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-effectiveness can lead to outcomes consistent with apparent data-based discrimination. We show data from a field 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 reflect underlying measurements of gender equity such as labor participation rates or female education within the country. Instead, it reflects 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-effective, will deliver ads that were intended to be gender-neutral in what appears to be a discriminatory way, due to crowding out.

*Anja Lambrecht is Associate Professor of Marketing at London Business School. Catherine Tucker is the Sloan Distinguished Professor of Marketing at MIT Sloan School of Management, Cambridge, MA and Research Associate at the NBER. Thank you to NSF CAREER Award 6923256 for financial support. All errors are our own.

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.

We explore this question using data from a field test of an ad that was intended to promote job opportunities and training in STEM (Science, Technology, Engineering and Math). The ad was intended to be gender-neutral and was targeted neutrally. This ad was tested in 191 countries across the world. We show empirically, that the ad was shown to 20% more men than women. Since the lack of women in STEM fields is a much debated topic (Cheryan et al., 2011; Shapiro and Williams, 2012), such apparent algorithmic discrimination presents a potential policy concern. Therefore the rest of the paper is devoted to exploring why an ad intended to be gender-neutral was shown to more men than women.

A leading explanation is simply that women were less likely to click on the ad. This would lead an algorithm that was trying to maximize the probability of a click for each ad impression to show the ad to more men than women. However, we present evidence that this was not the case, and instead that if they were shown the ad, women were more likely to click on the ad than men. Another potential explanation is that women spend less time than men on the social media platform where the ad was displayed, meaning they are less likely to see the ads. We present evidence from industry reports that this is not the case.

We then explore an explanation based on the inherent sexism of the host culture. Perhaps the ad algorithm simply reflects underlying gender prejudices in the culture of the host country and the algorithm has learned over time to present ads in a way which reflects that bias. We use country-specific data from the World Bank to show that levels of female education or female labor market participation or general female inequality in that country cannot explain why the STEM ad was more likely to be shown to men than women.

We finally explore an alternative explanation based on spillovers from other advertisers' decisions. We present evidence from a separate data collection effort that on average across the world, female 'eyeballs' are more expensive than male eyeballs.

We present further evidence that this price premium is mainly driven by the female 25-34 demographic. We relate this to the fact that it was the female 25-34 demographic who in our test of the STEM ad were 40% less likely to see the ad. We also use additional data from an online retailer that suggests that the reason this demographic may be prized in general is that they are more likely, if they click on an ad, to actually purchase.

Our results suggest that behavior that is not intended to be discriminatory, such as young women being a demographic prized by advertisers, can lead to apparently discriminatory outcomes in a world where there are spillovers from one advertiser's valuation of eyeball to another advertiser's advertising campaign.

This paper contributes to three literatures.

The first literature is a nascent literature on data-based discrimination. Sweeney (2013) shows that a background check service's ads were more likely to appear in a paid search ad displayed after a search for names that are traditionally associated with African-Americans. Datta et al. (2015) find that women were less likely to see ads for an executive coaching service in India. In general, this literature has focused on documenting empirical regularities rather than understanding the underlying causes of the discriminatory outcomes. Indeed, methodological work in computer science into data-collection to investigate discrimination foregrounds the finding of empirical patterns consistent with discrimination, rather than understanding why disparate outcomes occurred (Ruggieri et al., 2010). For example Datta et al. (2015) states, 'We cannot determine who caused these findings due to our limited 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 on the portrayal of women in advertising content, especially across different international contexts (Lysonski and Pollay, 1990; Ford et al., 1998) and more recently on pricing (Busse et al., 2016). The closest paper to ours that we are aware of is Morton et al. (2003), who explore how the internet channel has affected the potentially discriminatory pricing of new cars to women and minorities, and who find smaller effects for women than minorities. Our contribution to this literature is that this is the first paper in marketing to explore how online algorithms, rather than simply the internet in general, affect potentially discriminatory outcomes.

