# Computer algorithms prefer headless women

Grazia Cecere[1], Clara Jean[2,3], Matthieu Manant[3], and Catherine Tucker[4]

[1]Institut Mines-Télécom Business School, [2]Epitech, [3]University of Paris-Sud,
[4]MIT Sloan School of Management

January 5, 2019

**Abstract**

Advertising algorithms power the online digital economy. However, it is unclear whether they may end up distorting the kind of information people are exposed to. To explore this we ran a randomized online ad campaign on Snapchat on behalf of a French computer science school that explored how the ad algorithm allocated pictorial content representing gender. Our results show that pictures depicting a complete male torso was shown more to teens, while the female picture that was displayed most by the algorithm depicted the woman as not having a head. We present suggestive evidence that these algorithms are driven by preferences in large population centers in Paris as it appears the algorithm determines which images are the most "engaging" on the first day for the places with the largest numbers of users and replicates this pattern going forward elsewhere.

This paper is presented at the ASSA, Atlanta (2019). We thank the conference participants at MIT Code Conference, Boston (2018) for their helpful comments. This work was supported by grant 17-MA-04 from the Paris-Saclay Maison des Sciences de l'Homme (MSH). We also thank the computer science school EPITECH for its support. All errors are our own. E-mails: grazia.cecere@imt-bs.eu, clara.jean@epitech.eu (corresponding author), matthieu.manant@u-psud.fr, and cetucker@mit.edu.

[1]Institut Mines-Télécom Business School, 9, rue Charles Fourier - 91000 Évry, France.

[2]Epitech, 24, rue Pasteur - 94270 Le Kremlin-Bicêtre, France.

[3]University of Paris-Sud, Paris-Saclay, 54, bd Desgranges - 92330 Sceaux, France.

[4]MIT Sloan, Cambridge, MA.

# 1 Introduction

Ad algorithms are designed to maximize efficiency for the advertiser but they also control the kind of ads and information to which users are exposed. Though algorithms are determining who sees billions of ad impressions little work has been done to explore what influences algorithmic-decision making. Previous work shows that an ad algorithm can reproduce apparent discriminatory biases due to the behavior of other advertisers (Lambrecht and Tucker, 2018). Given empirical evidence of discrimination in online markets (see Fisman and Luca (2016) for an overview), it is useful to explore how the process by which algorithms make determinations could also lead to apparent bias. In this study, we look at how the pictorial content that an ad campaign contains can lead algorithms to show different types of ads to different groups in ways that may initially appear unsettling and discriminatory.

To explore this we ran a randomized online ad campaign on Snapchat, a social network used by many teenagers, with 78% of teenagers aged 18 to 24 using it regularly (Pew Research Center, 2018a).[1] With 191 million daily users (SnapInc, 2018), Snapchat is the favorite platform of 45 percent of teens[2] and get much visual attention as TV ads.[3]

The campaign was conducted on behalf of a French computer science school. The field experiment used a $2 \times 2$ design with four treatments to explore how the algorithm reacted to different ways of portraying gender in pictorial content. The images in these ads differed in the message written on the individual's t-shirt, and whether or not the photo included a person's head. The first two photos of a woman and a man were taken from the back but their gender was displayed quite clearly to both the consumers and the algorithm. The second two photos showed the same individuals' headless images and taken from the back which made it difficult to identify the gender. We also varied whether or not the image

[1]Only 51% of American teens aged 13 to 17 now declare using Facebook which is a significantly lower proportion than Instagram and Snapchat users (Pew Research Center, 2018b).

[2]According to 2018's Taking Stock With Teens survey, See http://www.piperjaffray.com/3col.aspx?id=5383, last retrieved December, 2018. In the case of Instagram, this figure is only of 26%

[3]See https://blog.hootsuite.com/social-media-advertising-stats/, last retrieved December, 2018

included a message that emphasized female or male empowerment. We ran the ad campaign over a period of nine days, targeting high schoolers aged between 16 and 19 in different French towns. We used data from a field experiment involving more than 268 different ad campaigns, two for each town targeted.[4]

The image that included a female-oriented message and included the female's head was less likely to be shown more to all users compared to the other photos. We provide suggestive evidence that ad algorithm is driven by preferences in large population centers such as Paris. Our results suggest that algorithmic-decision making can lead to outcomes that can distort the kind of information people are exposed to.

Our paper grows to the nascent literature which tries to understand the economics behind apparent algorithmic bias. In the context of online advertising, early work revealed a biased distribution of advertisements based on ethnic origin and gender (Sweeney, 2013; Datta *et al.*, 2015). Some of the most recent articles propose economic explanations for such biases. Lambrecht and Tucker (2018) conducted a field test on a social media regarding a gender-neutral ad for STEM jobs, and showed that women were less likely to see an ad because the ad algorithm seeks to minimize the advertiser's costs by avoiding expensive female eyeballs. Similarly, Cecere *et al.* (2018) ran a field experiment at the high school level in relation to a gender-neutral ad for STEM education distributed by the algorithm on a social network. They identified a bias against women (not explained by the behavior or prices of other advertisers). The treatment ad, which was intended to be more popular with women, faced a crowding-out effect and was generally less well displayed to both men and women. We contribute to this literature by exploring the role of the distribution of user eyeballs from which the algorithm learns behavior may contribute to bias.

We also contribute to the larger literature on bias in general in online platforms, which highlights that, instead of reducing market discrimination by reducing information asymmetries

[4]Campaigns are ran separately for male and female.

among users, many online markets such as the housing or labor markets, are marked by bias against some users (Edelman *et al.*, 2017; Doleac and Stein, 2013; Fisman and Luca, 2016; Manant *et al.*, 2018). This stream of research highlights the role of platform design choices in these discriminations. Our research highlights that by building algorithms that attempt to learn quickly what content appeals most to users, platforms may be inadvertently creating situations where population centers determine what the rest of the network sees.

Finally, we contribute to the literature that attempts to optimize ad distribution from a managerial perspective. This literature has highlighted that multi-armed bandit (MAB) method can help determine which advertising content to display (Schwartz, 2013). Originally, MAB problems are defined as decision models where agents seek to optimize their decisions based on the information they have (Robbins, 1985).[5] The key idea behind a multi-armed bandit problem is that there is a tradeoff between exploration and efficiency. The algorithm tries to determine when it has explored enough (for example by showing multiple different versions of an ad) to understand relative preferences, and then at that point persists in showing the most preferred ad as there are fewer gains to exploration. In the context of crowdsourcing and using real data, Tran-Thanh *et al.* (2014) shows that an algorithm based on a MAB model exceeds the existing crowdsourcing methods by more than 300%. Johnson *et al.* (2017) provides evidence that algorithm targeting can improve ad distribution, and Schwartz *et al.* (2017) using MAB methods show that the implementation of this method allows for an 8% improvement in the customer acquisition rate compared to a control group, without any additional investments for the company. However, our article shows how the process by which an algorithm may learn to display content and then persistently replicate what it learns can lead to advertising outcomes that appear to be discriminatory and unsettling.

The article is organized as follows. Section 2 describes the design of our field experiment and Section 3 presents the descriptive statistics. Section 4 presents the results, and Section 5 concludes.

