Algorithmic Bias? A study of data-based discrimination in the serving of ads in Social Media PRELIMINARY

Anja Lambrecht and Catherine Tucker

September 12, 2016

Abstract

The delivery of online ads has changed, so that rather than choosing to deliver advertising via a certain medium, instead within the same medium advertisers can choose which users their ads are shown to or allow an algorithm to pick the `right' users for their campaign. In this paper we show initial data that suggests this shift in optimizing delivery based on cost-e ectiveness can lead to outcomes consistent with apparent data-based discrimination. We show data from a eld test of a social media ad for STEM jobs that was explicitly intended to be gender-neutral in its delivery. We show that women were far less likely to be shown the ad, but not because they were less likely to click on it - if women ever saw the ad, they were more likely than men to click. We present evidence of the mechanism by which this apparent data-based discrimination occurs. The likelihood of showing ads to men rather than women does not re ect underlying measurements of gender equity such as labor participation rates or female education within the country. Instead, it re ects the fact that younger women are a prized demographic and as a consequence are more expensive to show ads to. This means that an ad algorithm which simply optimizes ad delivery to be cost-e ective, will deliver ads that were intended to be gender-neutral in what appears to be a discriminatory way, due to crowding out.

1 Introduction

Recently, the policy discussion of the potential for privacy harms of big data has shifted towards a discussion for the potential for data-based discrimination and in particular data-based discrimination in online advertising. Though the existence of outcomes that appear to be discriminatory have been documented (Sweeney, 2013; Datta et al., 2015), there have been few attempts to try to understand why ad algorithms can produce apparently discriminatory outcomes. This paper attempts to redress that gap.

visibility into the ad ecosystem, which includes Google, advertisers, websites, and users.' Our paper intends to be a first step at uncovering why ad algorithms may lead, in this case unintentionally, to outcomes which appear to be discriminatory.

The second literature is a literature on the delivery of ads by algorithm. There is a huge literature in computer science and machine learning devoted to better construction of such algorithms.[1] The actual study of algorithms in marketing has generally focused on the question of how to proceed when the underlying machinations of such algorithms may challenge causal inference (Johnson et al., 2015). Some work in marketing asks how traditional operations techniques, such as the lens of the multi-arm bandit problem, can help ad algorithms learn (Schwartz et al., 2016). Other work in marketing also documents when traditional ad algorithms can actually under-perform (Lambrecht and Tucker, 2013). Our paper, to our knowledge, is the first in marketing to evaluate the potential for ad algorithms to discriminate.

The third literature is a literature on discriminatory outcomes in marketing. The majority of this has documented discriminatory behavior in offline environments (Harris et al., 2005; Baker et al., 2005, 2008; Busse et al., 2016). Work on gender-based discrimination has focused

There are multiple policy implications of this paper. First and foremost, it highlights that occurrences of apparent data-based discrimination may neither be intentional nor reflective of underlying cultural prejudice. Instead, apparent data-based discrimination may simply reflect spillovers from the behavior of other advertisers. This means that regulators need to be cautious about assuming discrimination on the part of the platform or firm if there is the possibility that other people's behavior could explain an apparently discriminatory outcome.

Second, this phenomenon itself highlights an important insight about privacy online. Often privacy online is conceptualized as an individual right. However, the interconnectedness of data online and the potential for spillovers such as those documented in this paper highlight the extent to which issues in privacy online should be thought of through the lens of potential for spillovers, rather than restriction of the actions of a particular firm or platform independent of its effect on others in the ecosystem.

Third, there are questions about what should be done about such unintended discriminatory consequences of spillovers in ad algorithms. As shown by Dwork et al. (2011), ensuring algorithmic outcomes are `fair' can come into conflict with data privacy concerns as well as requiring human intervention. It also sheds lights on recent EU initiatives such as the push towards algorithmic transparency.[2] Our results highlight that algorithmic transparency may not be sufficient to prevent outcomes occurring that appear discriminatory. Without knowledge of how different actors behave whose behavior is governed by the algorithm, it is difficult to predict what may be the outcome of an algorithm that on its face of it looks reasonable and merely efficiency-maximizing.

