STEM and teens: An algorithm bias on a social media

Cecere, G. [1, 2] **Jean, C.** [2, 3] Le Guel, F.[2] Manant, M.[2]

[1]Institut Mines Telecom, TEM

[2]University of Paris-Sud, RITM

[3]EPITECH, PARIS

March 29, 2018

Motivation and contribution

Motivation

- An increasing number of teens use social media as a source of information: a bias against a certain group of individuals may then reinforce biased choices by social media's users
- an emerging literature identifies a bias on social media: as an example, Lambrecht and Tucker (2017) show a bias against girls in favor of boys via a field experiment
- this may offset policy initiatives that aim to reduce other bias such as the gender gap in Science, Technology, Engineering and Mathematics (STEM) education.

Contribution

- We ran a field experiment to establish a counterfactual element that can explain differences in ad distribution
- A link between offline education environment (administrative data) with online content

Research question

Are social media algorithms biased?

Is it possible to reduce the bias by prompting the algorithm?

Our work in one slide

Field experiment data on STEM education ad in France
Settings:

- Run an ad campaign at the high school level $\Rightarrow$ 101 high schools
- Set up a gender neutral ad
- Target teens between 16-19
- Randomization at the high school level (Treatment vs Control)

Results:

- Dependent variable: *Impressions*
- Girls were shown less impressions compared to boys
- Girls between 18 and 19 clicked more if they came across the ad rather than boys
- Crowding-out effect of the treatment ad

Contribution to the existing literature

We contribute to two literature strands:

- **Algorithm bias**:
  While they can improve ad effectiveness or reduce human biases, they show apparent discriminatory outcome
  $\Rightarrow$ Identify and explain potential bias (counterfactual and education data), evaluate negative spillovers from a user's point of view
- **Economics of privacy**
  We meet a trade-off between ad effectiveness and a lock-in situation
  $\Rightarrow$ Disclosed information may keep us in a category of information based on prejudice

Algorithm bias

Algorithms may improve effectiveness and be fairer

- Machine learning algorithm (MLA) can help prevent human biases (Kleinberg, et al. 2018)
- MLA can improve ad distribution (Stitelman, et al. 2011)

Nevertheless, they can also provide apparent discriminatory outcomes.

- Ethnic bias (Sweeney, 2013)
- Gender bias
  - Datta et al. (2014) identify gender bias in cases of ad for high paid jobs.
  - Lambrecht and Tucker (2017) show gender discrimination in STEM jobs ad explained by eyeballs and spillovers.

Economics of privacy

Disclose personal data: for what?

- Immediate gratification (Acquisti et al., 2016)
- Improve ad effectiveness (Goldfarb and Tucker, 2011)

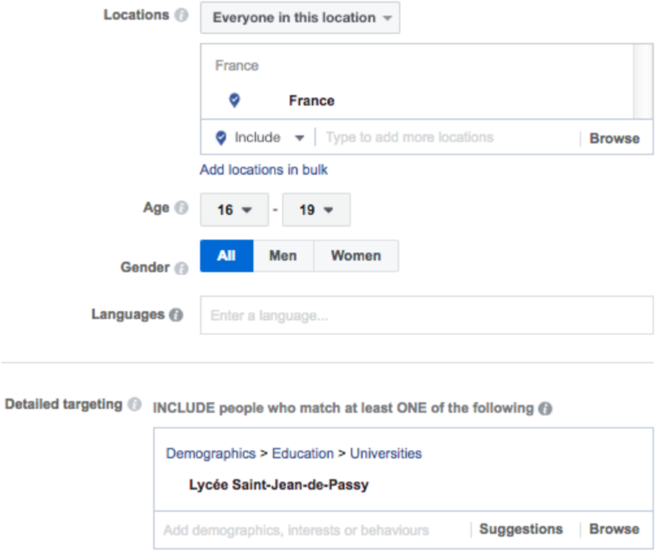
But, disclosing data might have other unexpected implications

- Data collection can generate unintended spillovers, which might have potential negative effects on short- and long-term (Tucker, 2017)
- Lock-in situation

Design

Target	French high school students
Age	16-19
Location	France
Gender	All
Campaign level	High school (101)
	Run simultaneously 101 ad campaigns
Randomization process	Yes
Treatment	Girl content in text heading
Duration	2 weeks
Daily budget CPM	2€
Optimization goal	Impressions
Analysis unit	High school-Gender-Age