[5]See Bergemann and Välimäki (2006) for a complete review applied to economics

## 2 Research design

We ran ad campaigns on behalf of a French computer school. The ads displayed the back views of individuals (from the waist up, including and excluding the head) wearing black t-shirts bearing the advertising message. We used a $2 \times 2$ design which varied the individuals' genders (woman / man) and type of photo (with / without head). In the photos depicting a young woman and a young man including the head, gender was clearly recognizable; in the two photos depicting the same individuals but without their heads, the genders were less clear (see Fig. 1) as only the torso was present. The advertising message on the back of the black t-shirts worn by the young woman read "50% Woman 50% Machine 100% Epitech" (female message), and the message on the back of the t-shirts worn by the young man read "50% Man 50% Machine 100% Epitech" (male message).[6] In the headless images gender can be inferred only from the message.


(a) Female *complete* (b) Male *complete* (c) Female *headless* (d) Male *headless*

Figure 1: Images of the four different ads

To obtain an insight into how an algorithm might interpret the photos, we analyzed them using the online Google Cloud Vision API tool which uses artificial intelligence to categorize the photos (see Figures 19,20, 17, 18 in appendix). Google's algorithm was able to clearly identified the gender of the individuals in photos with heads but failed to identify them in

[6]Epitech is the name of the computer school

the headless photos.

The ad campaigns ran for nine consecutive days during mid-June 2018. We targeted 16 to 19 year old high schoolers in 134 French cities; each city was associated to a particular photo, identified using an ex ante randomization procedure based on the cities' demographic and socio-economic characteristics (see Appendix 7 for the results of randomization procedure). The treatments female complete and female headless were displayed in 34 cities and the treatments male complete and male headless were displayed in 33 cities.

We conducted a total of 268 simultaneous but different ad campaigns i.e. two per city - one targeting women and one targeting men. This allowed us to test a potential difference in the distribution of the ads between women and men. For three cities - Paris, Lyon and Marseille which are the three largest French cities - we targeted districts which allowed us to display all four treatments in each city.[7]

Lastly, following the automated Snapchat platform suggestion, for each ad campaign we allowed a daily budget of €50, and a bid of €2.6 per day. The cost of the campaigns was based on CPM.[8] Although this might suggest that the campaign would be optimized based on the number of impressions, we show that the algorithm considered user engagement; the Snapchat algorithm now only allows advertisers to bid based on engagement.

# 3 Data

The ad campaign received 2,174,513 impressions in total and it resulted in a total of 2,412 observations. Our key outcome variable is the number of impressions received by each city. For each ad campaign, we received aggregate data from Snapchat which included the total number of impressions, total number of swipe ups, swipe up rate, and amount spent. A

[7]Paris, Lyon and Marseille are the three largest French cities based on population numbers. In France, districts are called *arrondissements*. The Paris inner city includes more than 2 million people while Paris suburbs include more than 10 million people and several departments.

[8]CPM is Cost per mille (i.e. per thousand) impressions.

swipe up[9] occurs when the snapchatter swipes up to view the ad attachment, in our case the computer school's official website. In addition to ad performance data, we collected information on the number of snapchatters in each city for each gender group at the start of the ad campaign.[10] We matched city-level administrative data to the ad campaign data provided by the platform.

## 3.1 Ad Distribution

Table 1 shows how the different images were distributed for the entire sample and for each treatment group and includes F-tests of equality of means. Table 8 presents the full descriptive statistics. Each advertising campaign included 901 impressions and 2.6 swipe ups giving a swipe up rate of 0.293%.[11] The numbers of impressions and swipe ups were statistically different among the four groups, highlighting the unequal distribution of the ads among the groups. On average, the complete male and the headless male photos were displayed more often by the algorithm than the complete female photo. The different ad distribution patterns appear not to be explained by differences in the unit costs since these differences are not statistically different.

Table 1: Ads performance overall and for the four treatment groups

| | **Overall** | **Female complete** | **Female headless** | **Male complete** | **Male headless** | **F-test** |
|---|---|---|---|---|---|---|
| | Mean (sd) | Mean (sd) | Mean (sd) | Mean (sd) | Mean (sd) | p-value |
| Impressions | 901.5 | 713.2 | 812.7 | 1137.8 | 938.9 | 0.000 |
| | (917.8) | (424.4) | (733.6) | (1127.9) | (1133.5) | |
| Swipe ups | 2.619 | 2.038 | 2.397 | 3.285 | 2.745 | 0.000 |
| | (3.044) | (1.796) | (2.799) | (3.565) | (3.538) | |
| Swipe up rate | 0.293 | 0.290 | 0.294 | 0.294 | 0.291 | 0.984 |
| | (0.293) | (0.245) | (0.266) | (0.226) | (0.251) | |
| Unit cost spent | 1.217e-3 | 1.227e-3 | 1.214e-3 | 1.208e-3 | 1.219e-3 | 0.180 |
| | (0.1e-3) | (0.1e-3) | (0.1e-3) | (0.1e-3) | (0.1e-3) | |
| N | 2,412 | 612 | 612 | 594 | 612 | |

[9] A swipe up is equivalent to a click on other online platforms.
[10] We computed the snapchatter mean based on the estimation provided by the platform.
[11] The swipe up rate for Paris was around 0.375% compared to 0.281% for all the other cities.

## 3.2 Impressions

To measure whether the algorithm ultimately could distort the information to which individuals are exposed, we first calculate the distribution of the ad images displayed over time. Figure 2 shows the overall distribution of impressions normalized by the number of Snapchat users in the focal city. Overall, the distribution pattern appears stable over time i.e. the differences in the distribution of the four ads was the same every day. The most widely distributed ad contained the image of a complete male torso, however the ad that was least widely distributed was the ad that displayed a complete female image. Figures 3 and 4 respectively depict the overall distribution of the impressions displayed to women and men. It seems that although the distribution patterns are identical, women received fewer ads.