Last, our results also have insights for advertisers who themselves wish to avoid their ads being shown in a way which may favor one demographic group over another unintentionally. There are a few reasonably easy steps to take. First, advertisers themselves should realize that in an ecosystem where other advertisers' advertising decisions can have implications for

[2] http://fusion.net/story/321178/european-union-right-to-algorithmic-explanation/

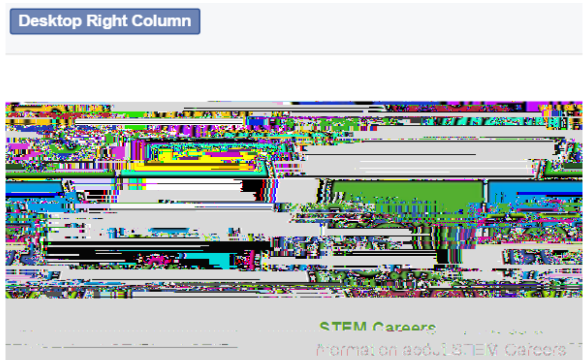

Figure 1: Sample Ad

to whom an ad is displayed, they may need to take additional verification steps to ensure that their campaigns are being shown equally to the groups they intend to show it to after the campaign is launched. Second, if advertisers are particularly concerned about striking a particular balance between age groups, genders or other common demographic groupings it may be worth separately constructing such campaigns, and adjusting bid values, rather than relying on an algorithm to allocate them.

2 Field Test

The field test that is the focus of the paper is very straightforward. We use the term `field test' rather than `field experiment' as there was no inherent randomization in ad delivery. Instead an ad was `tested' in 191 countries. We use the word `test' to reflect the fact that there was no strategy underpinning the selection of countries, ad format, or wording of the ad which could provide an alternative explanation of the results.

The field test was for an ad that promoted careers in STEM. The text of the ad was very simple; it said `Information about STEM careers' accompanied by a picture that represented the different fields in STEM. Figure 1 displays a mock-up of the ad.

The field test was conducted on a major social media platform. A separate ad campaign was created with an identical ad for 191 countries. We use this cross-national variation later in the paper to explore whether the differences in ad allocation we observe can be ascribed to

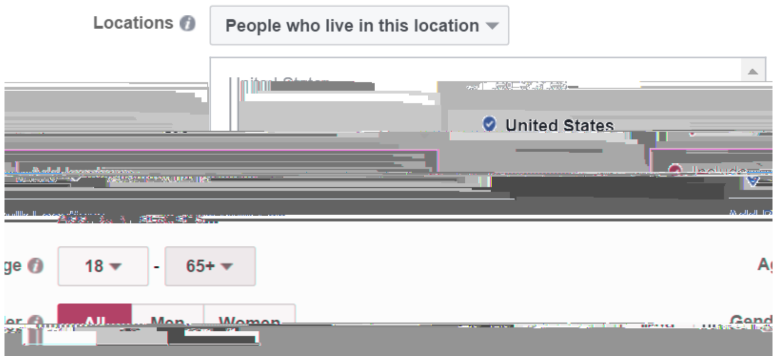

Figure 2: Ad Targeting Settings - Ad intended to be shown to both men and women aged 18-65.

different economic and cultural conditions regarding the role of women in different nations. In all cases the ad was targeted at both men and women between the ages of 18-65. The only variation for each of the 191 ad campaigns was the country it was targeted towards. Figure 2 displays the ad targeting settings for a typical ad.