[1]This paper does not seek to do justice to this literature, but instead refers the interested reader to the proceedings of the ACM conference on Electronic Commerce.

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

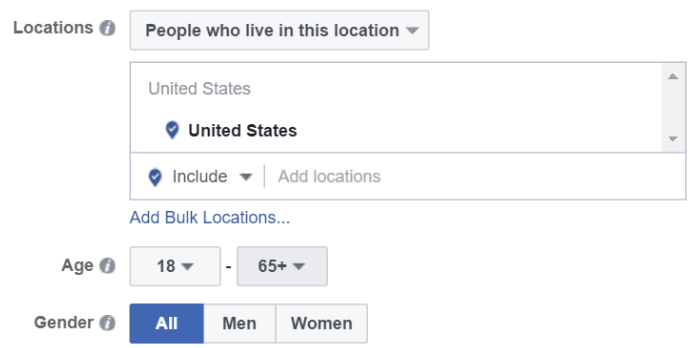

Figure 2: Ad Targeting Settings - Ad intended to be shown to both men and women aged 18-65.

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.

[3] `http://www.businessinsider.com/the-six-countries-that-block-social-media-2015-4` Though Turkey is sometimes mentioned as a country that does block social media and has in the past banned Twitter, we were still able to collect advertising data on it.

	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 different 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 reflects the distribution of costs per click paid by the campaign. Relative to other studies of the cost of

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

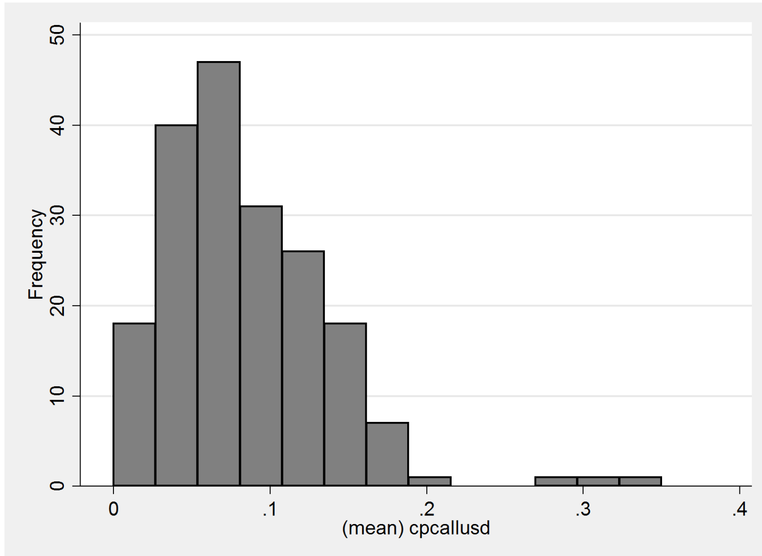

Figure 3: Histogram of average cost per country

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 field test were visible even on the platform-supplied dashboard. Figure 4 supplies a screenshot of the dashboard.

