STEM and teens: An algorithm bias on a social media
PRELIMINARY VERSION, DO NOT QUOTE

Grazia Cecere*† Clara Jean* Fabrice Le Guel* Matthieu Manant*

January 22, 2018

Abstract

We study whether online platforms might reproduce offline stereotypes of girls in the STEM disciplines. The article contributes to work that aims to shed light on the possible bias generated by algorithms that replicate offline stereotypes. We run a field experiment based on launching an online ad campaign on a popular social media platform on behalf of a French computer science engineering school. We randomize the ad campaign in order to estimate whether a message aimed at prompting girls is more displayed to girls than to boys. The ad campaign targets students from randomly selected high schools in the Paris area. Our results show that on average girls received 24 fewer impressions than boys, but were more likely to click on the ad if they come across it. The treatment ad aimed at targeting more girls has a crowding-out effect caused by an ad that was shown less often to the whole audience. This bias is moderated for high schools with a large majority of girls on a science track.

Keywords: Gender-gap, discrimination, algorithm bias, STEM education.
JEL Codes: J16, I24

*RITM, University of Paris-Sud, University of Paris-Saclay
†KIND, Telecom Ecole de Management

1 Introduction

While the presence of women in higher education has increased over the years in the OECD countries, the numbers of women enrolled in science, technology, engineering, and mathematics (STEM) education programs continue to be significantly lower than the numbers of men (OECD, 2017). One explanation for this societal trend might be that women's educational choices possibly follow stereotypes (Bordalo *et al.*, 2016), or social norms (Akerlof and Kranton, 2010). Nevertheless, in a context of digitized information which allows easier access to information for all, we might expect a better matching between women and education programs, and particularly, increasing participation of women in STEM education. However, the growing use of data-driven algorithms on online platforms might also reproduce offline stereotypes, thereby offsetting the potential benefits of digitized information. We conduct a field experiment on a popular social media to study whether the social media algorithm reproduces stereotypes related to women in STEM fields.

To improve ad targeting on social media, firms employ machine learning algorithms (in most cases without human intervention) that use individuals personal information such as gender, age, and education, and users' interactions. However, use of intelligent algorithms can lead also to unanticipated correlations and may reproduce offline discrimination or stereotypes which generate negative spillovers (Tucker, 2017).

We set up an online ad campaign related to STEM education on a popular social media platform, targeting high school students, on behalf of a French computer science school. The aim of this field experiment was twofold. First, to measure whether, overall, algorithms distribute STEM ads equally to girls and boys. Second, to estimate how a girl-oriented message is distributed to girls and boys. We extend a previous experiment conducted by Lambrecht and Tucker (2017), and contribute additional findings on this topic.

The ad campaign ran for a two-week period, and was aimed at French high school students

aged between 16 and 19 years with accounts on the social network platform making no distinction between girls and boys. The computer science school's goal was to attract high school students generally. The ad provided information on job market access and average first wage after graduating from the computer science school. We wanted to study the behavior of the social media ad algorithm, and how the ad was distributed by gender and by high school. We conducted a field experiment that included 104 simultaneous ad campaigns, one for each high school targeted. In order to test our hypothesis about the effect of "girl" content on the distribution of the ad by the platform algorithm , we randomized the high schools into two groups. The ad images displayed were the same for both groups but the heading descriptions differed slightly: the treatment group received ads with headings that included "girl" content, while the ads displayed to the control group made no reference to gender content. Our results are robust to different specifications and control groups.

We contribute to three literature strands. First, we complement the literature on algorithmic bias. Recent research shows puzzling effects of the use of intelligent algorithms. On the one hand, algorithms can improve ad effectiveness (Kleinberg *et al.*, 2017), on the other hand they can reproduce offline stereotypes (Lambrecht and Tucker, 2017).[1] We add to the findings on the latter effect by testing the effect of the ad content if it includes reference to gender. Second, a large number of approaches have been developed to study the gender gap in STEM education (e.g. Fryer and Levitt, 2010; Croson and Gneezy, 2009; Guiso *et al.*, 2008). We are interested in how social media indirectly might influence participation of girls in STEM education. Third, we investigate the relative importance of personal data to improve the matching between individuals and services. The literature on privacy economics highlights the importance for Internet companies of exploiting personal data to target consumers but our understanding of the potential spillovers from this activity is limited (Acquisti *et al.*, 2016).

[1]The literature describes these processes as 'feature engineering'. See Datta *et al.* (2014) and OECD (2017) for more information.

Our paper is related to Lambrecht and Tucker (2017) but differs from it in several ways. First, we focus on teenagers – specifically high school pupils aged between 16 and 19 – and whether the social media ad algorithm discriminates among them during ad distribution. Second, we use a randomization process to check whether minor manipulation of the ad content that is addressed to girls (rather than boys) might affect the behavior of the ad algorithm. Third, we run the ad campaigns at the high school level in order to match experimental data with administrative data. Fourth, in order to reach the largest number of individuals, the cost paid for the experiment was based on the cost per thousand impressions (CPM) rather than the cost per click.

Policymakers and education institutions strive to spark girls interest in STEM education, and eventually STEM jobs, and the dissemination to teenagers of information on STEM is an important issue. More generally, policymakers need a better understanding of the extent to which gender or racial biases might be driven by online platforms use of algorithms. Use of digital solutions might be hindering policies aimed at reducing stereotypes.

The remainder of the paper is organized as follows. The next section provides a review of the relevant literature and is followed by a description of our research design. Succeeding sections describe the data set and empirical methods, and present our main results. The final two sections provide a discussion and offer some conclusions.

2 Literature review

There is a large literature on the gender gap in STEM education. Here, we are interested in the role of data-driven social media algorithms. To our knowledge, the link between algorithm bias resulting from machine learning algorithms and the persistence of the gender gap has not been investigated in much depth.