The 191 countries were chosen to try and span the entire world. According to the United Nations, there are 195 countries. According to the social media platform, there are 213 countries and regions it marks as territories, such as American Samoa. The missing countries in our dataset are ones where the social media platform did not reach. For example, North Korea attempts to ensure that its citizens do not browse the broader web, meaning that it is not part of our dataset.[3]

For each country, the maximum bid for a click was set at $0.20. If after a week that campaign had not been viewed by 5,000 unique viewers, the bid was raised up to $0.60. Countries for which this occurred included Switzerland, the UK, the US and Canada. We account for any differences this time variation may have introduced in our regressions with time fixed effects.

	Mean	Std Dev	Min	Max
Impressions	1930.6	2288.7	1	24980
Clicks (All)	3.03	4.48	0	42
Unique Clicks (All)	2.81	4.11	0	40
CPC (All) (USD)	0.085	0.091	0	0.66
CPM (Cost per 1,000 Impressions) (USD)	0.18	0.32	0	4.33
Reach	621.6	815.8	1	11200
Frequency	4.33	4.29	1	53
Clicks Impressions	0.15	0.17	0	1.52
Clicks Reach	0.0063	0.013	0	0.25
Female	0.50	0.50	0	1
highgdp	0.50	0.50	0	1
High % Female labor part	0.50	0.50	0	1
High % Female primary	0.49	0.50	0	1
High % Female secondary	0.50	0.50	0	1
High Fertility Rate	0.50	0.50	0	1
High Female Equality Index (CPIA)	0.23	0.42	0	1
High % Internet Users	0.51	0.50	0	1

Table 1: Summary statistics

3 Data

For each of the 191 campaigns for each of the di erent countries, the social media platform released extensive data on their performance. This data is summarized in Table 1. We augmented this advertising data with data from the World Bank about each of the countries we had data for that pertain to the status of women and the female labor force in that country. This data was collected from the World Bank data repository for the most recent year that data was available.[4]

As shown in Table 1, in general bids for a click in each campaign were very low and were set to try and pay the minimum amount possible in that country for the ad to be shown to at least 5,000 social media platform users in that country. Figure 3 re ects the distribution of costs per click paid by the campaign. Relative to other studies of the cost of

[4] `http://data.worldbank.org/`

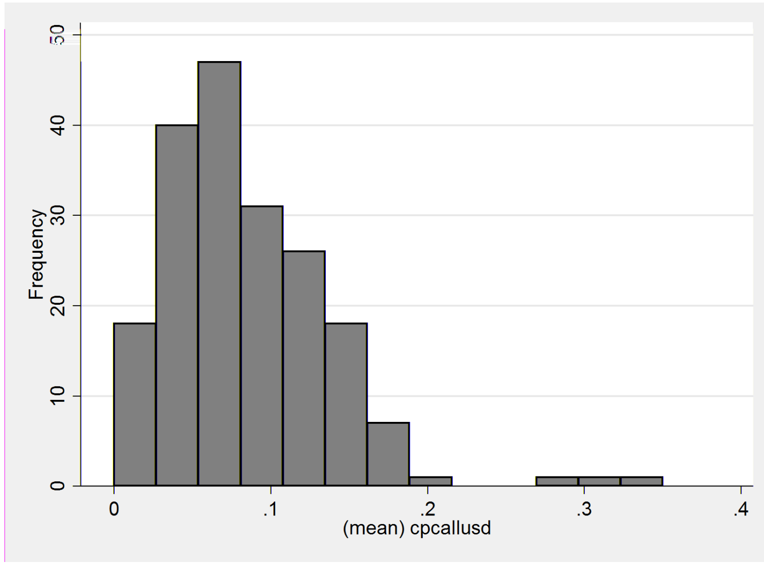

Figure 3: Histogram of average cost per country

social media campaigns, these click prices are obviously low (Tucker, 2014b,a). We discuss in detail the implications of this when we turn to the role of pricing in explaining the outcomes we observe.

3.1 Model Free Evidence

The main results of the eld test were visible even on the platform-supplied dashboard. Figure 4 supplies a screenshot of the dashboard.