Settings


Locations
Everyone in this location
France
France
Include
Type to add more locations
Browse
Add locations in bulk
Age
16
-
19
Gender
All
Men
Women
Languages
Enter a language...
Detailed targeting
INCLUDE people who match at least ONE of the following
Demographics > Education > Universities
Lycée Saint-Jean-de-Passy
Add demographics, interests or behaviours
Suggestions
Browse

Ad view

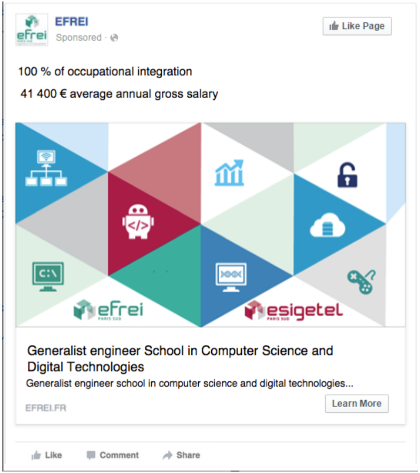


Figure: Control ad

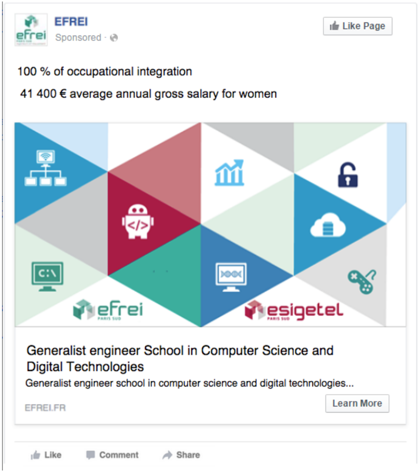


Figure: Treatment ad

Data

2 Sets of data:

- Administrative: students enrollment, proportion of girls, graduation rate, ect.
- Experimental: ad performances such as impressions, clicks, frequency, amount spend at the high school level

Context

- In France, 12th grade students are required to state their education preferences for post-secondary education on a government platform.
- In 2017, the deadline was March, 21st. We conducted the experiment between March, 11th and March, 26th.
- Goal of the campaign :
  - for the school: to encourage enrollment of new students, especially girls in the school
  - for us: to look at the effect of counterfactual on ad display

Statistical evidence

Randomization procedure: T-test

Table: Pre-treatment summary statistics

	Control			Treatment			p-value
Variable	**Mean**	**Std. dev.**	**N**	**Mean**	**Std. dev.**	**N**	
Proportion of girls overall	0.517	(0.094)	52	0.554	(0.129)	49	0.097
Science track	0.458	(0.106)	49	0.475	(0.091)	44	0.383
Economics and Social Science track	0.591	(0.083)	49	0.610	(0.089)	40	0.316
Literacy track	0.777	(0.102)	42	0.788	(0.102)	39	0.606
Graduation rate	0.912	(0.806)	52	0.914	(0.100)	49	0.920
Science track	0.900	(0.115)	49	0.921	(0.102)	44	0.368
Economics and Social Science track	0.916	(0.78)	49	0.912	(0.117)	40	0.833
Literacy track	0.940	(0.062)	42	0.923	(0.113)	39	0.403
Educational stages distribution							
12th-grade	0.320	(0.027)	52	0.325	(0.023)	49	0.354
11th-grade	0.340	(0.032)	52	0.332	(0.023)	49	0.156
10th-grade	0.340	(0.043)	52	0.343	(0.036)	49	0.670
Tracks distribution							
Science track	0.446	(0.169)	49	0.438	(0.136)	44	0.802
Economics Social Science track	0.252	(0.082)	49	0.279	(0.087)	40	0.141
Literacy track	0.120	(0.053)	42	0.144	(0.082)	39	0.113
Public schools[1]	0.788	(0.412)	52	0.776	(0.442)	49	0.876
Parisian schools [2]	0.308	0.466	52	0.306	0.466	49	0.987
Average enrollment[3]	852.192	(375.519)	52	845.469	(410.623)	49	0.932

Notes: This table reports mean estimates for the variables in our data set for both the treatment and control groups. Standard deviations are in parentheses.
4 Public school takes value 1 if the high school is public
5 Parisian school takes value 1 if the high school is located in Paris area
6 Average enrollment