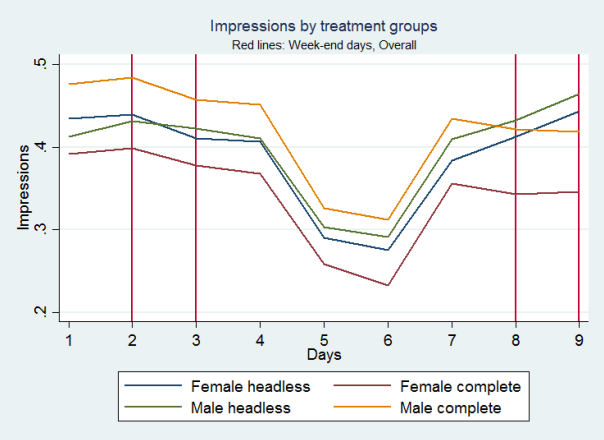


Figure 2: Impressions overall

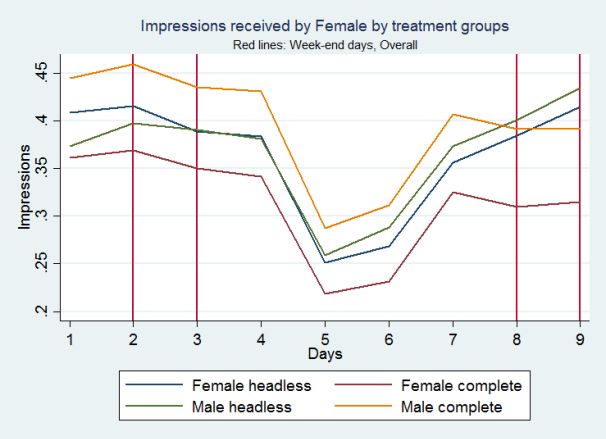


Figure 3: Impressions distributed to women

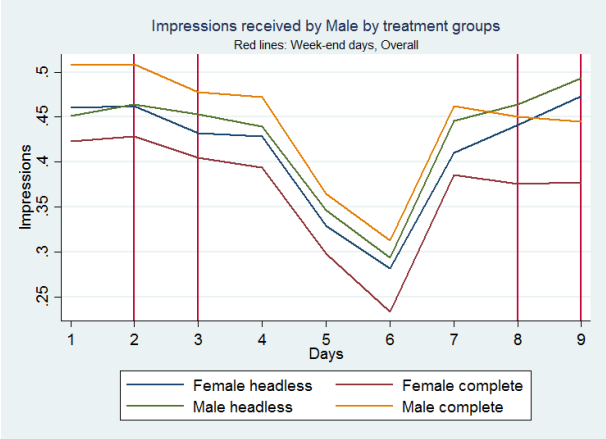


Figure 4: Impressions distributed to men

Figure 5 shows the distribution of the ads in Paris normalized by the average number of Snapchat users. Figure 6 depicts the distribution of pictorial content in the other cities normalized by the average number of snapchatters. The ad distribution patterns for all the ads are similar for all the cities in the experiment; the pattern of distribution of the male complete is similar for the cities in the experiment excluding Paris (see Fig. 5). The female complete photo was the least displayed.

One hypothesis therefore is the reason why there was this distortion in how gender was represented in the ad images was that the algorithm learned from a population-center such as Paris what images appeared to engage users the most on the first day and then replicated that pattern going forward. In the following we investigate this hypothesis by testing differences in swipe up rate between Paris and all other cities.

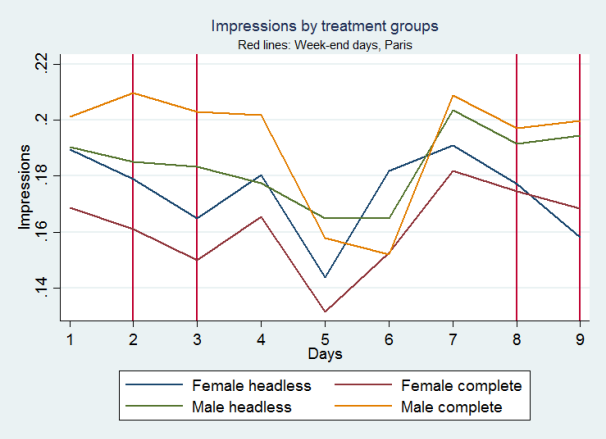


Figure 5: Impressions in Paris

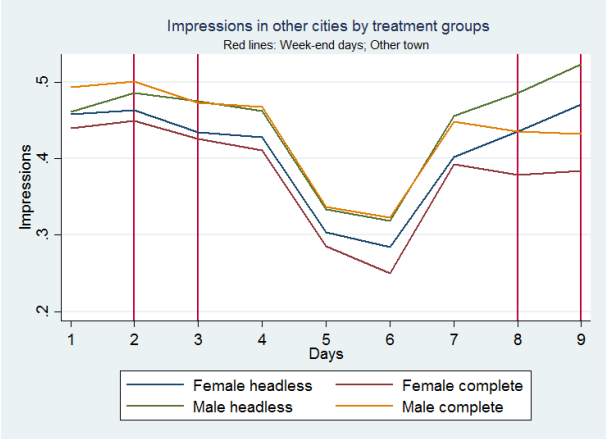


Figure 6: Impressions in other cities

## 3.3 Cities with few and many snapchatters

We analyze whether the algorithm may end up distorting the ad distribution once we consider the number of present snapchatters in each city. Figure 7 shows the overall distribution of impressions normalized by the number of snapchatters in the high end areas of Snapchat users namely more than 1,750. [12] Figure 8 shows the distribution of the number of snapchatters in the low end areas of users (inferior to 1,750). Figures 9 and 10 show respectively the cost per unit spent overall according to the low and high number of snapchatters in the area. Overall, we observe that the gap between the distribution of impressions among treatments is higher for cities where there is a high number of Snapchat users than cities where there is a low number of users (Figure 8). This pattern suggests a higher competition related to advertisers for areas where the number of Snapchat users is important (Figures 9 and 10).

[12]The average number of snapchatters is 2,369 and the median corresponds to 1,750. In Paris, the average number of snapchatters is 4,191 and in the other cities the users' average is 2,104 see Figure 9.