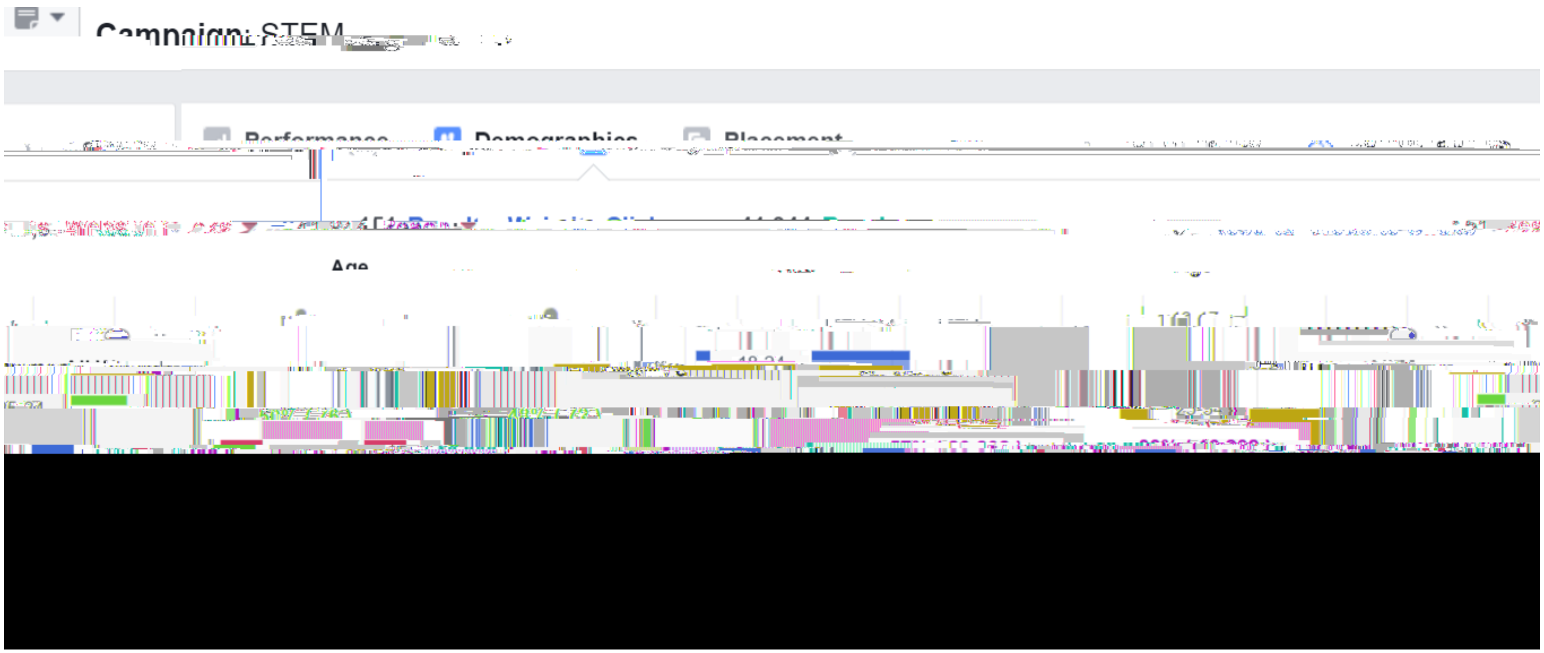

Figure 4: Main results of Field Test Visible on Dashboard

For readability, we also report these aggregate statistics in Table 2. Table 3 reports these aggregate statistics as an average at the country level. A comparison of Table 2 and 3 makes it clear that the pattern of impressions across di erent age groups is di erent at the aggregate level than at the average country level. This is because the larger countries where there were more impressions also tended to be the ones where the ad was shown more to younger people.