Literature on machine learning algorithms

The empirical literature on machine learning algorithms relies largely on field tests and experiments to analyze algorithm behaviors. A growing strand of this literature is aimed at justifying the role and effectiveness of these tools. For instance, Stitelman *et al.* (2011) conducted an experiment during a marketing campaign for a fast-food company, and found a positive effect of using machine learning for identifying potential new consumers. From a different perspective, Kleinberg *et al.* (2017) study the use of a calibrated and trained algorithm in the context of legal court decisions. They show that machine learning algorithms can help reducing criminality of about 25%, and generate much more fair decisions in favor of afro-American and Hispanics compared to human decisions.

From yet another perspective, the literature investigates bias produced by algorithms on online platform, and attempts to provide explanations for it. For example, Sweeney (2013) highlights that algorithms with machine learning capability might express unintentional biases linked to individual sociodemographic data. She focuses on ad displays based on an ethnicity inferred by the algorithm. Performing Google searches, she observes that compared to female white-identifying names, female black-identifying names receive more displays of an ad for criminal records services. Following this study, others have confirmed the existence of such biases (see e.g. Datta *et al.*, 2014; Angwin, 2016). O'Neil (2016) argues that algorithms might generate biases because they are trained with biased data. In particular, most algorithms can reproduce discrimination or stereotypes observed and learned from individual data. Lambrecht and Tucker (2017) investigate and try to explain these biases. They ran a country level field test on a social media with a gender-neutral ad for STEM jobs. Their results suggest apparent discriminatory outcomes to the detriment of women. They underline two possible causes for this bias. First, women are valued more highly, suggesting more expensive "eyeballs". Second, there are spillovers associated to targeting by other advertisers, which might incite the algorithm to reproduce gender biases against women. We

rely on this work in conducting our experiment, which provides slightly different findings.

Overall, the results in the literature on algorithm biases are ambivalent. Some studies provide evidence of apparent discriminatory outcomes, which cannot be explained using an economic rationale, while others find positive effects due to better matching between services and individuals, and fairer decisions toward certain categories of persons.

Literature on gender gap in education

A large literature in economics attempts to explain why a gender gap issue in STEM education occurs, and how it can be reduced. From a social point of view, policy makers' goal is to improve men / women diversity, which should lead to better social productivity. Fryer and Levitt (2010) investigate how gender gap in mathematics rises at kindergarten. They show that there is no gender gap between boys and girls upon entry to kindergarten, but the gender gap rises over the first six years of school, where girls do significantly worse than boys on math tests. The literature has also investigated the link between this gap on educational choices by teens, how it shapes their careers choices, and largely impact the labor market (Chetty *et al.*, 2011; Leibbrandt and List, 2015).

The literature identifies children's environment as an explanation for gender disparities in test scores. Using data from the international survey PISA, Guiso *et al.* (2008) highlight the importance of cultural beliefs about girls performances, and their effects on both math and reading tests. They show that the math gender gap disappears in countries with a more gender-equal culture. Nollenberger *et al.* (2016) find similar conclusions, underlining that cultural beliefs about women's role in society has a significant impact on the math gender gap. Pope and Sydnor (2010) show that gender differences are accentuated in areas with lower median income, less educated adult population, and high index of stereotypes adherence. Carrell *et al.* (2010) find that, while female students perform worse than males in the same introductory math and science courses, the gap is lower when female students

are taught by female faculty members, and even eliminated if female students handle high math abilities. This effect is minimal for male students. There is therefore an "homophily" effect that impacts girls' performances, as if girls identify their peers as a model, and want, in return, to investigate this field.

School quality is also partly responsible for gender gap in educational and behavioral outcomes. Especially, Autor *et al.* (2016) observe a strong and positive relationship between school quality and students' academic achievement both for girls and boys, but a major influence of the school quality on boys performance, where lower performing schools bring boys to perform worse than girls.

Although there is a large empirical literature studying gender gap in STEM disciplines, research on the mediating role of social media is still limited. Our paper contributes to this literature by bringing attention toward implications of the use of intelligent algorithms on this topic.

Literature on economics of privacy

The use of personal data to better target consumers is essential for Internet companies, but little is known about the possible associated spillovers (Acquisti *et al.*, 2016). Individuals disclose their personal data in order to obtain immediate access to services or to get in contact with peers. These data can be used by algorithms to target particular groups of consumers, and might generate unexpected negative spillovers. In particular, since data are likely to persist in digital identities, data disclosed at one point in time can be used in the future (Tucker, 2017).

Goldfarb and Tucker (2011) show that limiting the data collection capabilities of firms can induce negative effects traduced by less effective advertising campaigns. However, using algorithm to collect personal data is raising some concerns. The main worry is that an algorithm might reinforce certain individuals traits and induce a lock-in situation whereby

the individual is categorized by the algorithm and means that his or her past behavior on a platform will influence the information they are able to access in the future.

Our paper contributes to this literature by highlighting that the personal information disclosed can influence how the algorithm categorizes the individual and the type of information he or she can access.

3 Research design

We want to understand the role of a social media algorithm in ad displays, and more particularly, how an ad message affects the ads distribution to girls and to boys. We designed a field experiment with simultaneous ad campaigns distinguishing between neutral ads (addressed to the control group) and ads with "girl" content (addressed to the treatment group). We hypothesize that the algorithm takes account of the "girl" content and displays more ads to girls than to boys. We first outline the French education system, and then describe our experimental design and the data used.

3.1 Context

To access post-secondary education in France, 12th-grade students are required to register on a government platform to state their education preferences. In 2017, the closing date for this was March 21st ; we ran our field experiment from March 11th to March 26th, 2017. The aim of the engineering school was to encourage the enrollment of new students, and especially girls who are under-represented in STEM disciplines. Our experiment targeted French teenagers (aged 16 to 19) enrolled in high schools in the Paris area.

High school system in France

The French educational system comprises three stages. Most children attend primary school

(6 to 10 years) for five years, spend four years in middle school (age 11 to 14), and three years in high school (15 to 17 years). Education is compulsory up to age 16.