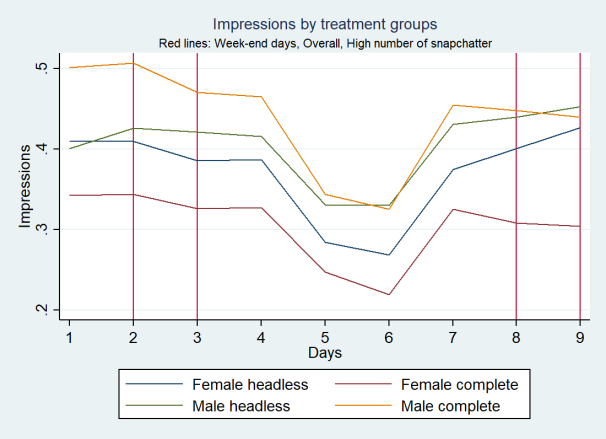


Figure 7: Impressions in cities with high number of snapchatter

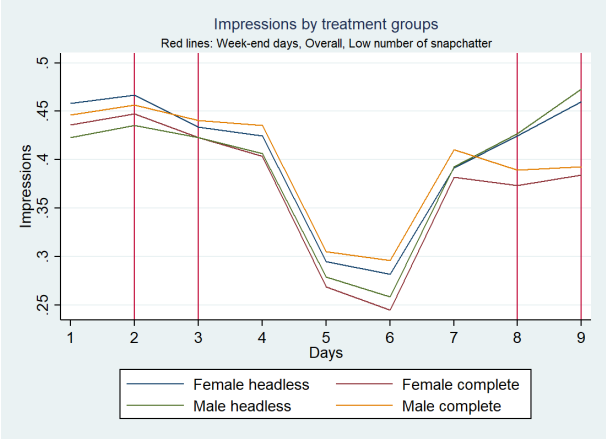


Figure 8: Impressions in cities with low number of snapchatter

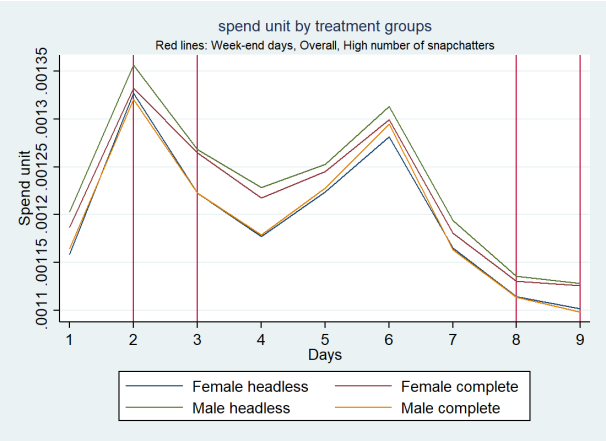


Figure 9: Unit cost spent overall in cities with high number of snapchatter

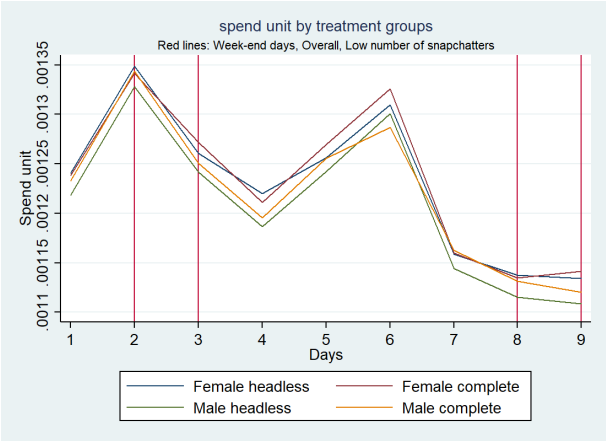


Figure 10: Unit cost spent overall in cities with low number of snapchatter

## 3.4 Swipe-up rates

To explain the distribution differences between ads, we examine whether these differences can be explained by the differences in the swipe up rates for each ad. As discussed above, we observe that though we purchased ads on an impression basis, the ad algorithm appears to reflect engagement too - and therefore reflects initial swipe up rates in its decision about what advertising content to allocate. Figure 11 shows the overall swipe up rate distribution. Though overall swipe up rate of the complete male photo decreases over time as same as in

other cities 13, the algorithm persists in showing this ad.

Similarly, the complete female photo was the least displayed but the swipe up rate related to this photo was overall larger at the beginning of the experimentation and in the last three days. We conclude from this there seems to be no correlation between the display of the complete female photo and the swipe up on the photo. While preferences for the overall campaigns are not drastically different between the complete male photo and the headless photos (male and female), we observe a different pattern for Paris (Fig. 12) wherein snapchatters had a preference for the complete male photo.

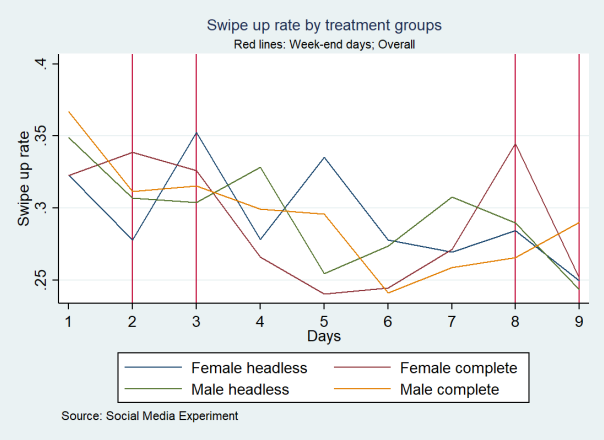


Figure 11: Swipe up rates overall

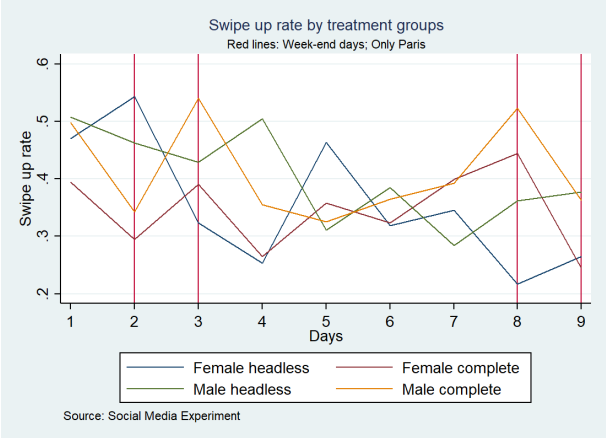


Figure 12: Swipe up rates in Paris

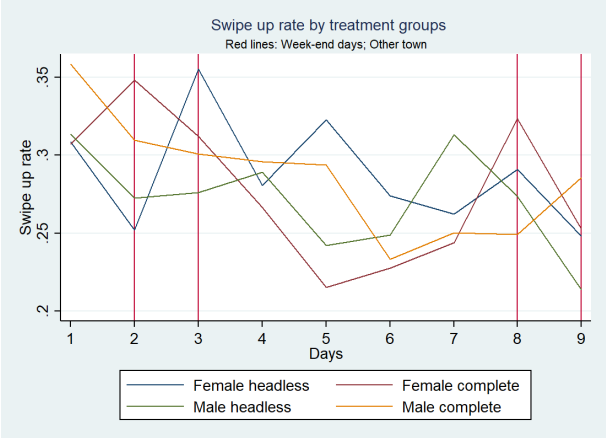


Figure 13: Swipe up rates in other cities

# 4 Results

In this section, we conduct an econometric analysis to understand how the social media algorithm takes account of the ad content in its distribution decision. First, we want to understand whether the ad content influenced the number of impressions, and how Snapchat users reacted to this content. We are interested also in how the algorithm displays the different treatments based on the target group gender. We also emphasize that the algorithm seems to be calibrated on the preferences of users located in Paris.

## 4.1 Did the ad content have an impact on the distribution?

Here we test the effect of advertising content on the ad display. We propose an OLS regression analysis to estimate the number of impressions by city $i$ and gender $g$, at time $t$:

$$Impressions_{igt} = \alpha + \beta X_{igt} + \eta Z_{ig} + \lambda_t + \epsilon_{igt},$$

where $X_{igt}$ is the vector of the variables for the treatment associated to the observation, $Z_{ig}$ is the vector of gender fixed effects and the average number of snapchatters per city, and $\lambda_t$ is a vector of time fixed effects.

Table 2 reports the results of our regressions. Column (1) provides estimates of the number of impressions overall, column (2) presents estimates of the numbers of impressions for the city of Paris, column (3) presents estimates of the numbers of impressions for all other cities. Columns (4) and (5) respectively show the number of swipe ups and the overall swipe up rate. The coefficients of the variables related to the female photos (with and without head) are negative and significant both overall and for the city of Paris (columns (1) and (2)) suggesting that the male related ad content was displayed more than the female content. However, this pattern changes for the other French towns where male headless was shown

more compared to male complete. Compared to the male photo with head (Male complete), none of the treatments seems to influence the swipe up rate (see columns (5)). This result suggests that the ad distribution by the algorithm is not influenced by individual preferences. The estimated number of swipe ups confirms that swipe ups are related to the number of impressions.

