230 poses the most obvious hurdle for holding online platforms such as Google liable, it turns out that on the facts of Datta et al's research the narrow scope of Title VII itself is a more formidable hurdle. By only applying to a tightly scoped set of employment-related entities, none of which Google appears to be acting as while serving ads, Title VII would not apply.

Our analysis reveals that advertisers covered by Title VII and the FHA using online algorithmically driven black-box advertising platforms face a dilemma: on the one hand they are bound to avoid advertising that infers a preference, but on the other, they cannot independently control the demographic makeup of those receiving their advertisements. The assumption has been that the advertising platforms which have the capacity to control the demographics of an advertising campaign were beyond the reach of antidiscrimination law due to Section 230's preclusion of holding interactive computer service providers liable for content created and developed by others. Our analysis reveals that Section 230 may not preclude liability in all instances. This is because targeting produced by platform algorithms that contributes to the illegality of an advertisement—its expression of a preference for or against a protected class—could be considered development under existing case law, thereby opening up the possibility of advertising platforms being found liable under the FHA. However, the inherent coverage limits in Title VII constrain the types of advertising platforms that might face liability (they would need to meet the statutory definition of an employment agency or other entity listed in Section 704(b)). Advertisers should be aware of the ways in which advertising platform algorithms can introduce bias into advertising campaigns, advertising platforms should provide ways to ensure advertisers can reach demographically diverse audiences where the law demands that they do so, and policymakers should consider whether the narrow scope of Title VII's advertising provision is fit for purpose in today's advertising ecosystem.

Acknowledgments

This paper has benefited from discussion at PLSC 2015 and the Unlocking the Black Box Conference, 2016. We gratefully acknowledge funding support from the National Science Foundation (Grants 1514509, 1704845, and 1704985). The opinions in this paper are those of the authors and do not necessarily reflect the opinions of any funding sponsor or the United States Government.

References

Julia Angwin and Terry Parris, Jr. Facebook lets advertisers exclude users by race. *ProPublica*, 2016.

Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias: There's software used across the country to predict future criminals. and its biased against blacks. *ProPublica*, May 2016.

Solon Barocas and Andrew Selbst. Big data's disparate impact. *California Law Review*, 104: 671, 2016.

Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in Neural Information Processing Systems*, pages 4349–4357, 2016.

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.

Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2015.

Tarleton Gillespie. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, 167, 2014.

Saikat Guha, Bin Cheng, and Paul Francis. Challenges in measuring online advertising systems. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, pages 81–87, New York, NY, USA, 2010.

Matthew Kay, Cynthia Matuszek, and Sean A. Munson. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3819–3828. ACM, 2015.

Pauline T. Kim. Data-driven discrimination at work. *Wm. & Mary L. Rev.*, 58:857, 2016.

Joshua A. Kroll, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. Accountable algorithms. *U. Pa. L. Rev.*, 165:633, 2016.

Anja Lambrecht and Catherine E. Tucker. Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads. *Social Science Research Network (SSRN)*, August 2017.

Latanya Sweeney. Discrimination in online ad delivery. *Commun. ACM*, 56(5):44–54, May 2013.

Deborah M. Todd. CMU researchers see disparity in targeted online job ads. *Pittsburgh Post-Gazette*, July 2015.

Catherine A Tremble. Wild westworld: The application of Section 230 of the Communications Decency Act to social networks' use of machine-learning algorithms. *Social Science Research Network (SSRN)*, 2017.

Michael Carl Tschantz, Amit Datta, Anupam Datta, and Jeannette M. Wing. A methodology for information flow experiments. In *Computer Security Foundations Symposium*, 2015.

Valentina Zarya. The Percentage of Female CEOs in the Fortune 500 Drops to 4%. *Fortune*, June 2016.