Table 2: Initial Dashboard Results reported as a Table

Age Group	Male Impr.	Female Impr.	Male Clicks	Female Clicks	Male ClickRate	Female ClickRate
Age18-24	746719	649590	1156	1171	.0015	.0018
Age25-34	662996	495996	873	758	.0013	.0015
Age35-44	412457	283596	501	480	.0012	.0017
Age45-54	307701	224809	413	414	.0013	.0018
Age55-64	209608	176454	320	363	.0015	.0021
Age 65+	192317	153470	307	321	.0016	.0021
Total	421966	330652	595	585	.0014	.0018

Table 3: Initial Dashboard Results in Table Reported as an Average per Country

Age Group	Male Impr.	Female Impr.	Male Clicks	Female Clicks	Male ClickRate	Female ClickRate
Age18-24	3909	3401	6	6	.0015	.0018
Age25-34	3471	2597	5	4	.0013	.0015
Age35-44	2159	1485	3	3	.0012	.0017
Age45-54	1611	1177	2	2	.0013	.0018
Age55-64	1097	924	2	2	.0015	.0021
Age 65+	1007	808	2	2	.0016	.0021
Total	2209	1732	3	3	.0014	.0018

Three immediate patterns in the data are obvious. First, men see more impressions of the ad than women. Second, the fact that men see more ads than women is particularly true in younger cohorts. Third, women and men click on ads in similar numbers. The rest of the paper is devoted to exploring the robustness of these empirical regularities and providing suggestive evidence about why they occur.

4 Results

4.1 Do men indeed see more STEM ads than women?

Though these empirical regularities may seem obvious in Table 4, we do check that our results are robust to a standard regression framework which allows us to control for di erent country characteristics.

For campaign i and demographic group j in country k on day t, the number of times an ad is displayed is modeled as a function of:

$$
\begin{aligned}
AdDisplay_{ijkt} = & \\
& + \; {}_1 Female_j \\
& + \; {}_2 Age_j \\
& + \; {}_3 Female_j \quad Age_j \\
& + \; {}_k + \; {}_{jk}
\end{aligned}
\qquad (1)
$$

Female

the fact that some groups may have had individuals who saw more than one ad on any one day. Columns (5)-(6) explore the effects of gender on ad frequency, that is, the number of ads any one individual saw. We find that conditional on seeing an ad, a woman is more likely to see it multiple times. This suggests that in general our measure of impressions may

Table 4: Women Are Shown Fewer Ads Than Men

		(1) Impressions	(2) Impressions	(3) Reach	(4) Reach	(5) Frequency	(6) Frequency
Female		-469.4	-205.3	-223.4	-96.90	0.715	1.253
		(94.21)	(43.23)	(34.00)	(19.95)	(0.147)	(0.300)
Female	Age18-24		-292.8		-229.5		-0.513
			(188.4)		(73.48)		(0.262)
Female	Age25-34		-651.1				

as a robustness check.

Let F denote the logistic likelihood function. Due to the aggregate nature of the social media platform data, which does not have user-level variables, all individuals i in demographic group j in country k have the same vector of x control variables. The likelihood of observing each observation of of