Students enrolled in high school attend for three years before graduating. At the end of the first year (11th grade), they choose a specialized track which they follow for the remaining two years in high school. In general education, the choice is among three main tracks: Literary (L), Economics and Social sciences (ES), and Science (S).[2]

3.2 Experimental design

We ran our ad campaign over a two-week period on a social network platform, targeting French high school students aged between 16 and 19. We target both girls and boys without distinction. The field experiment involves 104 different ad campaigns, one for each high school targeted.[3] The engineering schools goal was to attract high school students via these campaigns. Our interest was in the behavior of the social media ad algorithm. The ad provided information on job market access, and average first job wage after graduating. We randomized two groups of high schools. The ad provided information about job market access, and average first job wage after graduating from the school. We created two random groups of high schools treatment and control - which received the same pictures but slightly different text. The text presented to the control group was: «100% of occupational integration (and) €41,400, average annual gross salary» (Fig. 1), the text presented to the treatment group was: «100% of occupational integration (and) €41,400, average annual gross salary *for women*» (Fig. 2).(see ad designs in Fig. 1 and Fig. 2). [4] We set a daily budget of €2 for each high school.[5]

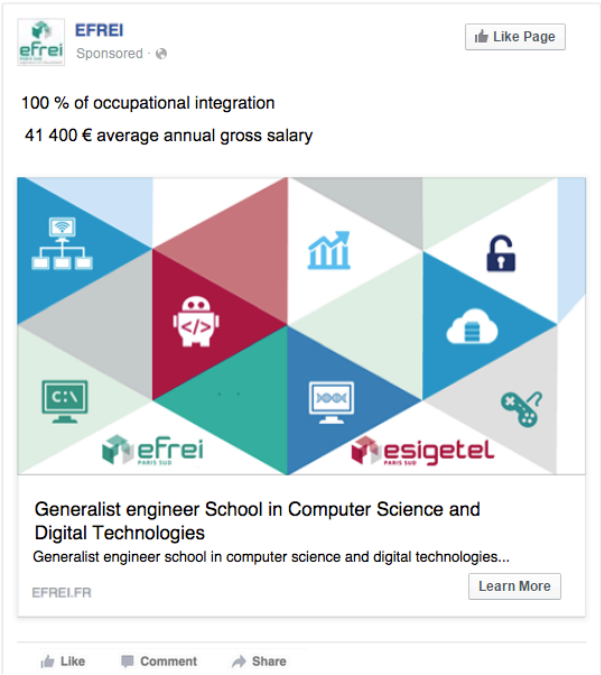
[2]Students may also choose a vocational track such as Management science and technologies track (STMG), Biotechnologies and physics in laboratory track (STL), Music and dance techniques track (TMD), or others.

[3]Our sample includes 38 high schools located in Paris and 66 in the suburbs; 27 are private high schools and 77 are public high schools.

[4]The original French ads are available on request.

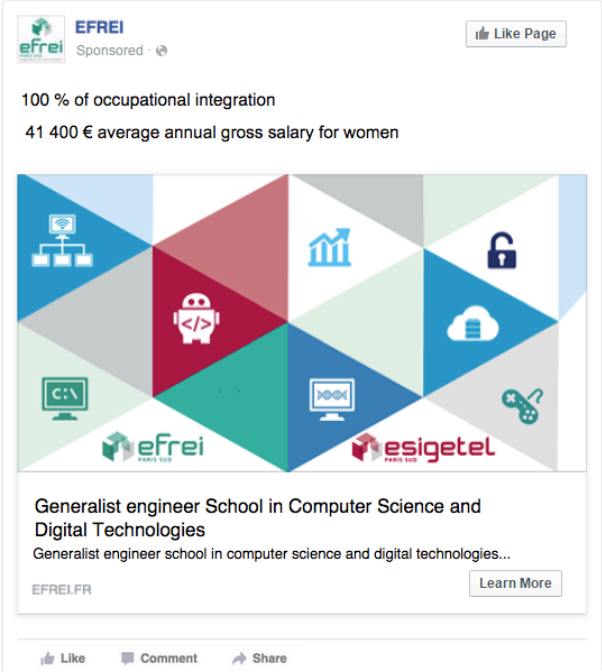
[5]For each campaign, we spent an amount lower than our budget, with an overall average cost per thousand impressions of €1.54.

The experiment was conducted over 14 days within the period of post-secondary education registration, and involved a gender neutral ad design. We had two objectives. First, we were interested in how ad is displayed between girls and boys (control ad). Second, we were interested in the behavior of the algorithm if prompted to target more girls (treatment ad).



Notes: Translation of the neutral ad we presented to high school students in the control group

Figure 1: Control ad



Notes: Translation of the ad with "girl' content we presented to high school students in the treatment group

Figure 2: Treatment ad

3.3 Randomization procedure

We get high school administrative data from Etalab, a French national project that provides detailed information for each high school in France[6]. As, the engineer school used to target students from Paris area, we restricted our sample to high schools in this area with a social media page[7]. This produced a sample of 104 high schools, corresponding to 79,955 students.

[6]Data are available upon request from Etalab. They correspond to the academic year 2015-2016. For more information about Etalab data, see https://www.data.gouv.fr/fr/search/?q=lyc%C3%A9es, last retrieved January 12th, 2018.

[7]The total number of high schools in Paris area is **465**.

Table 1 presents the summary statistics of high school data. We have information on total number of students enrolled, number of girls enrolled, number of students enrolled in each track, number of girls in each track, graduation rate, and high school status (public or private).[8] In our sample, girls are over-represented in literary track, and under-represented in the science track (44.4%) which is representative of national trends.

Table 1: Overall sample: high school data

Variable	**Mean**	**Std. Dev.**	**N**
Proportion of girls overall	0.523	0.087	103
L track	0.772	0.080	86
ES track	0.589	0.090	94
S track	0.444	0.096	97
Graduation rate			
L track	0.729	0.389	104
ES track	0.827	0.305	104
S track	0.838	0.269	104
Enrollment per high school (proportion)			
10th-grade	0.364	0.051	104
11th-grade	0.320	0.028	104
12th-grade	0.316	0.032	104
L track	0.112	0.091	104
ES track	0.240	0.112	104
S track	0.390	0.184	104
Other tracks	0.258	0.275	104
Paris downtown[6]	0.365	0.484	104
Public schools[6]	0.260	0.441	104

Notes: This table reports overall mean estimates for the 104 high schools in our sample. We abbreviate each track by the first letter of their name: L stands for Literary, ES for Economic and Social sciences, and S for Sciences. Girls in L (resp. ES, and S) track represents the proportion of girls enrolled in literary (resp. economic ans social sciences, and sciences) tracks for all high schools. Graduation rate is the proportion of students who passed the final exam for all tracks. Enrollment per high school in 10th-grade (resp. 11th-grade and 11th-grade) is the average number of 10th-grade (resp. 11th-grade and 12th-grade) students enrolled in each high school. Enrollment per high school in L (resp. ES and S) track is the average proportion of students enrolled in L (resp. ES and S) tracks. Variables indicated by [6] are dummy variables.