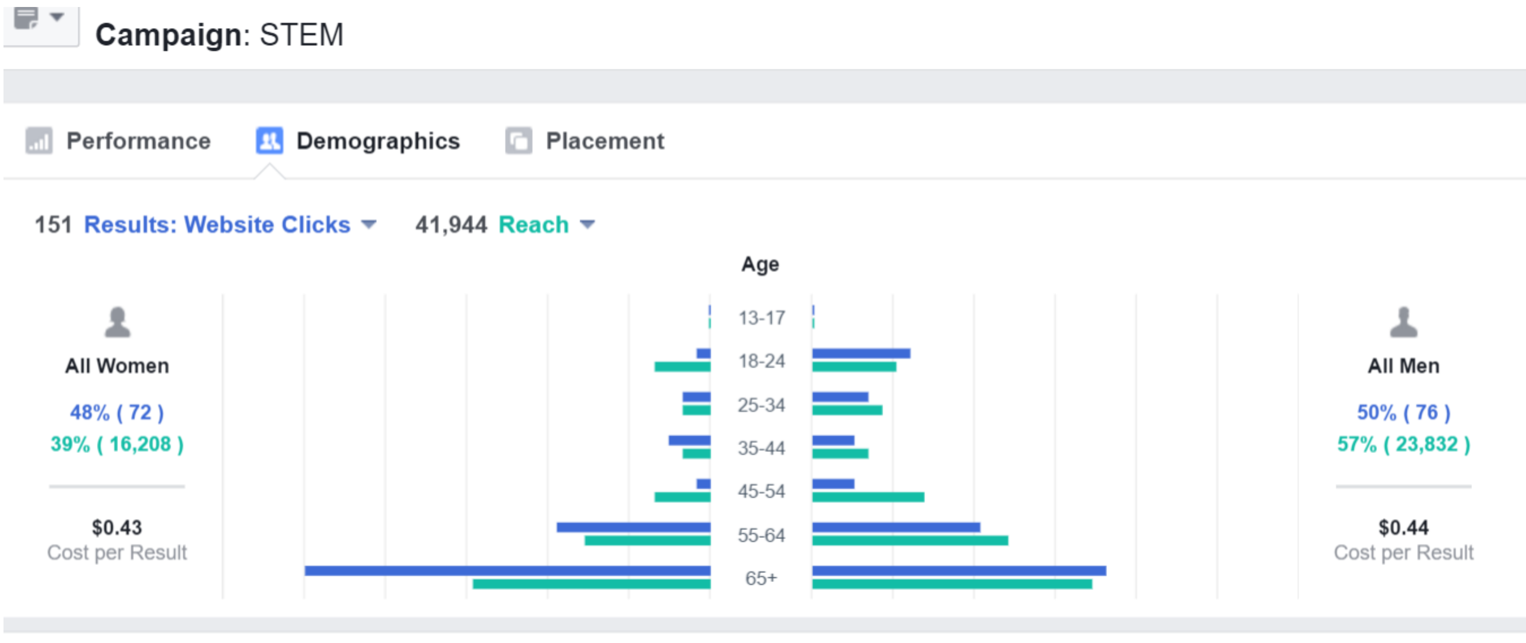

Figure 4: Main results of Field Test Visible on Dashboard

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 different age groups is different 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 different 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:

$$\begin{aligned} AdDisplay_{ijkt} = & \\ & + \beta_1 Female_j \\ & + \beta_2 Age_j \\ & + \beta_3 Female_j \times Age_j \\ & + \alpha_k + \epsilon_{jk} \end{aligned} \tag{1}$$

$Female_j$ is an indicator for whether or not this was a demographic group consisting of women. Age_j is a vector of fixed effects that capture the different age groups of the social media platform's users. We include a vector of country fixed effects α_k to capture variation in number of impressions due to country size and other country characteristics, such as technological sophistication.

Column (1) of Table 4 shows the results of a simple regression with no interactions. It suggests that women indeed, as suggested by Figure 4, were less likely to see the ad. Column (2) reports the full specification suggested by equation (1) and suggests that this unequal distribution of impressions is driven especially by women aged between 25-35 seeing the ad far less then comparable men.

Columns (3)-(4) replicates the results for reach (rather than impressions). This reflects 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 understate the extent to which women were not shown our ad, and therefore for the rest of the paper we focus on the ad reach or the number of unique people who saw our ad as the main dependent measure.

The coefficients for age suggest that fewer older people saw our ad at the country-observation level. This is despite the fact that, in aggregate, more older people saw our ad (Table 2). The reason for this discrepancy is that, as shown by Table 3, at the average-country level fewer older people saw our ad. This reflects the fact that smaller countries or countries also tended to have fewer older people using the social media platform.