Table 5: If They See The Ad, Women Are More Likely To Click Than Men

	(1) Clicks (All)	(2) Unique Clicks (All)	(3) Clicks Impressions	(4) Clicks Reach	(5) Clicks (All)	(6) Unique Clicks (All)	(7) Clicks Impressions	(8) Clicks Reach
Female	0.221	0.303	0.0348	0.00273	0.264	0.399	0.0414	0.00359
	(0.0271)	(0.0290)	(0.00733)	(0.000590)	(0.0932)	(0.0875)	(0.0230)	(0.00173)
Female Age18-24					-0.137	-0.166	-0.0180	-0.00110
					(0.0975)	(0.0956)	(0.0264)	(0.00161)
Female Age25-34					-0.0899	-0.135	-0.0246	-0.00218
					(0.113)	(0.109)	(0.0286)	(0.00205)
Female Age35-44					0.0822	-0.0289	-0.0139	-0.00241
					(0.113)	(0.109)	(0.0269)	(0.00192)
Female Age45-54					0.0633	0.000689	-0.00464	-0.00176
					(0.119)	(0.117)	(0.0282)	(0.00175)
Female Age55-64					0.0465	-0.0573	0.0214	0.00231
					(0.136)	(0.129)	(0.0304)	(0.00216)
Age18-24	-0.175	-0.214	-0.0223	-0.00117	-0.105	-0.129	-0.0133	-0.000621
	(0.0576)	(0.0557)	(0.0136)	(0.000808)	(0.0731)	(0.0704)	(0.0150)	(0.000573)
Age25-34	-0.375	-0.460	-0.0482	-0.00265	-0.332	-0.394	-0.0359	-0.00156
	(0.0593)	(0.0572)	(0.0130)	(0.000836)	(0.0823)	(0.0785)	(0.0186)	(0.000671)
Age35-44	-0.341	-0.409	-0.0510	-0.00190	-0.379	-0.392	-0.0441	-0.000693
	(0.0712)	(0.0657)	(0.0133)	(0.000885)	(0.0902)	(0.0839)	(0.0172)	(0.00110)
Age45-54	-0.190	-0.222	-0.0290	-0.00165	-0.220	-0.220	-0.0267	-0.000769
	(0.0613)	(0.0605)	(0.0121)	(0.000847)	(0.0865)	(0.0843)	(0.0155)	(0.000666)
Age55-64	-0.0186	-0.0199	-0.00140	0.00146	-0.0426	0.00913	-0.0121	0.000305
	(0.0682)	(0.0666)	(0.0147)	(0.000893)	(0.0955)	(0.0879)	(0.0166)	(0.000846)
Country Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

equality. In all columns these interactions are insigni cant and the signs are inconsistent. This general lack of signi cance suggests that the particular cultural prejudices of the country towards women cannot explain the fact that more ads are being shown to men than women. Table A1 in the appendix shows that these results (or at least the general lack of measured signi cant e ects) hold for impressions as well.

Table 6: Women Being Exposed To Fewer Ads Than Men Is Not Driven Entirely By Underlying Gender Disparity In Labor Market Conditions In That Country

	(1) Reach	(2) Reach	(3) Reach
Female	-321.5	-253.4	-324.8
	(86.17)	(44.95)	(56.52)
Female High % Female labor part	61.78		
	(95.81)		
Female High % Female primary		-58.59	
		(95.25)	
Female High Female Equality Index (CPIA)			140.6
			(162.3)
Age18-24	1011.0	983.6	1057.3
	(145.4)	(144.8)	(150.5)
Age25-34	606.2	596.4	1181.9
	(95.13)	(94.50)	(106.1)
Age35-44	173.3	169.1	460.9
	(57.59)	(57.01)	(42.14)
Age45-54	63.04	54.88	150.9
	(44.01)	(43.33)	(32.05)

7 Do our results simply reflect competitive spillovers?

We now explore how competitive spillovers and pricing pressure for certain demographic groups may explain our results.

The firm bid for advertising impressions by specifying a maximum price it was willing to pay per click (CPC). This number was specific to a country and did not vary by age group or gender. Across all campaigns, the average cost per click was nearly identical for men and women (0.089 and 0.086), (t=.50). This by itself might seem to suggest that price itself does not play a role.

However, that still leaves the possibility that the budget caps and bid caps that the field test of the STEM ad deployed simply meant that the algorithm did not charge the advertiser the higher amount that would have been required to reach more women.

To explore this possibility, w(maxim)n-326(/l-37327(ted)fu6dh5269(r(fu606(w)au606(whAibilit)2dhlit)

	Mean	Std Dev	Min	Max
Avg Suggested Bid	0.60	1.16	0.010	37.8
Min Suggested Bid	0.30	0.53	0.010	6.69
Max Suggested Bid	0.95	1.45	0.017	43
Female	0.50	0.50	0	1

Table 7: Summary statistics

precisely interpret the economic implications of a price.