Table 2: OLS estimations on ad performance : Overall, Paris, and other cities

| | (1) Impressions Overall | (2) Impressions Paris | (3) Impressions Other cities | (4) Swipe ups Overall | (5) Swipe up rate Overall |
|---|---|---|---|---|---|
| Female headless | -173.058*** | -154.976*** | -47.700 | -0.377** | 0.010 |
| | (42.517) | (48.501) | (40.405) | (0.147) | (0.014) |
| Female complete | -300.642*** | -175.849*** | -107.588*** | -0.831*** | 0.004 |
| | (36.336) | (50.381) | (29.931) | (0.126) | (0.014) |
| Male headless | -116.610** | -90.255* | 144.937** | -0.267 | 0.006 |
| | (57.227) | (47.513) | (59.812) | (0.173) | (0.014) |
| Constant | 384.259*** | 167.892** | -134.539* | 0.956*** | 0.278*** |
| | (60.309) | (73.535) | (76.700) | (0.210) | (0.020) |
| Snapchatters fixed effects | Yes | Yes | Yes | Yes | Yes |
| Gender fixed effects | Yes | Yes | Yes | Yes | Yes |
| Time fixed effects | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.259 | 0.768 | 0.396 | 0.280 | 0.029 |
| N | 2,412 | 306 | 2,106 | 2,412 | 2,412 |

*Notes*: OLS Estimates. Columns (1), (2) and (3) present estimates of the number of impressions, respectively for the overall sample, Paris and other cities. Columns (4) and (5) present respectively estimates of the number of swipe ups and the swipe up rate overall. Robust standard errors are reported in parentheses. The omitted treatment category is *Male Complete*. Significance levels at 1%; 5%, and 10% are indicated respectively by ***, ** and *.

## 4.2 Did the algorithm learn from ads performance?

The algorithm is supposed to seek to maximize swipe ups by relying on real-time advertising performance data, including impressions and swipe ups. The algorithm can learn from how snapchatters have reacted to the different ads by swiping them up (the number of swipe ups) and for a given distribution (the number of impressions for a given target). This information can be collected progressively during the campaigns of the previous days and be regularly updated by the algorithm. Table 3 reports the results of regressions that include lagged impressions and swipe ups for each treatment. Columns (1), (2) and (3) show the estimates of the number of impressions and they include the lagged number of impressions for each treatment. It suggests that the algorithm considers the number of impressions distributed to each treatment during the day before. Columns (4), (5) and (6) show the estimates of the number of impressions and they include the lagged number of swipe up rate for each treatment. Interestingly, the result suggests that the algorithm did not consider the previous swipe up rates when distributing the ad (see column (4) and column (6)). Only the individuals' preferences of male complete (in Paris) seem to influence the algorithm decision-making.

Table 3: Previous impressions and swipe ups can affect current impressions

| | Lagged number of impressions | | | Lagged number of swipe up rate | | |
|---|---|---|---|---|---|---|
| | Overall | Paris | Other cities | Overall | Paris | Other cities |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Male headless | -52.633 | -0.795 | -49.537 | -51.388 | 25.326 | 149.002 |
| | (35.665) | (38.594) | (35.664) | (79.764) | (99.562) | (78.618) |
| Female headless | -121.429** | 26.412 | -124.425** | -202.237*** | -3.715 | -95.745 |
| | (50.808) | (47.949) | (51.873) | (62.318) | (96.231) | (59.982) |
| Female complete | -12.941 | -15.090 | -16.031 | -291.495*** | 19.793 | -151.868*** |
| | (20.713) | (41.490) | (21.151) | (56.487) | (98.571) | (50.746) |
| Male complete impressions$_{t-1}$ | 0.958*** | 0.913*** | 0.960*** | | | |
| | (0.024) | (0.042) | (0.027) | | | |
| Female complete impressions$_{t-1}$ | 0.964*** | 0.921*** | 0.968*** | | | |
| | (0.017) | (0.040) | (0.024) | | | |
| Female headless impressions$_{t-1}$ | 1.122*** | 0.830*** | 1.126*** | | | |
| | (0.070) | (0.062) | (0.072) | | | |
| Male headless impressions$_{t-1}$ | 1.048*** | 0.909*** | 1.050*** | | | |
| | (0.045) | (0.031) | (0.046) | | | |
| Male complete swipe up rate$_{t-1}$ | | | | -225.657** | 373.678* | -246.927** |
| | | | | (99.047) | (202.331) | (106.305) |
| Female complete swipe up rate$_{t-1}$ | | | | -227.473*** | -110.657* | -79.772 |
| | | | | (57.160) | (61.071) | (57.807) |
| Male headless swipe up rate$_{t-1}$ | | | | -372.011*** | 100.561 | -202.542** |
| | | | | (79.119) | (74.689) | (80.087) |
| Female headless swipe up rate$_{t-1}$ | | | | -92.758 | 2.635 | -55.980 |
| | | | | (63.299) | (49.013) | (69.409) |
| Constant | 50.263** | -15.839 | 57.744** | 461.974*** | -12.579 | -35.333 |
| | (22.309) | (40.976) | (24.373) | (69.271) | (107.685) | (81.625) |
| Snapchatters fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Gender fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.940 | 0.956 | 0.940 | 0.257 | 0.774 | 0.386 |
| N | 2,144 | 272 | 1,872 | 2,144 | 272 | 1,872 |

*Notes*: OLS Estimates. Dependent variable is the number of impressions. Columns (1), (4) present estimates of the number of impressions. Columns (2) and (5) present estimates of the number of impressions only in Paris. Columns (3) and (6) present estimates the number of impressions for other cities. Robust standard errors reported in parentheses. Omitted treatment category is *Male Complete*. Significance at 1%; 5% and 10% indicated respectively by ***, ** and *.

## 4.3 Did the number of snapchatters matter?

We are interested here in the decisions of the algorithm according to the number of snapchatters in the city. To investigate this question, we distinguish cities with high number of snapchatters from cities with low number of snapchatters. When the number of users is small, one can expect decisions from the algorithm that do not depend on user preferences. Indeed, in this case the data on which the algorithm can rely are a priori insufficient and one can therefore anticipate a different behavior of the algorithm according to the number of users. Columns (1), (2), and (3) show the regressions of the number of impressions in areas with low numbers of snapchatters (less than 1,750 users for a given target [13]), distinguishing Paris (column 2) from other cities (column 3). Columns (4), (5), and (6) show the regressions of the number of impressions in low density areas of snapchatters (more than 1,750 users). We show that the algorithm behaves differently depending on the number of users in a given target.

[13]The average number of snapchatters is 2,369 and the median corresponds to 1,750.