Randomization procedure

Our choice to prompt the algorithm to target more girls implied that before commencement of the ad campaign**s**, we needed to randomize the high schools to allow comparison of two

[8] Our sample includes 38 high schools located in Paris and 66 in the suburbs; 27 are private high schools and 77 are public high schools.

similar groups, and to minimize selection bias. We randomly assigned each of the high schools represented in the social network to the control or the treatment groups. Table 2 presents the pre-treatment statistics, there were no statistical differences between the two groups.

Our randomization process produced two groups: 53 high schools in the control group, and 51 high schools in the treatment group. Table 2 tests regressor balance by providing means for all pre-treatment variables by high schools which ensures we have two statistically identical groups of high schools.

3.4 Data description

The social platform provides experimental data which includes detailed information collected over a two-week period on the ad campaigns performance. The unit of analysis is the high school, by gender and by age group. The dataset includes two age categories (16-17 and 18-19), and for each category data are displayed for boys and girls. Ad performance includes five main variables: *Impressions, Reach, Frequency, Clicks*, and *CPM.*

Impressions refers to the number of ad display while *Reach* refers to the number of different individuals who saw the ad. Since a single individual could see the ad several times, thus *Frequency* measures the average number of times the individual saw the ad . We introduced a dummy variable *Click ad* to distinguish between high schools whose students clicked or not on the ad. Finally, *CPM*[9] indicates how much we spent on impressions per high school per day.

[9]Cost per thousand impressions

4 Results

4.1 Ad campaign performance

During the ad campaigns, we reached a total of 253,865 teens, to whom ads were displayed 1,226,92 times.[10] In this data set, each observation is the high school performance by gender and age category (16-17 and 18-19). The total sample size includes 5,484 observations. Although we budgeted for €2 per day per high school, we paid €1.564 per day on average as the ad system of the social network platform is based on a second price auction. For impressions, we recorded a very low number of clicks, with only 24% of high school with at least one click which is in line with the results in the literature. [11]

On average, the algorithm displays 224 impressions per day to each high school, and the average number of reaches is 46 individuals. Table 3 presents the overall performance of the relevant variables.

4.2 Are girls and boys treated equally?

Table 4 presents a summary of the average ad performance measures by gender, and also presents the t-stat to measure the statistical difference in these indicators between girls and boys.

On average, girls received fewer impressions than boys (209 vs 239), and benefited from fewer ad repetitions (4 vs almost 5). Similarly, the difference in CPM is not significant suggesting no cost differences between girls and boys. Thus, the difference in number of impressions between girls and boys cannot be attributed to the difference in price.

Table 5 depicts the CPM per each gender and age category. Across all 104 campaigns, the

[10] A single teenager could be reached several times during the period of the experiment, and have been recorded several times.

[11] Chatterjee *et al.* (2003) and Lambrecht and Tucker (2017) find also an individuals' low click probability.

Table 2: Pre-treatment summary statistics

Variable	**Control**		**Treatment**		**p-value**
	Mean	Std. Dev.	Mean	Std. Dev.	
Proportion of girls overall	0.513	(0.079)	0.534	(0.095)	0.234
S track	0.439	(0.087)	0.4450	(0.106)	0.582
ES track	0.581	(0.090)	0.599	(0.090)	0.343
L track	0.774	(0.079)	0.769	(0.081)	0.779
Graduation rate overall	0.926	(0.067)	0.923	(0.101)	0.862
S track	0.884	(0.199)	0.791	(0.321)	0.077
ES track	0.842	(0.285)	0.811	(0.327)	0.614
L track	0.732	(0.388)	0.725	(0.394)	0.926
Enrollment per high school[12]	804.585	(392.836)	731.608	(343.888)	0.317
12th-grade	0.315	(0.031)	0.319	(0.034)	0.565
11th-grade	0.321	(0.026)	0.318	(0.030)	0.635
10th-grade	0.364	(0.050)	0.363	(0.054)	0.916
S track	0.433	(0.173)	0.345	(0.185)	0.0135
ES track	0.244	(0.111)	0.236	(0.114)	0.697
L track	0.104	(0.071)	0.120	(0.109)	0.352
Other tracks	0.219	(0.233)	0.299	(0.310)	0.138
Paris downtown[6]	0.380	(0.489)	0.350	(0.483)	0.798
Public schools[6]	0.208	(0.409)	0.314	(0.469)	0.221
Sample size	53	51			

Notes: This table reports mean estimates for variables in our data set for both treatment and control group. Standard deviations are indicated in parentheses. We abbreviate each track by the first letter of their entitled. L stands for Literary, ES for Economic and Social sciences, and S for Sciences. Statistics for enrollment by grade and series are in proportion. Variables indicated by [6] are dummy variables.

Table 3: Social media data: raw data

Variable	Mean	Std dev.
Impressions	223.729	237.601
Reach	46.292	37.337
Frequency	4.432	4.231
Female users[6]	0.504	0.500
Age 18-19[6]	0.509	0.500
Treated high schools[6]	0.473	0.499
Click ad[6]	0.242	0.429
Sample size	5,484	

Notes: This table reports overall mean of the row data collected during the experiment. Variables indicated by [6] are dummy variables.