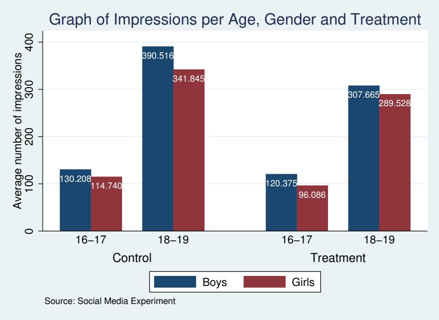


Figure: Bar graph of impressions

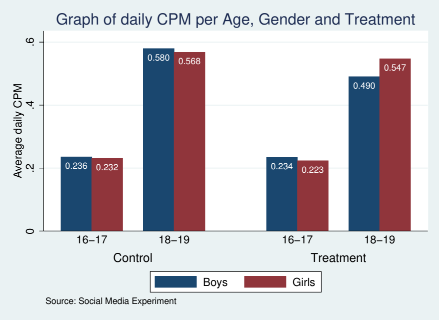


Figure: Bar graph of daily CPM

Ad display according to high schools groups

Table: Distribution of the ad: control vs treatment

	Control		**Treatment**		**p-value**
	Mean	Std. dev.	Mean	Std. dev.	
Impressions	245.702	(253.268)	205.301	(217.423)	0.000
Ad clicks	0.261	(0.439)	0.230	(0.421)	0.009
CPM daily	0.406	(0.273)	0.377	(0.283)	0.000
CPM daily 16-17	0.234	(0.136)	0.228	(0.153)	0.356
CPM daily 18-19	0.574	(0.269)	0.519	(0.306)	0.000
Reach	50.302	(37.729)	43.619	(36.560)	0.000
Frequency	4.251	(2.207)	4.610	(5.759)	0.002
Sample size	2,851		2,482		

Are girls and boys treated equally (1)

Table: Ad display according to gender (boys vs girls)

	Overall		**Boys**		**Girls**		**p-value**
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	
Impressions	226.899	(238.093)	241.247	(245.479)	212.739	(229.740)	0.000
Ad clicks	0.247	(0.431)	0.254	(0.435)	0.239	(0.427)	0.208
CPM overall	0.392	(0.278)	0.389	(0.271)	0.395	(0.285)	0.618
CPM 16-17	0.231	(0.144)	0.228	(0.131)	0.225	(0.156)	0.223
CPM 18-19	0.548	(0.288)	0.558	(0.301)	0.538	(0.275)	0.067
Reach	47.191	(37.335)	47.262	(37.573)	47.121	(37.106)	0.890
Frequency	4.418	(4.251)	4.815	(5.312)	4.026	(2.783)	0.000
Sample size	5,333		2,684		2,649		

Econometric specification and results

Econometric specification

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

$$\begin{aligned} Impressions_{ijt} = & \beta_0 + \beta_1 GirlsSN_i + \beta_2 Age_i + \beta_3 Treatment_j + \\ & \beta_4 (GirlsSN_i \times Age_i) + \beta_5 (Girls_i \times Treatment_j) + \alpha_i + \lambda_t + \epsilon_{ijt}, \end{aligned} \quad (1)$$

$Treatment_j$= dummy variable equal to 1 if treatment ad
λ_t = vector of time fixed effects
α_i = vector of high school fixed effects
ϵ_{ijt}= the error term

Baseline results

Table: OLS estimations: Number of impressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Girls SN	-25.591***	-23.984***	-26.704***	-13.319***	-18.785***	-24.262***	-32.171***
	(4.768)	(3.583)	(5.688)	(4.472)	(3.997)	(6.544)	(5.995)
Age 18-19		226.253***	219.280***	236.802***	227.126***	227.165***	227.165***
		(3.617)	(5.628)	(5.176)	(8.210)	(8.215)	(8.216)
Treatment			-41.259***		-41.273***	-47.228***	-47.228***
			(5.653)		(5.654)	(8.276)	(8.277)
Girls SN x Age 18-19				-20.938***	-15.588	-15.643	0.063
				(7.125)	(11.239)	(11.247)	(13.745)
Girls SN x Treatment						11.830	28.837***
						(11.315)	(9.602)
Girls x Treatment x Age 18-19							-33.674**
							(15.295)
High school fixed effects	Yes	Yes	No	Yes	No	No	No
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	438.223***	325.692***	219.588***	320.360***	215.619***	218.379***	218.391***
	(15.532)	(14.782)	(13.883)	(14.761)	(13.640)	(13.963)	(13.958)
Observations	5,333	5,333	5,333	5,333	5,333	5,333	5,333
R-squared	0.480	0.705	0.242	0.705	0.242	0.243	0.243