Table 4: Estimations of the number of impressions in area with low and high number of users

| | Low numbers of users | | | High numbers of users | | |
|---|---|---|---|---|---|---|
| | Overall | Paris | Other cities | Overall | Paris | Other cities |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female headless | -2.415 | -211.630*** | -2.473 | -433.579*** | -701.249*** | -398.298*** |
| | (23.207) | (20.330) | (23.210) | (70.375) | (77.788) | (77.537) |
| Male headless | -103.211*** | 12.644 | -103.054*** | -90.945 | -307.587*** | 55.778 |
| | (21.962) | (31.981) | (21.949) | (91.640) | (78.009) | (117.643) |
| Female complete | -11.152 | -223.467*** | 2.706 | -574.507*** | -559.810*** | -536.996*** |
| | (24.564) | (32.744) | (24.487) | (61.959) | (81.042) | (68.589) |
| Constant | 540.761*** | 611.126*** | 541.632*** | 1260.806*** | 1067.508*** | 1353.340*** |
| | (30.196) | (42.266) | (30.148) | (93.209) | (79.775) | (108.178) |
| Time fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Gender fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.233 | 0.407 | 0.239 | 0.091 | 0.479 | 0.077 |
| N | 783 | 135 | 774 | 1,629 | 171 | 1,332 |

*Note:* OLS Estimates. Columns (1) to (3) present estimates of the number of impressions for cities or districts with the low number of snapchatters. Columns (4) to (6) present estimates of the number of impressions for cities or districts with high number of snapchatters. Robust standard errors reported in parentheses. Omitted treatment category is *Male Complete*. Significance at 1%; 5% and 10% indicated respectively by ***, ** and *.

## 4.4 Did Paris data drive the algorithm?

To further explore the evidence that ad algorithm is driven by preferences in Paris, we introduce a set of interaction effects between each treatment and Paris. Table 5 presents the estimates. We observe that the *Female complete* photo is always less displayed compared to the male complete photo.

Table 5: Algorithm decision to target Paris

| | **Impressions** |
|---|---|
| | Overall |
| Female complete | -398.513*** |
| | (49.725) |
| Female headless | -301.303*** |
| | (56.049) |
| Male headless | -171.986** |
| | (72.664) |
| Paris | -8.069 |
| | (72.307) |
| Female complete × Paris | -142.329* |
| | (79.697) |
| Female headless × Paris | -266.873*** |
| | (80.777) |
| Male headless × Paris | -76.662 |
| | (93.948) |
| Constant | 974.083*** |
| | (68.389) |
| Time fixed effects | Yes |
| Gender fixed effects | Yes |
| R-squared | 0.126 |
| N | 2,412 |

*Note:* OLS Estimates. Dependent variable is the number of impressions. Robust standard errors reported in parentheses. Omitted treatment category is *Male Complete*. Significance at 1%; 5% and 10% indicated respectively by ***, ** and *.

## 4.5 Do price differences explain the distribution of impressions?

Another potential explanation for our results is that the ad of complete male is cheaper to display. Thus, we tested whether differences in ad display among cities can be explained by the different amounts spent on the ad campaign. According to the figures below, it seems that difference in ad costs do not affect or explain the displays of the photos. The graphical evidence suggests that there are no significant differences in the ad costs associated to each treatment. In particular, Figure 15 suggests that, after the 5th day of the campaign, the cost related to showing the male complete photo compared to the other portrayals, was slightly higher in Paris.

To investigate this in more depth, we ran a robustness check (see table 6). We estimated the impact of each treatment on the related unit cost. Column (1) presents estimations for the overall sample; column (2) presents the unit cost estimations for the city of Paris and column (3 ) presents the same costs for the other cities. Overall, the male complete photo was the cheapest; however, in the case of Paris, the headless content (women and men) costs the same as the male complete photo. Only the female complete photo was significantly less expensive than the male complete photo. For the other cities, the female complete photo was the most expensive to advertise compared to the male complete photo. These findings suggest that the algorithm decides about the ad distribution based on the Paris preferences and regardless of the costs of the ad.

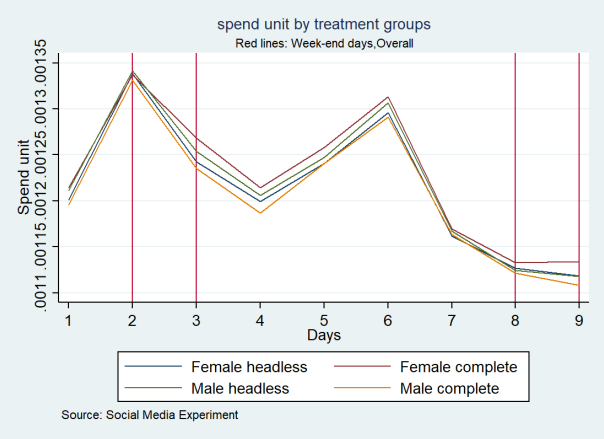


Figure 14: Unit cost spent for impressions overall

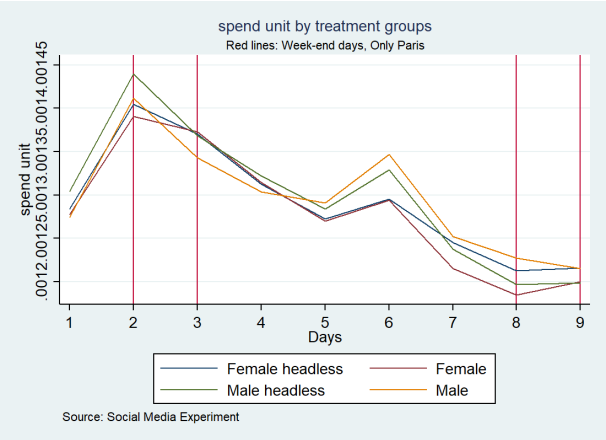


Figure 15: Unit cost spent for impressions in Paris

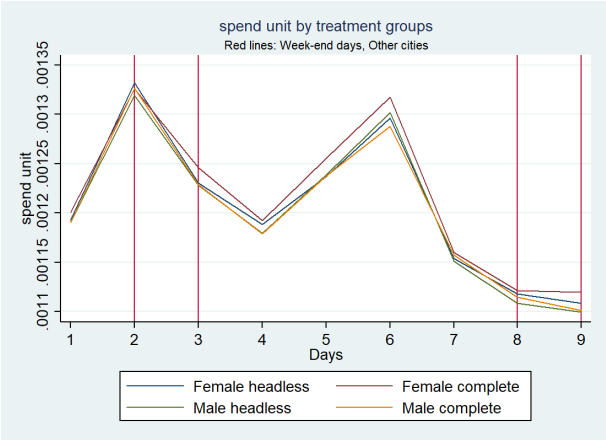


Figure 16: Unit cost spent for impressions in other cities

Table 6: Unit cost spent for one impression

| | Overall | Paris | Other cities |
|---|---|---|---|
| Female headless | 0.015*** | -0.011 | 0.006** |
| | (0.003) | (0.010) | (0.003) |
| Female complete | 0.026*** | -0.021** | 0.016*** |
| | (0.003) | (0.009) | (0.003) |
| Male headless | 0.015*** | -0.000 | 0.002 |
| | (0.003) | (0.007) | (0.003) |
| Constant | 1.198*** | 1.324*** | 1.218*** |
| | (0.005) | (0.015) | (0.005) |
| Snapchatters fixed effects | Yes | Yes | Yes |
| Gender fixed effects | Yes | Yes | Yes |
| Time fixed effects | Yes | Yes | Yes |
| R-squared | 0.707 | 0.767 | 0.742 |
| N | 2,412 | 306 | 2,106 |

*Note:* OLS estimates. The table presents estimates of the unit cost spend. Robust standard errors reported in parentheses. Omitted treatment category is *Male Complete*. Significance at 1%; 5% and 10% indicated respectively by ***, ** and *.