Table 4: Ad display according to gender

	Mean	Std Dev.	t-stat	p-value
Impressions			4.602	0.000
Boys	238.591	245.434		
Girls	209.114	228.742		
Reach			0.299	0.765
Boys	46.444	37.569		
Girls	46.143	37.113		
Frequency			7.196	0.000
Boys	4.845	5.275		
Girls	4.027	2.798		
Click ad			1.265	0.206
Boys	0.250	0.433		
Girls	0.235	0.424		
CPM			-0.396	0.692
Boys	0.387	0.275		
Girls	0.390	0.287		

Notes: This table reports ttest result conducted on experimental data to identify differences between boys and girls. All of our variables have a number of observations equal to 5 484 (N= 5,484). We notice a significant difference between boys and girls in impressions and frequency, with a lower ad display and frequency toward girls.

average price cost per CPM was €0.231 for boys aged between 16-17, and €0.222 for girls in the same age category. The average CPM for 18-19 year olds was €0.534 for boys and €0.553 for girls.

4.3 Graphical evidence

Figure 3 displays the average number of impressions by gender and age group and shows that the number of impressions displayed is always lower for girls compared to boys regardless of age or treatment group. We observe that this gap tends to decrease in the treatment group although this group seems to experience a crowding-out effect with fewer impressions for all age groups and genders.

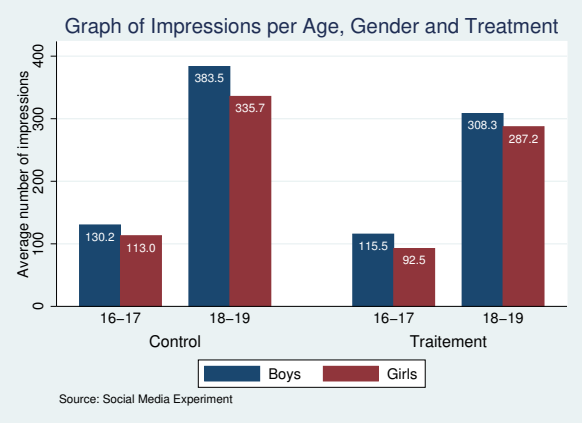
Figure 4 depicts CPM by age and gender and shows the difference between the treatment and control groups. In the control group, boys both in 16-17 age group and in 18-19 age group

Table 5: Daily CPM

	Daily CPM	
	Mean	Std Dev.
Age 16-17		
Girls	0.222	0.132
Boys	0.231	0.157
Age 18-19		
Girls	0.553	0.304
Boys	0.534	0.280

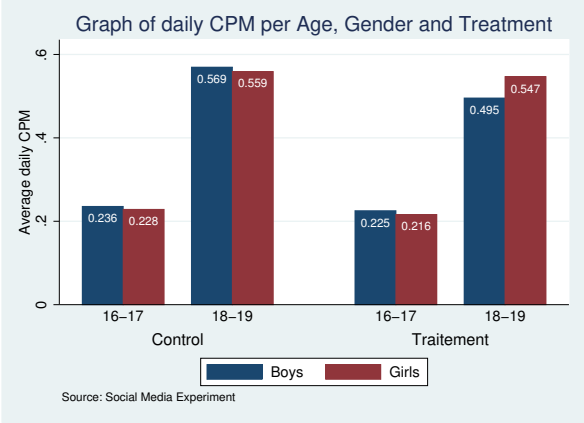
Notes: This table presents detailed summary statistics about daily CPM according to sociodemographic feature such as age and gender. All amounts are expressed in euro. The sample size is 5,484.

cost in average slightly more than girls. The difference were respectively €0.008 and €0.010 more expensive than girls. While girls in the treatment group aged 18–19 were €0.019 more expensive than the corresponding boys, boys were €0.010 on average more expensive than girls in the control group.



Notes: We present differences in the number of impressions according to two sociodemographic categories: age and gender. On the left in the horizontal axis, we show the statistics of the control group, on the right we indicate the treatment group. Standard deviations are respectively: (102.770), (96.917), (307.139), (284.966) for the control group and (120.577), (81.617), (251.487), (254.080) for the treatment group.

Figure 3: Graph bar of impressions



Notes: We present differences in CPM according to two sociodemographic categories: age and gender. On the left in the horizontal axis, we show the statistics of the control group, on the right we indicate the treatment group. Standard deviations are respectively: (0.141), (0.132), (0.569), (0.559) for the control group and (0.172), (0.133), (0.290), (0.326) for the treatment group.

Figure 4: Graph bar of daily CPM

Table 6 depicts the difference between treatment and control. There are statistical difference showing that the treatment ad was less displayed to all. It suggests that prompting women in STEM field results in reducing the ad display.

Table 6: Distribution of the ad: control vs. treatment

	Control	Treatment	p-value
Impressions	242.580	202.727	0.000
	(252.969)	(217.345)	
Frequency	4.236	4.651	0.000
	(2.210)	(5.686)	
Reach	49.683	42. 514	0.000
	(37.847)	(36.397)	
CPM daily	1.602	1.522	0.000
	(0.657)	(0.677)	
Sample size	2,890	2,594	

Notes: This table reports mean estimates for ad display variables for both treatment and control groups. Standard deviations are summarized in parentheses.

5 Econometric analysis

In this section, we propose further results to investigate the effects of the treatment, and of enrollment of girls in science track. We explore a variety of potential explanations that justify the difference in number of impressions between girls and boys and the effect of treatment on the ad performance.

5.1 Empirical models

To estimate whether the ad algorithm distributes the ad equally, and allows for the effect of the treatment, we estimate the following pooled fixed effects model for a high school k, and a demographic group j (gender and age group), at time t:

$$Impressions_{kjt} = \beta_0 + \beta_1 X_j + \beta_2 Girls_j \times Age_j + \beta_3 Treat_k + \beta_4 Treat_k \times Girls_j + \alpha_k + \lambda_t + \epsilon_{kjt}, \quad (1)$$

where X_j is a vector of the demographic controls including the two dummy variables *Girls* and *Age*, and $Treat_k$ is a dummy variable indicating whether high schools received the treatment

ad. We included two interaction terms. $Girls_j$ x Age_j estimates the joint effect of being a girl and aged between 18-19. $Treat_k$ x $Girls_j$ estimates the joint effect of being a girl and receiving the treatment ad. We include a vector of high school fixed effects α_k to capture heterogeneity among high schools, and include a vector of time fixed effects λ_t to account for variation due to different time period (14 days). ϵ_{kjt} is the error term.