5 Do our results reflect the fact that women were less likely to click on the ad?

We now turn to think about whether or not this divergence in impressions simply reflects an accurate prediction for the algorithm that women are less likely to click on ads. This seems a natural explanation of our results, given the prevalence of the use of 'quality scores' in ad algorithms. A quality score is a predictive method whereby an advertising algorithm is more likely to choose to show ads that will be clicked on (Athey and Nekipelov, 2010). The use of quality scores is profit-maximizing for the advertising platform in cases where advertisers pay per click, and consequently the advertising platform wants to make sure that it shows ads that are most likely to get clicks. Such a quality score could explain our findings if the underlying ad algorithm detected that women were less likely to click on our ad campaign, and consequently was less likely to show it.

Therefore, we next explore whether those who were shown the ad actually clicked. The

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***		-296.0***		-0.619*
		(150.6)		(46.96)		(0.266)
Female × Age35-44		-455.5***		-156.6***		-0.882***
		(107.7)		(30.13)		(0.241)
Female × Age45-54		-219.8**		-95.30***		-0.887**
		(67.95)		(23.77)		(0.294)
Female × Age55-64		35.28		18.50		-0.323
		(38.83)		(14.09)		(0.403)
Age18-24	2697.1***	2843.1***	890.8***	1005.4***	-0.464*	-0.208
	(244.8)	(278.0)	(106.1)	(127.4)	(0.203)	(0.170)
Age25-34	2088.6***	2413.7***	549.8***	697.6***	-0.672***	-0.364*
	(204.1)	(234.9)	(66.72)	(82.03)	(0.160)	(0.141)
Age35-44	901.5***	1128.9***	193.4***	271.5***	-0.545***	-0.105
	(116.5)	(133.7)	(40.06)	(46.50)	(0.142)	(0.163)
Age45-54	482.2***	591.7***	97.02**	144.5***	-0.462***	-0.0195
	(83.22)	(84.08)	(30.40)	(34.35)	(0.106)	(0.163)
Age55-64	106.7*	88.67	16.20	6.769	0.0119	0.173
	(50.22)	(51.46)	(18.51)	(19.26)	(0.178)	(0.144)
Country Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2339	2339	2339	2339	2339	2339
R-Squared	0.487	0.490	0.435	0.440	0.778	0.779

Ordinary Least Squares Estimates. Dependent variable as shown. Omitted demographic groups are those aged 65+ and men. Robust standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

social media platform reports data by grouping all connections, unique clicks, clicks, reach and impressions by country and age and gender group. This means that while a social media platform user makes a binary choice over whether to click, our data is aggregated across consumers, and we observe a number of successes (unique clicks) out of a number of trials (impressions) for each campaign-day. When we turn to think about differences in whether women actually click, we of course want to account for the discrepancy by gender in the number of impressions shown. As a result, we estimate an aggregate logit model using maximum likelihood (Flath and Leonard, 1979) as well as a simple linear clickthrough rate 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 the sum of positive unique clicks as a function of the sum of reach for that campaign that day is:

$$F(\beta x)^s \{1 - F(\beta x)\}^{r-s} \tag{2}$$

where s is the number of unique clicks and r is the population of social media platform users exposed to the messages.

Table 5 reports the result of our investigation of clickthroughs. Column (1) presents results of a simple specification for clicks as a function of impressions. It suggest that women are more likely to click on the ad. Column (2) repeats this but instead of using impressions it uses reach (that it the number of unique users exposed to a message) as the measure of population. Again, it suggests women are more likely to click on the ad. Column (3) and (4) show that our results replicate even when using a linear clickthrough rate estimated via ordinary least squares. We repeat this analysis with the same age and gender interactions that we used in Table 4 and suggested in equation (1). However, as shown by Columns (5)-(8), these interactions were not significant, indicating that click propensity did not differ by age group and gender. However, we do observe across most columns that in general younger people are less likely to click. In the final two columns where we use an OLS specification we lose significance, which may reflect the binary functional form of our data.