# 5 Conclusion

This paper explores how the ad distribution algorithm of an online social media can be influenced by the content of ads. We examined how four ads with different photos were displayed in French cities across time. We found that while there is no significant relationship overall between the type of photo and the swipe up rate, the male related ad content was displayed more compared to the female complete ad. We explore why this occurred and present suggestive evidence that because the algorithm tries to learn quickly what content is most engaging and then replicate this pattern it based its allocation of impressions on swipe up rates in Paris - which has a far larger population than the other cities in our sample.

Our article shows that the attempt of advertising algorithms to try and learn quickly what content is most engaging can lead to advertising results that might seem (at first glance) to be less effective, and which appear unsettling - such as the algorithmic decision to show photos of headless females rather than females with heads. The results of our experiment

highlight the sensitivity of online advertising distribution to elements that advertisers cannot anticipate including the treatment of advertising content by algorithms. We want to alert ad platforms and policy makers to the effects that standard mechanism by which algorithms learn can have and inadvertently generalize from population centers to the rest of a country.

# References


Bergemann, D. and Välimäki, J. (2006). *Bandit Problems*. Cowles Foundation Discussion Papers 1551. Cowles Foundation for Research in Economics, Yale University.

Cecere, G., Jean, C., Le Guel, F. and Manant, M. (2018). *STEM and teens: An algorithm bias on a social media*. Working paper.

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

Doleac, J. L. and Stein, L. C. (2013). The visible hand: Race and online market outcomes. *The Economic Journal*. 123(572), F469–F492.

Edelman, B., Luca, M. and Svirsky, D. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*. 9(2), 1–22.

Fisman, R. and Luca, M. (2016). Fixing Discrimination in Online Marketplaces. *Harvard Business Review*.

Johnson, G. A., Lewis, R. A. and Nubbemeyer, E. I. (2017). Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness. *Journal of Marketing Research*. 54(6), 867–884.

Lambrecht, A. and Tucker, C. E. (2018). Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science, forthcoming*.

Manant, M., Pajak, S. and Soulié, N. (2018). Can social media lead to labor market discrimi-

nation? Evidence from a field experiment. *Journal of Economics & Management Strategy, forthcoming*.

Pew Research Center, P. (2018a). Social Media Use in 2018. March.

Pew Research Center, P. (2018b). Teens, Social Media and Technology 2018. March.

Robbins, H. (1985). Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*. (pp. 169–177). Springer.

Schwartz, E. M. (2013). Optimizing adaptive marketing experiments with the multi-armed bandit.

Schwartz, E. M., Bradlow, E. T. and Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*. 36(4), 500–522.

SnapInc (2018). *Reports First Quarter 2018 Results*. Https://investor.snap.com/news-releases/2018/05-01-2018-211516272.

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

Tran-Thanh, L., Stein, S., Rogers, A. and Jennings, N. R. (2014). Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. *Artificial Intelligence*. 214, 89 – 111.

# 6 Appendix

## 6.1 Analysis of the ad photos by Google Cloud Vision

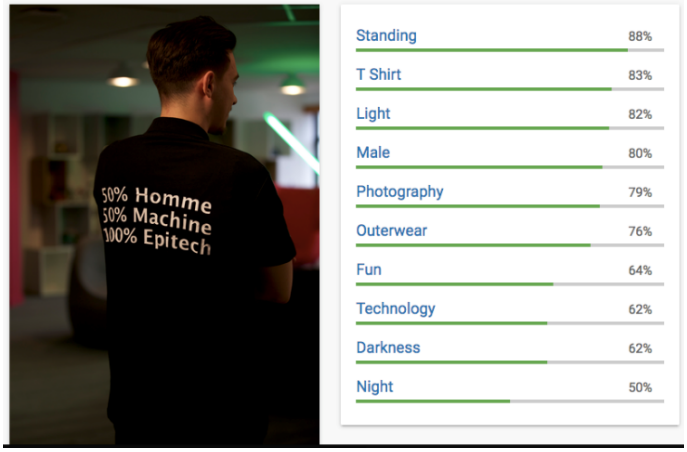


Figure 17: Male complete analysis

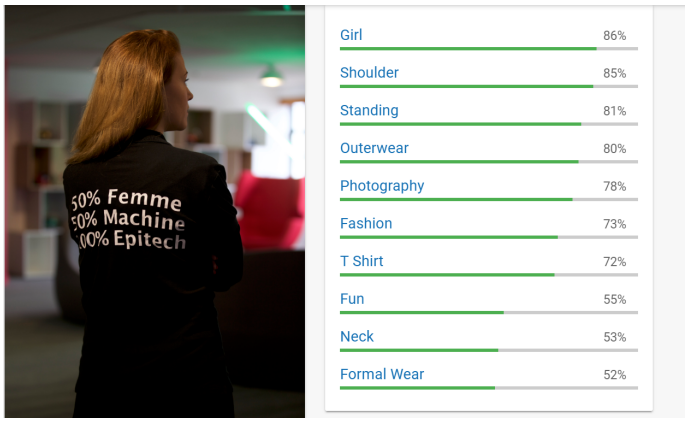


Figure 18: Female complete analysis

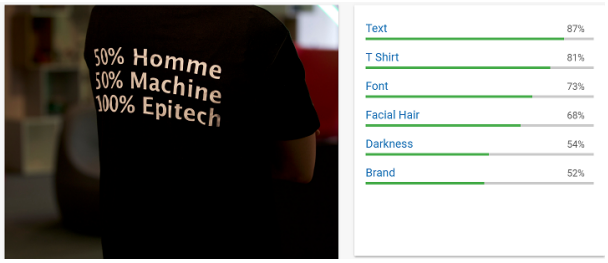


Figure 19: Male headless analysis

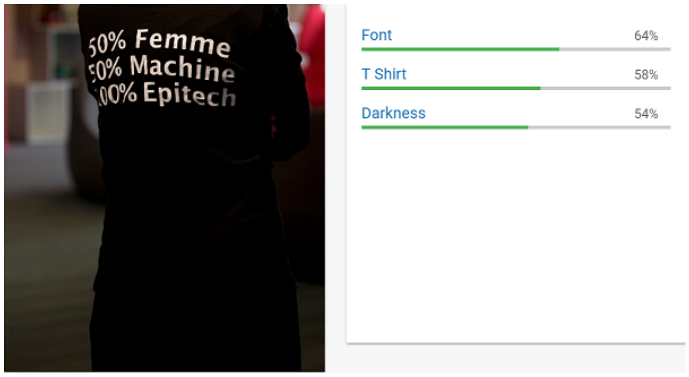


Figure 20: Female headless analysis

## 6.2 Results of randomization procedure

We randomized four groups of cities that received unique photos. The randomization procedure is based on taking account of these cities' socioeconomic characteristics. Table 7 presents the average means for each group, and the results of the F-tests that show that