Note that this data also deviates from our original data in terms of age cohorts. In general, to avoid the restrictions on advertising to children inherent under COPPA and other privacy regulations designed to protect children, the eld test of the ad was not shown to anyone under the age of 18. However, we were able to collect pricing data on this group and use them as a baseline for the analysis. Furthermore, because in some countries there was too sparse a population of those who were 65+ for us to be able to get separate estimates, we combine the 55-64 and 65+ cohorts in this analysis.

7.1 Analysis of Secondary Pricing Data

Table 8 shows the results of our analysis of this secondary data. Columns (1) and (2) show that on average the platform suggests that advertisers bid 10 cents more to advertise to women. In terms of age, those in the 25-44 year old age group are also more expensive to advertise to. Columns (3) explores how this changes when we include interactions between gender and age. It shows strikingly that women between 25 and 45 are more expensive to advertise to than men, and this is particularly true for women aged 25-34. Columns (4)-(5) show that this result replicates if we look at the minimum or maximum suggested bid rather than the average. However, since there is large variation in the maximum bid as shown by Table 7, it is likely that Columns (1)-(3) are more reliable estimates.

We speculate that one reason behind this price premium may be that this group of women is traditionally a highly prized demographic for advertisers. Indeed, as stated by the business press, it is precisely this demographic of 25-34-year-old women which should be most prized by online advertisers, both because they are likely to engage with advertising and because they traditionally control household expenses.[6]

Therefore, a potential explanation behind our result is not that the ad-algorithm itself is discriminating actively against women or reflecting the local audience's cultural prejudices against women. Instead, it is reflecting spillovers from the behavior of other advertisers. As long as these other advertisers prize the `eyeballs' of young women, it means that any employment-related ad algorithm designed to allocate advertising impressions in a cost-effective manner will not display ads that are intended to be gender-neutral in a gender-neutral manner, but instead will favor cheaper male eyeballs.

8 Why are Women such a Prized Demographic?

The next question is why women are such a prized demographic that such crowding out occurs. To investigate this we use completely separate data from a large retailer that sold a broad range of fashionable consumer items that were largely intended to be decorative. It used social media advertising to try and generate demand for its one-day sales. It specifically divided its advertising campaigns so that it separately targeted men and women in different campaigns. We focus on the instances where the campaigns were identical in terms of product, behavioral targeting and wording.

This data is on the campaign level and include information on the number of impressions per campaign as well as the number of clicks and the number of instances when, upon arrival on the website, consumers added products to their shopping carts. Unlike our earlier data, this data is focused on the US.

[6] http://www.businessinsider.com/young-women-are-most-valuable-mobile-ad-demographic-2012-2

Table 8: In General, Women Are More Expensive To Advertise To On Social Media And The Competitive Spillover From Other Advertisers' Decisions May Explain Our Finding

	(1) Avg Suggested Bid	(2) Avg Suggested Bid	(3) Avg Suggested Bid	(4) Min Suggested Bid	(5) Max Suggested Bid
Female	0.112	0.112	-0.0464	-0.0130	-0.0155
	(0.0339)	(0.0329)	(0.0373)	(0.0288)	(0.0396)
Female Age18-24			0.0645	0.0226	-0.224
			(0.0372)	(0.0292)	(0.275)
Female Age25-34			0.258	0.0699	0.185
			(0.0890)	(0.0287)	(0.0497)
Female Age35-44			0.150	0.0609	0.177
			(0.0423)	(0.0291)	(0.0462)
Female Age45-54			0.0746	0.0193	0.119
			(0.0537)	(0.0397)	(0.0804)
Female Age55+			0.129	0.0476	0.190
			(0.0440)	(0.0342)	(0.0544)
Age18-24	-0.0102	-0.0102	-0.0420	-0.0438	0.335
	(0.0279)	(0.0271)	(0.0399)	(0.0303)	(0.276)
Age25-34	0.171	0.191	0.0419	0.00799	0.231
	(0.0445)	(0.0527)	(0.0397)	(0.0299)	(0.0524)
Age35-44	0.0738	0.0738	-0.000705	-0.0426	0.179
	(0.0359)	(0.0348)	(0.0438)	(0.0313)	(0.0582)
Age45-54	0.0587	0.0596	0.0217	-0.0220	0.235
	(0.0400)	(0.0389)	(0.0550)	(0.0373)	(0.0863)
Age55+	0.0194	0.0210	-0.0445	-0.0520	0.107
	(0.0343)	(0.0333)	(0.0429)	(0.0320)	(0.0556)
Country Controls	Yes	No	Yes	Yes	Yes
Observations	3048	3048	3048	2777	2776
Log-Likelihood	-3970.7	-4506.3	-3966.3	700.9	-3716.3
R-Squared	0.303	0.00897	0.305	0.718	0.492