Interest on the ad: Probability to clicks

Table: Probability to clicks: Marginal effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Girls SN	-0.012	-0.012	-0.015	-0.036**	-0.041**	-0.041**	-0.048**
	(0.012)	(0.011)	(0.012)	(0.017)	(0.018)	(0.020)	(0.022)
Age 18-19		0.164***	0.152***	0.144***	0.129***	0.129***	0.129***
		(0.011)	(0.011)	(0.016)	(0.016)	(0.016)	(0.016)
Treatment			-0.031***		-0.031***	-0.030*	-0.030*
			(0.012)		(0.012)	(0.016)	(0.016)
Girls SN x Age 18-19				0.041*	0.046**	0.046**	0.058**
				(0.023)	(0.023)	(0.023)	(0.028)
Girls SN x Treatment						-0.001	0.015
						(0.023)	(0.030)
Girls SN x Treatment x Age 18-19							-0.026
							(0.034)
High school fixed effects	Yes	Yes	No	Yes	No	No	No
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,113	5,113	5,333	5,113	5,333	5,333	5,333

Robustness check

Table: Estimations of the number of impressions: Robust standard errors clustered at high school level

	(1)	(2)	(3)	(4)	(5)	(6)
Girls	-28.180***	-26.794***	-18.951**	-26.704***	-24.262***	-32.171***
	(9.496)	(9.412)	(7.728)	(5.688)	(6.544)	(5.995)
Age 18-19		219.087***	226.859***	219.280***	227.165***	227.165***
		(20.398)	(22.273)	(5.628)	(8.215)	(8.216)
Girls SN x Age 18-19			-15.439		-15.643	0.063
			(14.249)		(11.247)	(13.745)
Treatment				-41.259***	-47.228***	-47.228***
				(5.653)	(8.276)	(8.277)
Girls SN x Treatment					11.830	28.837***
					(11.315)	(9.602)
Girls SN x Age 18-19 x Treatment						-33.674**
						(15.295)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
High school fixed effects	Yes	Yes	Yes	No	No	No
Constant	311.002***	200.475***	196.537***	219.588***	218.379***	218.391***
	(23.093)	(16.443)	(15.897)	(13.883)	(13.963)	(13.958)
Observations	5,333	5,333	5,333	5,333	5,333	5,333
R-squared	0.023	0.235	0.235	0.242	0.243	0.243

Notes: OLS estimations. The dependent variable is number of *Impressions*. Columns (1), (2), (4) include high school characteristics. Robust standard errors clustered at high school level are reported in parentheses. Columns (3), (5), (6), (7) do not include high school fixed effects since they are collinear with the dummy variable of treatment. All the regressions include day fixed effects. Significance at 1%; 5% and 10% levels indicated respectively by ***,** and *.

Table: Impressions (OLS) and Ad Click (Probit estimation):

	Impressions		**Ad click**	
Girls SN	-25.942***	-25.900***	-0.124*	-0.123*
	(4.394)	(4.399)	(0.064)	(0.064)
Age 18-19	221.783***	221.809***	0.394***	0.395***
	(8.427)	(8.425)	(0.060)	(0.060)
Girls SN x Age 18-19	-8.715	-8.757	0.149*	0.148*
	(11.892)	(11.895)	(0.085)	(0.085)
Treatment	-26.746***	-48.210	-0.047	-0.606**
	(6.188)	(33.037)	(0.042)	(0.273)
Public	-69.987***	-69.200***	-0.132**	-0.112**
	(7.536)	(7.692)	(0.055)	(0.056)
Girls in Science	-2.099***	-2.294***	-0.010***	-0.016***
	(0.354)	(0.459)	(0.003)	(0.004)
Girls in Economics and social science	1.737***	1.703***	0.005	0.004
	(0.400)	(0.407)	(0.003)	(0.003)
Girls in Literacy	-0.622*	-0.583	-0.003	-0.002
	(0.360)	(0.365)	(0.002)	(0.003)
Professional High school	-45.643	-32.681	-0.147	0.188
	(34.662)	(39.450)	(0.272)	(0.323)
Girl in Science x Treatment		0.441		0.012**
		(0.675)		(0.006)
Time fixed effects	Yes	Yes	Yes	Yes
Constant	323.367***	331.061***	-0.166	0.036
	(27.906)	(30.361)	(0.213)	(0.226)
Observations	4,269	4,269	4,269	4,269

Significance at 1%; 5% and 10% indicated respectively by ***, ** and *.