In a second stage, we implemented a slightly different version of equation (1) to estimate whether the numbers of girls enrolled in different tracks are considered equally by the algorithm in relation to the number of impressions:

$$Impressions_{kjt} = \beta_0 + \beta_1 X_j + \beta_2 Girls_j \times Age_j + \beta_3 Treat_k + \beta_4 Treat_k \times Girls_j + \rho Y_k + \lambda_t + \epsilon_{kjt}, \quad (2)$$

which includes a vector Y_k of a set of high school characteristics including the dummy variable Paris, whether or not it is a public high school, the number of girls enrolled in the high school, the number of girls enrolled in each main tracks (science, literature and social sciences and economics).

In the previous estimations, we compared girls with boys. To check the robustness of our model, we also estimated the model for the subsample of girls. This allows us to compare girls enrolled in different tracks while controlling for high school characteristics and time period. This robustness check is shown in the estimation below:

$$Impressions_{kijt} = \beta_0 + \beta_1 Age_j + \beta_2 Treat_k + \beta_3 Girls\_in\_track_i + \beta_4 Treat_k \times Girls\_in\_track_i + \alpha_k + \lambda_t + \epsilon_{kijt} \quad (3)$$

where *Girls in track*$_i$ is the number of girls enrolled in the different tracks i and $Treat_k \times$ *Girls in track*$_i$ is the joint effect of the numbers of girls enrolled in one of the three main tracks

that received the treatment ad. We next included a vector of time fixed effect λ_t to capture variation due to different time period, and a vector to control for school characteristics α_k.

5.2 Treatment effect: a crowding-out effect

Table 7 presents the estimates by gender, age group, and inclusion in the treatment group. Columns (1), (2), (4) include high school characteristics and time period. Columns (3), (5), (6) exclude high school fixed effect to avoid collinearity between the variable *Treat* and the vector controlling for high school heterogeneity.

The impact of being a girl is statistically significant for the number of *Impressions*. More precisely, as shown in Table 7 column (1), on average, girls receive 26 fewer impressions than boys. Column (2) suggests that the age group 18 to 19 receives many more impressions (226 on average) than those in the 16 to 17 age group. However, although this age category receives markedly more impressions, girls still receive fewer impressions (column (4)). Column (3) adds the interaction effect of the treatment ad on the ad display. It shows that there is a crowding out effect of treatment – high schools in the treatment group received overall fewer impressions. According to the results in column (6) in Table 7, it would seem that, even if this result is not significant, the treated ad was not shown more to girls than to boys suggesting that unless the message directly prompts girls, the algorithm does not show the ad more to girls compared to boys.

Effect of gender and treatment on impressions

Table 7: Analysis for pooled impressions by treatment and student characteristics

	OLS estimations: Impressions					
	(1)	(2)	(3)	(4)	(5)	(6)
Girls	-25.865***	-23.653***	-27.381***	-11.785***	-19.345***	-24.719***
	(4.692)	(3.523)	(5.610)	(4.416)	(3.937)	(6.538)
Age 18-19		225.825***	217.430***	237.558***	225.387***	225.394***
		(3.565)	(5.536)	(5.112)	(8.090)	(8.091)
Treat			-40.039***		-40.036***	-45.774***
			(5.574)		(5.574)	(8.185)
Girls x Age 18-19				-23.225***	-15.777	-15.792
				(7.002)	(11.060)	(11.063)
Girls x Treat						11.380
						(11.157)
High school fixed effect	YES	YES	NO	YES	NO	NO
Time Fixed effect	YES	YES	YES	YES	YES	YES
cons	436.945***	324.376***	215.518***	318.423***	211.465***	214.190***
	(15.398)	(14.696)	(13.699)	(14.674)	(13.458)	(13.794)
N	5,484	5,484	5,484	5,484	5,484	5,484
R-Squared	0.480	0.705	0.239	0.705	0.239	0.239

Notes: This table reports estimates for impressions according to student's sociodemographic and treatment. Columns (1), (2), (3) include high school characteristics and time period fixed effects. Standard errors, reported in parentheses are robust to heteroskedasticity. Columns(3)(5)(6) do not contain high school fixed effects due to collinearity between this vector of control and the Treatment variable. Significance at 1%; 5% and 10% levels are indicated by ***,** and * respectively.

Click probability, gender, and treatment

A possible interpretation of our main result is that the ad algorithm distributes more likely the ad to individuals who may more interested by the content of the ad. To measure individuals' interest for the ad, we measure the probability to click on the ad. Table 8 presents the probability of clicking on the ad based on the students socio demographics (columns (1)–(2)) and our treatment (columns (3)-(6)). Table 9 shows the marginal effects of these estimations. While girls receive statistically less impressions than boys, girls between 18 and 19 are significantly more likely to click on the ad (Column (4)). Columns (5–6) of Table 9 show that the interaction term *Girls X Age 18 19* which is still positive and significant, girls between 18 and 19 have about 4.5% of probability to click on the ad. The results in column (6) suggest that the treatment does not have a significant impact on the click probability for girls.

Table 8: Analysis for probability of click by treatment and student characteristics

	Probit estimations: click ad					
	(1)	(2)	(3)	(4)	(5)	(6)
Girls	-0.038	-0.039	-0.047	-0.119*	-0.133**	-0.135**
	(0.039)	(0.040)	(0.038)	(0.061)	(0.058)	(0.067)
Age 18-19		0.591***	0.498***	0.522***	0.425***	0.425***
		(0.040)	(0.038)	(0.056)	(0.053)	(0.053)
Treat			-0.106***		-0.106***	-0.108**
			(0.038)		(0.038)	(0.053)
Girls x Age 18-19				0.139*	0.149*	0.149*
				(0.081)	(0.076)	(0.076)
Girls x Treat						0.005
						(0.076)
High school fixed effect	YES	YES	NO	YES	NO	NO
Time Fixed effect	YES	YES	YES	YES	YES	YES
cons	-0.109	-0.409**	-0.805***	-0.370*	-0.763***	-0.762***
	(0.183)	(0.188)	(0.076)	(0.190)	(0.079)	(0.081)
N	5,264	5,264	5,484	5,264	5,484	5,484
Log lik.	-2684.096	-2576.894	-2934.269	-2575.431	-2932.367	-2932.366
Chi-squared	460.656	631.173	199.953	630.920	201.872	201.907
p-value	0.000	0.000	0.000	0.000	0.000	0.000

Notes : This table reports estimates for probability to click according to student's sociodemographic and treatment. Columns (1-2 and 4) include fixed effect high school characteristics and time period. Standard errors, reported in parentheses are robust to heteroskedasticity. Columns(3 and 5-6) do not contain high school fixed effects due to collinearity between this vector of control and the Treatment variable. Significance at 1%; 5% and 10% levels are indicated by ***,** and * respectively.