Table 7: **Results of randomization procedure**

| | **Complete female N=34** | | **Headless female N=34** | | **Complete male N=33** | | **Headless male N=33** | | **F-test** |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | (p-values) |
| Hourly wage | 14.878 | (2.791) | 14.741 | (2.442) | 14.801 | (2.544) | 15.114 | (2.999) | 0.947 |
| Senior managers' hourly wage | 25.282 | (2.733) | 25.059 | (2.408) | 25.172 | (2.526) | 25.467 | (2.979) | 0.935 |
| Middle managers' hourly wage | 14.830 | (0.944) | 14.807 | (0.828) | 14.810 | (0.876) | 15.005 | (0.939) | 0.773 |
| Employees' hourly wage | 10.803 | (0.664) | 10.811 | (0.623) | 10.780 | (0.643) | 10.846 | (0.747) | 0.983 |
| Hourly wage worker | 11.360 | (0.663) | 11.386 | (0.680) | 11.385 | (0.589) | 11.551 | (0.696) | 0.622 |
| Women's hourly wage | 13.191 | (2.445) | 13.040 | (2.026) | 13.110 | (2.198) | 13.297 | (2.577) | 0.973 |
| Senior executive women's hourly wage | 21.583 | (2.348) | 21.453 | (1.985) | 21.512 | (2.151) | 21.689 | (2.473) | 0.977 |
| Middle manager women's hourly wage | 13.672 | (0.986) | 13.681 | (0.880) | 13.697 | (0.940) | 13.747 | (1.021) | 0.989 |
| Women's hourly wage employee | 10.607 | (0.733) | 10.601 | (0.660) | 10.589 | (0.719) | 10.644 | (0.805) | 0.992 |
| Women's hourly wage worker | 10.040 | (0.799) | 9.9 | (0.734) | 9.925 | (0.710) | 10.141 | (0.848) | 0.632 |
| Men's hourly wage | 16.087 | (3.126) | 15.936 | (2.814) | 15.998 | (2.872) | 16.388 | (3.415) | 0.935 |
| Men's hourly wage senior executive | 27.187 | (3.371) | 26.835 | (2.918) | 27.023 | (3.058) | 27.368 | (3.648) | 0.921 |
| Middle manager men's hourly wage | 15.736 | (1.034) | 15.681 | (0.892) | 15.684 | (0.940) | 15.953 | (1.049) | 0.635 |
| Men hourly wage employee | 11.270 | (0.497) | 11.304 | (0.542) | 11.230 | (0.478) | 11.329 | (0.612) | 0.887 |
| Men hourly wage worker | 11.621 | (0.650) | 11.666 | (0.693) | 11.653 | (0.589) | 11.824 | (0.682) | 0.593 |
| 18-25 hourly wage | 9.823 | (0.642) | 9.772 | (0.493) | 9.775 | (0.542) | 9.882 | (0.615) | 0.848 |
| 26-50 hourly wage | 14.650 | (2.688) | 14.501 | (2.304) | 14.560 | (2.452) | 14.846 | (2.873) | 0.952 |
| 50+ hourly wage | 17.754 | (3.922) | 17.539 | (3.532) | 17.639 | (3.603) | 18.060 | (4.192) | 0.950 |
| 18-25 women hourly wage | 9.463 | (0.628) | 9.373 | (0.493) | 9.409 | (0.542) | 9.471 | (0.636) | 0.884 |
| 26-50 women hourly wage | 13.274 | (2.525) | 13.123 | (2.080) | 13.205 | (2.277) | 13.364 | (2.664) | 0.980 |
| 50+ women hourly wage | 14.712 | (3.017) | 14.528 | (2.536) | 14.587 | (2.717) | 14.860 | (3.160) | 0.967 |
| 18-25 men hourly wage | 10.111 | (0.698) | 10.094 | (0.532) | 10.071 | (0.580) | 10.209 | (0.646) | 0.812 |
| 26-50 men hourly wage | 15.611 | (2.863) | 15.448 | (2.523) | 15.499 | (2.634) | 15.858 | (3.112) | 0.934 |
| 50+ men hourly wage | 19.941 | (4.716) | 19.639 | (4.331) | 19.779 | (4.357) | 20.313 | (5.070) | 0.941 |
| Teens 15-17 | 819.734 | (1257.662) | 1027.338 | (1457.429) | 1137.278 | (2063.484) | 1125.427 | (1728.995) | 0.849 |
| Teens 18-24 | 3063.610 | (4893.576) | 4753.539 | (9595.086) | 4618.144 | (10201.720) | 4102.805 | (6374.108) | 0.821 |
| High schoolers 18-24 | 2165.978 | (3704.839) | 3359.735 | (7373.131) | 3124.486 | (7385.333) | 2841.158 | (4640.866) | 0.860 |
| High schoolers 15-17 | 798.594 | (1231.321) | 986.370 | (1393.913) | 1085.640 | (1964.695) | 1089.338 | (1678.620) | 0.863 |

*Notes*: The table reports overall mean estimates for the 2015 administrative data for each treatment group. The last column shows the p-values of the computed F-statistic showing balanced groups in our sample.

Table 8: Descriptive statistics

| | Mean | Std.Dev. | Min | Max | N |
|---|---|---|---|---|---|
| Impressions | 901.539 | (917.839) | 121 | 9539 | 2,412 |
| Female complete | 0.254 | - | 0 | 1 | 2,412 |
| Male complete | 0.246 | - | 0 | 1 | 2,412 |
| Female headless | 0.254 | - | 0 | 1 | 2,412 |
| Male headless | 0.246 | - | 0 | 1 | 2,412 |
| Women targeting | 0.500 | - | 0 | 1 | 2,412 |
| Avg. snapchatters | 2369.403 | (1368.471) | 1250 | 8750 | 2,412 |

*Note:* On average, 901 impressions were displayed by day and city. Each of our treatment represents about 25% of our sample, 50% of snapchatters who were targeted were women and there is an average number of 2369 snapchatters by location.

Table 9: Descriptive statistics: Paris vs Other cities

| | **Paris** | | | | | **Other cities** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Min | Max | N | Mean | Std.Dev. | Min | Max | N |
| Impressions | 751.284 | (401.082) | 121 | 2059 | 306 | 923.371 | (968.412) | 125 | 9539 | 2,106 |
| Female complete | 0.353 | - | 0 | 1 | 306 | 0.239 | - | 0 | 1 | 2,106 |
| Male complete | 0.118 | - | 0 | 1 | 306 | 0.265 | - | 0 | 1 | 2,106 |
| Female headless | 0.176 | - | 0 | 1 | 306 | 0.265 | - | 0 | 1 | 2,106 |
| Male headless | 0.353 | - | 0 | 1 | 306 | 0.231 | - | 0 | 1 | 2,106 |
| Women targeting | 0.500 | - | 0 | 1 | 306 | 0.500 | - | 0 | 1 | 2,106 |
| Avg. Snapchatters | 4191.176 | (1742.376) | 1250 | 8250 | 306 | 2104.701 | (1073.551) | 1250 | 8750 | 2,106 |

*Note:* We present the full descriptive statistics for the subsamples of Paris and other cities.