Ordinary Least Squares Estimates. Dependent variable is average suggested bid in the Columns (1)-(3), minimum suggested bid in Column (4) and maximum suggested bid in Column (5). Omitted demographic groups are those aged between 13-17 and those of the male gender. Robust standard errors. * p < 0:05, ** p < 0:01, *** p < 0:001

8.1 Are Women of Higher Value to Advertisers?

We want to nd out whether women are indeed likely to be worth more than men to advertisers. Since the data is on the campaign level, we estimate an aggregate logit model. As before, our use of the aggregate logit model re ects the fact that ad performance is reported by grouping all successes and failures on each day without giving access to any information about an individual consumer. This means that while the consumer's decision is a binary

gender-neutral in its delivery. We show, though, that women were far less likely to be shown in terms of documenting not only the occasions when data-based discrimination may occur but also one of the likely (and unintentional) reasons why it occurs.

References

Athey, S. and D. Nekipelov (2010). A structural model of sponsored search advertising auctions. In *Sixth ad auctions workshop*, Volume 15.

Baker, S. M., J. W. Gentry, and T. L. Rittenburg (2005). Building understanding of the domain of consumer vulnerability. *Journal of Macromarketing 25*(2), 128{139.

Baker, T. L., T. Meyer, and J. D. Johnson (2008). Individual di erences in perceptions of service failure and recovery: the role of race and discriminatory bias. *Journal of the Academy of Marketing Science 36*(4), 552{564.

Busse, M. R., A. Israeli, and F. Zettelmeyer (2016). Repairing the damage: The e ect of price knowledge and gender on auto-repair price quotes*. *Journal of Marketing Research*.

Cheryan, S., J. O. Siy, M. Vichayapai, B. J. Drury, and S. Kim (2011). Do female and male role models who embody STEM stereotypes hinder womens anticipated success in STEM? *Social Psychological and Personality Science 2*(6), 656{664.

Datta, A., M. C. Tschantz, and A. Datta (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies 2015*(1), 92{112.

Dwork, C., M. Hardt, T. Pitassi, O. Reingold, and R. S. Zemel (2011). Fairness through awareness. *CoRR abs/1104.3913*.

Flath, D. and E. W. Leonard (1979). A comparison of two logit models in the analysis of qualitativemarket.unaddel.tSp88(ofSp88(ofSp88(of(e.)-mBdy)-272(STEM(ofSJ -179.67 eot)28Macr)50(omaacr)e6

Goldfarb, A. and C. Tucker (2011). Search engine advertising: Channel substitution when pricing ads to context. *Management Science 57*(3), 458{470.

Harris, A.-M. G., G. R. Henderson, and J. D. Williams (2005). Courting customers: Assessing consumer racial pro ling and other marketplace discrimination.

Tucker, C. (2014a). Social Advertising: Social Advertising: How Advertising that Explicitly Promotes Social In uence Can Back re. Mimeo, MIT.

Tucker, C. (2014b, October). Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research 51* (5), 546{562.

Table A1: Women Being Shown Fewer Ad Impressions Than Men Is Not Driven By Underlying Gender Disparity In Labor Market Conditions In That Country

	(1)	(2)	(3)