Table 9: Probability to click: marginal effect

	Probit margins effect: click ad					
	(1)	(2)	(3)	(4)	(5)	(6)
Girls	-0.011	-0.011	-0.014	-0.033*	-0.040**	-0.041**
	(0.011)	(0.011)	(0.011)	(0.017)	(0.017)	(0.020)
Age x 18 19		0.163***	0.150***	0.144***	0.128***	0.128***
		(0.011)	(0.011)	(0.015)	(0.016)	(0.016)
Treat			-0.032***		-0.032***	-0.033**
			(0.011)		(0.011)	(0.016)
Girls x Age 18 19				0.038*	0.045*	0.045*
				(0.022)	(0.023)	(0.023)
Girls x Treat						0.001
						(0.023)
High school fixed effect	YES	YES	NO	YES	NO	NO
Time fixed effect	YES	YES	YES	YES	YES	YES
N	5,264	5,264	5,484	5,264	5,484	5,484

Notes : This table reports marginal effect of probit regression. Columns (1-2 and 4) include high school characteristics and time period. Standard errors, reported in parentheses are robust to heteroskedasticity. Columns (3 and 5-6) do not contain high school fixed effects due to collinearity between this vector of control variables and the Treatment. Significance at 1%; 5% and 10% levels are indicated by ***,** and * respectively.

5.3 What about girls in science track?

Table 10 displays the estimated coefficient of girls enrolled in three main tracks namely Economics, Literary and Science, and the interaction effect of being a girl in the science track and receiving the treatment ad. Here, we are interested in whether the algorithm is able to recognize a more science oriented high school, and to adapt its display accordingly.

Impressions for girls in science oriented high schools

Table 10 reports impressions (columns(1-2)) and click probability (columns(3-4)) estimates by including the percentage of girls enrolled in each tracks (the reference category is *Other tracks*).

Columns (1-2) provide information on impressions and show that high school with a large majority of girls in science receive a higher number of impressions regardless of the treatment ad. We still obtain a negative and significant estimate for the treatment which confirms the

Table 10: Impressions for girls in science-oriented high schools

	Impressions		Click ad	
	(1)	(2)	(3)	(4)
Girls	-20.718***	-20.464***	-0.145**	-0.144**
	(4.589)	(4.551)	(0.067)	(0.067)
Age_18_19	214.676***	214.516***	0.439***	0.438***
	(8.139)	(8.154)	(0.061)	(0.061)
Girls x Age_18_19	-0.197	-0.514	0.171*	0.169*
	(11.459)	(11.485)	(0.088)	(0.088)
Treatment	-56.491***	-7.545	-0.427***	-0.069
	(12.170)	(5.702)	(0.102)	(0.045)
Paris	45.756***	47.312***	0.032	0.041
	(6.354)	(6.366)	(0.049)	(0.049)
Public	-79.643***	-75.330***	-0.104*	-0.068
	(7.532)	(7.532)	(0.059)	(0.058)
Girls_in_Etab	-0.074	-0.069	-0.002	-0.001
	(0.175)	(0.178)	(0.001)	(0.001)
Girls_in_S	1.518***	1.578***	0.005***	0.005***
	(0.169)	(0.169)	(0.001)	(0.001)
Girls_in_S x Treat	0.419***		0.003***	
	(0.101)		(0.001)	
Girls_in_ES	-1.157***	-1.029***	-0.002	-0.001
	(0.147)	(0.146)	(0.001)	(0.001)
Girls_in_L	0.589***	0.527***	-0.000	-0.001
	(0.118)	(0.118)	(0.001)	(0.001)
Time fixed effect	YES	YES	YES	YES
cons	200.174***	176.633***	-0.728***	-0.894***
	(19.085)	(18.614)	(0.137)	(0.131)
N	4,227	4,227	4,227	4,227
R-Squared	0.331	0.328		
Chi2			216.44	204.79
P-value			0.000	0.000

Notes: This table reports estimates of *impressions* (columns 1-2) and *click probability* (columns 3-4) with time fixed effect. Standard errors, reported in parentheses are robust to heteroskedasticity. Significance at 1%; 5% and 10% levels are indicated by ***, ** and * respectively.

presence of a crowding-out effect. The number of impressions is higher in Paris city high schools and private schools.

Based on click probability, columns (3) and (4) suggest that girls in the science track click more often on the ad whatever its type. We extended our statistical analysis to include additional controls such as interaction terms between treatment and the number of girls enrolled in each of the three tracks.

5.4 Is the treatment of girls equal among tracks?

Table **??** summarizes the effect of girls enrolled in the three tracks and the interaction terms between these terms and the variable *treatment*. It reports estimates of impressions (columns (1) and (3)) and click probability (columns (2) and (4)) controlling for the number of girl enrolled in the three tracks (science, literary , and economics) (columns(3-4)). We add the interaction terms between the numbers of girls enrolled in each track and the *Treatment* variable (columns (1) and (2)).

On average, girls enrolled in science receive significantly more impressions than girls enrolled in other tracks while the crowding-out effect of treatment disappears (column (3)). This might be because girls taking science have a higher probability of clicking on the ad, and in this case, treatment has a positive and significant impact suggesting that receiving the treated ad increases click frequency for girls (column (4)).

Our estimates are robust to changes in the specifications in this analysis with a dominant significant effect of receiving a lower number of impressions for girls compared to boys 7, and a higher probability of girls aged between 18-19 clicking on the ad.