Table: Impressions (OLS) and Ad click (Probit estimations): Sample with girls only

	Impressions		**Ad click**	
Age 18-19	213.104***	215.331***	0.544***	0.563***
	(8.468)	(8.115)	(0.061)	(0.062)
Treatment	-31.946***	-156.200**	-0.116*	-1.070*
	(8.476)	(70.678)	(0.060)	(0.588)
Girls in Science	-1.722***	-4.185***	-0.011***	-0.022***
	(0.458)	(0.568)	(0.004)	(0.005)
Girls in Economics and Social Science	1.909***	-6.321***	0.005	-0.018***
	(0.563)	(0.779)	(0.004)	(0.006)
Girls in Literacy	0.311	5.028***	-0.003	0.009*
	(0.487)	(0.665)	(0.003)	(0.005)
Girls in Science x Treatment		1.842**		0.013
		(0.937)		(0.008)
Girls in Economics and Social Science x Treatment		16.117***		0.046***
		(1.034)		(0.009)
Girls in Literacy x Treatment		-11.881***		-0.031***
		(0.940)		(0.008)
Time fixed effects	Yes	Yes	Yes	Yes
Constant	133.887***	375.352***	-0.328	0.606
	(28.342)	(45.258)	(0.224)	(0.395)
Observations	2,147	2,119	2,147	2,119
R-squared	0.253	0.330		

Significance at 1%; 5% and 10% are respectively indicated by ***,** and *.

Impressions: Household income level

	(1)	(2)	(3)	(4)	(5)
Girls SN	-30.440***	-30.411***	-30.653***	-31.441***	-31.498***
	(5.570)	(5.754)	(5.771)	(5.676)	(5.547)
Age 18-19	227.130***	226.217***	226.091***	227.415***	227.649***
	(5.533)	(5.695)	(5.718)	(5.640)	(5.517)
Girls in Science track	-2.287***	6.335***	-0.110	2.365***	-6.024***
	(0.415)	(0.623)	(0.712)	(0.482)	(1.317)
Treatment	-52.780***	-48.219***	-31.941***	-47.048***	-53.950***
	(6.059)	(5.933)	(5.996)	(5.833)	(6.220)
Prop. Household high income	-0.936**				-0.328
	(0.459)				(1.363)
Girls in Science track x Prop. Household high income	0.091***				**0.090***
	(0.010)				(0.015)
Household high middle income		14.510***			15.490***
		(2.391)			(2.852)
Girls in Science track x Prop. Household high middle income		-0.530***			-**0.366***
		(0.053)			(0.064)
Prop. Household middle income			-7.526***		-15.817***
			(1.326)		(2.053)
Girls in Science track x Prop. Household middle income			0.013		**0.366***
			(0.029)		(0.029)
Prop. Household low income				-1.826*	3.390*
				(0.946)	(1.730)
Girls in Science track x Prop. Household low income				-0.070***	-**0.055** *
				(0.020)	(0.028)
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Constant	178.997***	56.969*	370.619***	212.277***	309.357**
	(24.399)	(29.973)	(34.585)	(24.792)	(126.475)
Observations	4944	4944	4944	4888	4888
R-squared	0.347	0.303	0.299	0.334	0.364

Limitation

- The additional number of words in the treatment ad might also justify the decrease of the number of impressions.
- Can we talk about a bias? Algorithm replicates what it has learned from biased data (O'Neil, 2016).

Conclusion

- Field experiment suggests that a gender-neutral ad may not be allocated efficiently for the computer science school
- While **girls saw less impressions**, this difference cannot be explained by ad costs differences, **girls aged between 18-19 have a greater interest** in it as they have higher probability of clicking on the ad.
- Ambivalent results about the treatment: **crowding-out effect** but a positive impact on girls studying Economics and Science.
- Policy implications: **Algorithms bias might offset policy initiatives aiming to encourage enrollment of girls in STEM disciplines**.

Thank you for your attention !