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
Observations	4515014	1453890	2339	2339	4515014	1453890	2339	2339
Log-Likelihood	-52298.6	-40388.3	1043.2	7364.8	-52291.8	-40384.6	1045.9	7372.0
R-Squared			0.168	0.315			0.170	0.319

Aggregate Logit Estimates in Columns (1)-(2) and (5)-(6). Ordinary Least Squares Estimates in Columns (3)-(4) and (7)-(8). In Columns (2), (4), (6) and (8) the population variable is ad reach. In Columns (1), (3), (5), and (7) the population variable is ad impressions. The dependent variable is whether someone who was exposed to an ad clicked. Omitted demographic groups are those aged 65+ and men. Robust standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.1 Do women spend less time on social media?

Another related explanation is not so much driven by the clickthrough rate but instead by differences in the amount of time women relative to men spend on social media platforms such as Facebook. However, industry reports suggest both that women are more likely to use social media platforms then men and also that they are more likely to spend more time on the site and consequently be exposed to ads.[5]

6 Do our results reflect cultural prejudice or labor market conditions for women?

Another potential explanation for our results is that the underlying ad algorithm has learned the preferences of that host country and knows that in a particular country it is undesirable to show ads to women. Therefore, our results could simply reflect the sad fact that in most countries, women's labor market rights and careers lag behind men's. To explore this alternative explanation, we use data from the World Bank that illuminates a variety of facets of the culture surrounding gender in that country. Specifically we use data on female labor market participation, female education, and also a female equality index (CPIA). This index is a World Bank-constructed measure that assesses the extent to which the country has installed institutions and programs to enforce laws and policies that promote equal access for men and women in education, health, the economy, and protection under law. A higher index implies a more equal outcome.

Table 6 displays the results of this investigation. In each case, we estimate how the number of females who saw an ad campaign in a country was moderated by whether or not that country scored highly on that measure of gender-equality. We measure this by a binary indicator variable for whether that country was above the median by that measure of gender

[5]See `https://www.brandwatch.com/2015/01/men-vs-women-active-social-media/` and `http://www.pewresearch.org/fact-tank/2013/09/12/its-a-womans-social-media-world/`

equality. In all columns these interactions are insignificant and the signs are inconsistent. This general lack of significance 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 significant effects) 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)
Age55-64	-12.69	-17.48	-42.40
	(26.67)	(26.24)	(27.98)
Country Controls	Yes	Yes	Yes
Observations	1536	1548	588
Log-Likelihood	-12208.6	-12301.4	-4485.8
R-Squared	0.409	0.414	0.601

Ordinary Least Squares Estimates. Dependent variable is whether someone is exposed to an ad. Omitted demographic groups are those aged 65+ and men. Robust standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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, we collected further data on what the social media platform advised the correct bidding strategy would be for each of the demographic groups. Specifically, we separately collected advertising pricing data for the specific social media platform that we ran the field experiment on for each of our countries. Table 7 presents summary statistics for this secondary data. The data include the average suggested bid as well as a minimum and maximum of the suggested bid range.

This data collection was distinct from the data for the field experiment. The additional data we collected is the unconstrained amount that the social media platform recommends that an advertiser should pay to reach a certain demographic group. This recommended bid data has been used in previous scholarship such as Goldfarb and Tucker (2011). Though such recommended bid data does have the disadvantage that researchers have no information about the precise 'black box' that is used to calculate the values, in this particular study this is less of a concern, as we are using it simply to proxy for the likely competitive bidding environment for a particular gender-age group within a country, rather than trying to

	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 field 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 find 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 reflects 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 choice, we have access only to data aggregated across consumers, and observe a number of successes (e.g. clicks) out of a number of trials (e.g. impressions) for each campaign-day.

Column (1) of Table 9 displays the results of an estimation that uses the number of clicks out of the total number of impressions per campaign as the dependent variable. We also have data on the age group of those to whom ads were displayed which we include as before and also interact with an indicator for gender. The omitted age category in our specification is those who are 45+. We likewise control for the products advertised as well as including fixed effects for the week and day of week. Column (1) suggests that indeed women between 18 and 24 are more likely to click on ads than men of the same age, though in general that age group is unlikely to click. However, it suggests that women in the 25-34 age group are not particularly likely to click on ads. This result contrasts with the results of Table 5 which suggests that women are more likely to click if they are shown a STEM ad, rather than, as is the case here, an ad for a consumer item.