Table 11: Effect of the treatment on impressions and click for girls only

	Impressions		Click ad	
	(1)	(2)	(3)	(4)
Age 18-19	213.318***	213.596***	0.616***	0.610***
	(7.875)	(8.240)	(0.064)	(0.063)
Treatment	48.797***	-7.711	0.044	-0.065
	(18.697)	(8.146)	(0.152)	(0.063)
Girls in S	0.395	1.595***	0.002	0.006***
	(0.260)	(0.208)	(0.002)	(0.002)
Girls in ES	-2.971***	-1.463***	-0.009***	-0.003**
	(0.315)	(0.213)	(0.002)	(0.002)
Girls in L	2.596***	0.630***	0.007***	0.000
	(0.206)	(0.156)	(0.002)	(0.001)
Girls in S x Treat	2.906***		0.010***	
	(0.388)		(0.003)	
Girls in ES x Treat	2.574***		0.010***	
	(0.434)		(0.003)	
Girls in L x Treat	-3.977***		-0.014***	
	(0.286)		(0.002)	
Time fixed effect	YES	YES	YES	YES
cons	36.764*	67.135***	-1.199***	-1.105***
	(20.906)	(20.828)	(0.154)	(0.140)
N	2,126	2,126	2,126	2,126
R-Squared	0.366	0.309		
log-likelihood	-14080.761	-14173.023	-1063.069	-1083.716
Chi2			165.307	130.545
P-value	0.000	0.000	0.000	0.000

Notes: This table reports estimates for impressions and click probability for a subsample of girls. Columns (1 and 3) reports estimates for impressions and columns (2 and 4) reports estimates for the probability to click on the ad. Each regression includes time period fixed effects. Standard errors, reported in parentheses are robust to heteroskedasticity. Significance at 1%; 5% and 10% levels are indicated by ***,** and * respectively.

6 Conclusion

Girls enrolled in STEM higher education programs continue to be under-represented compared to boys. Our paper contributes by adding the role of social network as vector of information diffusion among teens. It contributes also to the literature on the economics of privacy. Teenagers who subscribe to a social media reveal their gender with no awareness that this information could potentially be used to discriminate them in the context of particular types of information such as ads related to STEM jobs.

We conducted a two-week, high school level, online experiment to test whether an algorithm might handle any bias. We relied on administrative data which provides us with detailed information on high schools, and conduct 104 simultaneous ad campaigns. High schools were split into two groups, control and treatment, to test the ad display formulated by the algorithm. In order to limit selection bias, we split the schools into groups based on a randomization procedure.

Our paper is closely related to Lambrecht and Tucker (2017). However, while they observe a cost difference between boys and girls which might explain the differences in ad delivery, we found no evidence of this. We also conducted several robustness check which showed that the decisions made by the algorithm towards girls enrolled in science were fairer. These additional findings allow us to (i) verify the existence of bias, but (ii) conflict with prior findings that cost explains ad display disparities between boys and girls. We favor and investigated the second explanation offered by Lambrecht and Tucker (2017) which is based on the concept of spillovers. Finally, we highlight that algorithms might be an effective and fairer means of targeting girls attending science oriented high schools who will likely be more interested in the ad set up.

This study makes several contributions. It adds to work on the existence of bias and the reproduction of stereotypes related to teenagers interest in STEM. Second, we tried to explain

this finding relying on the related economics literature. Third, we moderate our results by showing the effectiveness of the algorithm for targeting those girls who might be most interested in the ad, i.e. prioritizing girls attending science oriented high schools.

This research raises several questions about the role of machine learning algorithms for ad effectiveness and ad personalization. Although intelligent algorithms might generate apparent discrimination and lower levels of ad display to girls, they are able to decide to show the add to those girls likely to be most interested in the ad content. We cannot ignore the potential effectiveness of these tools and their attraction for firms and policy makers.

References

Acquisti, A., Taylor, C. and Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*. 54(2), 442–92.

Akerlof, G. A. and Kranton, R. E. (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press.

Angwin, J. (2016). Make algorithms accountable. *The New York Times*.

Autor, D. H., Figlio, D. N., Karbownik, K., Roth, J. and Wasserman, M. (2016). *School Quality and the Gender Gap in Educational Achievement*. Working Paper 21908. National Bureau of Economic Research.

Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*. 131(4), 1753–1794.

Carrell, S. E., Page, M. E. and West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics*. 125(3), 1101–1144.

Chatterjee, P., Hoffman, D. L. and Novak, T. P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science*. 22(4), 520–541.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*. 126(4), 1593–1660.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*. 47(2), 448–74.

Datta, A., Tschantz, M. C. and Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *ArXiv e-prints*.

Fryer, R. G. and Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*. 2(2), 210–40.

Goldfarb, A. and Tucker, C. E. (2011). Privacy Regulation and Online Advertising. *Management Science*. 57(1), 57–71.

Guiso, L., Monte, F., Sapienza, P. and Zingales, L. (2008). Culture, Gender, and Math.

Science. 320(5880), 1164–1165.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2017). Human decisions and machine predictions. *The Quarterly Journal of Economics*. forthcoming.

Lambrecht, A. and Tucker, C. E. (2017). Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. Working Paper.

Leibbrandt, A. and List, J. A. (2015). Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment. *Management Science*. 61(9), 2016–2024.

Nollenberger, N., Rodríguez-Planas, N. and Sevilla, A. (2016). The Math Gender Gap: The Role of Culture. *American Economic Review*. 106(5), 257–61.

OECD (2017). Going Digital: The Future of Work for Women. *Policy brief on the future of work*.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Crown Publishing Group.

Pope, D. G. and Sydnor, J. R. (2010). Geographic Variation in the Gender Differences in Test Scores. *Journal of Economic Perspectives*. 24(2), 95–108.

Stitelman, O., Dalessandro, B., Perlich, C. and Provost, F. (2011). Estimating The Effect Of Online Display Advertising On Browser Conversion, 816.

Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Queue*. 11(3), 10:10–10:29.

Tucker, C. (2017). Privacy, Algorithms and Artificial Intelligence. In *Economics of Artificial Intelligence*. University of Chicago Press.