However, we suggest that the price premium for young women may not stem simply from their propensity to click on ads but their ultimate profitability. Since advertisers ultimately care about sales more than about clicks, Column (2) displays a specification that uses as dependent variable the number of instances consumers added a product to their cart out of the total number of clicks. We use add to carts because the retailer tracks this as proxy for conversion rather actual purchases which are part of a separate billing system. We find that women in the 18-34 segment are more likely to convert, conditional on clicking. This makes them more valuable targets for advertisers, especially if a advertiser is paying for clicks rather than impressions as is common on many advertising platforms such as Facebook and Google. Column (3) looks at a similar specification that takes as the dependent variable the number of instances when products were added to shopping carts, out of the total number of impressions rather than clicks, and further confirms the results.

While we note that this is data from just one advertiser for a particular product category

Table 9: Younger women may be a valuable demographic as they appear more likely to convert conditional on clicking an ad

	Clicks out of impressions (1) Clicks	Add-to-cart out of clicks (2) Add to Cart	Add-to-cart out of impressions (3) Add to Cart
Female	-0.0522***	-0.0231	-0.0979
	(0.0152)	(0.186)	(0.185)
Age Group 18-24	-0.795***	-0.528	-1.392**
	(0.0379)	(0.558)	(0.548)
Age Group 25-35	-0.533***	-0.149	-0.742***
	(0.0194)	(0.265)	(0.264)
Age Group 35-44	-0.244***	-0.168	-0.430**
	(0.0155)	(0.202)	(0.201)
Female × Age Group 18-24	0.408***	1.078*	1.553***
	(0.0399)	(0.575)	(0.566)
Female × Age Group 25-35	-0.0602**	0.701**	0.709**
	(0.0272)	(0.326)	(0.324)
Female × Age Group 35-44	-0.000403	0.509*	0.508*
	(0.0220)	(0.264)	(0.263)
Week Controls	Yes	Yes	Yes
Day of week controls	Yes	Yes	Yes
Product Controls	Yes	Yes	Yes
Observations	127617816	67501	127605845
Log-Likelihood	-574304.1	-3339.4	-7802.1

Aggregate logit estimates. Dependent variable as listed. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Omitted demographic groups are men and those aged 45+.

which is consumer items, such as vases and decorative art, it does appear suggestive that we observe a pattern of increased conversion conditional of clicking on an ad which could explain why it is that advertisers treat younger women as such a highly prized demographic and potentially are willing to pay more to show ads to them.

9 Implications

The digital revolution has transformed how ads are delivered. Now, rather than an advertiser choosing a particular magazine or TV channel in which to place their ad, they can instead compete with other advertisers to have their ad shown to people on the same media. This paper investigates how this individualization of ad delivery can potentially have consequences in terms of unintended discriminatory outcomes.

Our analysis suggests this shift in optimizing delivery based on cost-effectiveness can lead to outcomes which are consistent with apparent data-based discrimination. We show data from a field test of an ad on social media for STEM jobs that was explicitly intended to be gender-neutral in its delivery. We show, though, that women were far less likely to be shown the ad. This is not because they were less likely to click on it - if women ever saw the ad, they were more likely than men to click. The likelihood of showing ads to men rather than women does not reflect World Bank measures of the culture of sexism within the country.

Instead, we present suggestive evidence that the apparent gender-discrimination reflects the fact that in the most advanced nations, women aged 18-35 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-effective, can deliver ads which are intended to be gender-neutral in what appears to be a discriminatory way. This suggests a nuanced view of the potential for apparently discriminatory outcomes even from neutral algorithms, and highlights the extent to which policymakers concerned about data and privacy online need to focus on the potential for unintended spillovers between players in an ecosystem rather than focusing on restricting the action of one firm or platform.

These results also emphasize that advertisers cannot rely on an algorithm to necessarily achieve neutrality across different demographic categories. Instead, to achieve a potentially neutral demographic allocation an advertiser may not only have to set different budgets for female and male advertising campaigns, but also further separate out bidding strategies by age as well as gender, to ensure that they do reach younger women.

There are of course limitations to our study. First, our field experiment consists of a single ad for STEM careers shown across multiple countries. Though it seems likely that our result would replicate across different ad designs, we do not have data to test this. Second, because we do not observe the workings of the actual ad algorithm, our result regarding the role of bidding decisions of other advertisers is suggestive rather than conclusive. Third, since our results are descriptive and focus on explaining an empirical regularity, we are not able to test or propose policy measures which may prevent the kind of outcomes we observe. Notwithstanding these limitations, we believe that this paper makes a useful contribution 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 differences 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 effect 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 qualitative marketing data. *Journal of Marketing Research 16*(4), pp. 533–538.

Ford, J. B., P. K. Voli, E. D. Honeycutt Jr, and S. L. Casey (1998). Gender role portrayals in Japanese advertising: A magazine content analysis. *Journal of Advertising 27*(1), 113–124.

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 profiling and other marketplace discrimination. *Journal of Public Policy & Marketing 24*(1), 163–171.

Johnson, G. A., R. A. Lewis, and E. I. Nubbemeyer (2015). Ghost ads: A revolution in measuring ad effectiveness. *Available at SSRN*.

Lambrecht, A. and C. Tucker (2013). When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research 50*(5), 561–576.

Lysonski, S. and R. W. Pollay (1990). Advertising sexism is forgiven, but not forgotten: Historical, cross-cultural and individual differences in criticism and purchase boycott intentions. *International Journal of Advertising 9*(4), 317–329.

Morton, F. S., F. Zettelmeyer, and J. Silva-Risso (2003). Consumer information and discrimination: Does the internet affect the pricing of new cars to women and minorities? *Quantitative Marketing and Economics 1*(1), 65–92.

Ruggieri, S., D. Pedreschi, and F. Turini (2010, May). Data mining for discrimination discovery. *ACM Trans. Knowl. Discov. Data 4*(2), 9:1–9:40.

Schwartz, E. M., E. Bradlow, and P. Fader (2016). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*.

Shapiro, J. R. and A. M. Williams (2012). The role of stereotype threats in undermining girls and womens performance and interest in stem fields. *Sex Roles 66*(3-4), 175–183.

Sweeney, L. (2013). Discrimination in online ad delivery. *ACMQueue 11*(3), 10.

Tucker, C. (2014a). Social Advertising: Social Advertising: How Advertising that Explicitly Promotes Social Influence Can Backfire. 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) Impressions	(2) Impressions	(3) Impressions
Female	-557.5**	-696.2***	-735.6**
	(194.0)	(140.5)	(249.0)
Female × High % Female labor part	-202.1		
	(241.5)		
Female × High % Female primary		145.5	
		(244.0)	
Female × High Female Equality Index (CPIA)			182.6
			(558.0)
Age18-24	2617.5***	2688.3***	3595.7***
	(304.7)	(320.4)	(440.1)
Age25-34	2044.9***	2145.2***	4419.8***
	(262.9)	(272.4)	(369.3)
Age35-44	732.4***	799.5***	1882.5***
	(140.6)	(150.9)	(224.2)
Age45-54	268.3**	290.9**	673.0***
	(90.74)	(96.08)	(150.4)
Age55-64	-31.10	-29.13	-94.77
	(58.27)	(59.08)	(100.9)
Country Controls	Yes	Yes	Yes
Observations	1536	1548	588
Log-Likelihood	-13604.4	-13751.7	-5180.0
R-Squared	0.439	0.465	0.640

Ordinary Least Squares Estimates. Dependent variable is whether someone sees an ad impression. Omitted demographic groups are those aged 65+ and men. Robust standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$