

Les algorithmes de réseaux sociaux préfèrent les femmes sans tête*

Projet d'article de recherche

Grazia Cecere¹, Clara Jean^{2,3}, Matthieu Manant³, and Catherine Tucker⁴

¹Institut Mines Télécom, Business School, ²Epitech, ³Université Paris-Sud, Paris-Saclay, ⁴MIT
Sloan School of Management

20 novembre 2018

1 Contexte : réseaux sociaux, publicité et algorithmes

Les adolescents utilisent aujourd'hui largement les réseaux sociaux pour trouver des informations, comme les actualités ou, plus pratiquement, des informations liées à leurs choix d'études ou de carrière. Sur ces plateformes en ligne, les algorithmes jouent un rôle clé dans l'accès à l'information des utilisateurs. Les campagnes publicitaires sont un exemple de distribution d'informations gérées par des algorithmes qui optimisent l'efficacité des campagnes publicitaires en personnalisant la distribution de la publicité selon les profils des utilisateurs. Néanmoins, il a été montré que la distribution des publicités peut être biaisée en raison de l'apprentissage par les algorithmes basé sur le comportement d'autres annonceurs.¹ Parallèlement, une littérature émergente montre l'existence de discriminations en ligne,² ainsi, il est important de comprendre si les algorithmes peuvent aider à réduire les discriminations ou au contraire les exacerber. Dans cette étude, nous nous intéressons à comment le contenu d'une campagne publicitaire peut amener un algorithme à distribuer différents types de publicités à différents groupes d'une façon qui peut apparaître discriminante. Ce travail fait suite à une expérimentation sur les biais des algorithmes dans un contexte publicitaire pour une publicité sur les filières STEM.³

2 Méthodologie

Nous testons l'effet de contenus "féminins" sur la distribution de publicités en menant des campagnes publicitaires sur la plateforme Snapchat pour le compte d'une école d'informatique française. Les visuels des publicités testés représentent des individus de dos avec un message publicitaire pour l'école visible sur leurs t-shirts. Pour tester la réaction de l'algorithme aux photos, nous adoptons un design 2×2 avec 4 traitements où nous faisons varier le genre des individus (femme / homme) et le type de photo (avec / sans tête) : les deux premières photos d'une jeune femme (Fig. 1(a)) et d'un jeune homme (Fig. 1(b)) permettent clairement de reconnaître leur genre, alors que ce n'est pas le cas pour les deux autres photos qui montrent les mêmes individus sans tête (Fig. 1(c) et Fig. 1(d)), rendant l'identification de leur genre plus subjective. Pour avoir une intuition sur l'interprétation

*Ce papier a été présenté à la 2018 MIT CODE (session "Health, Gender, and Education") et sera présenté au 2019 ASSA Annual Meeting de l'American Economic Association. L'expérimentation a été financée par la MSH Paris-Saclay dans le cadre du projet Maturation 17-MA-04. Contact : clara.jean@u-psud.fr

1. A. LAMBRECHT et C. E. TUCKER (2018). « Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads ». *Management Science*, à venir.

2. R. FISMAN et M. LUCA (2016). « Fixing Discrimination in Online Marketplaces ». *Harvard Business Review*.

3. G. CECERE et al. (2018). *STEM and teens : An algorithm bias on a social media*. Papier de travail SSRN.

de nos photos par l’algorithme, nous les avons testées en amont en utilisant l’outil en ligne Google Cloud Plarform qui utilise de l’intelligence artificielle pour catégoriser des photos. L’algorithme de Google identifie clairement le genre des individus sur les photos avec les têtes, mais ne parvient pas à les identifier sur les photos sans les têtes.



(a) Femme entière

(b) Homme entier

(c) Femme sans tête

(d) Homme sans tête

Les campagnes publicitaires ont duré 9 jours consécutifs ; nous avons ciblé des lycéens de 16 à 19 ans dans 134 villes françaises, chaque ville étant préalablement associée à une photo suite à une procédure de randomisation conduite ex-ante. Nous avons distingué les campagnes où nous avons ciblé des filles des campagnes où nous avons ciblé des garçons.⁴ Au total, nous avons donc mené simultanément 268 campagnes publicitaires différentes (deux campagnes par ville, une ciblant les filles, une autre ciblant les garçons).

3 Résultats

Nous montrons que : (i) la photo avec la femme entière a été la photo la moins distribuée aux utilisateurs par rapport aux autres photos, (ii) les contenus masculins (avec ou sans tête) sont plus distribués que les contenus féminins, et (iii) le *swipe up rate*⁵ réalisé à Paris le premier jour de la campagne influence la distribution globale de la publicité dans les autres villes en France, quelque soient les préférences des utilisateurs pour les différents photos dans ces villes. Notre travail s’inscrit dans la littérature sur la publicité en ligne et la littérature sur les biais des algorithmes. Notre article montre en particulier comment le contenu publicitaire lui-même peut conduire à des résultats publicitaires qui semblent (à première vue) être moins efficaces et qui s’apparentent à de la discrimination.

Les résultats de notre expérimentation mettent en évidence la sensibilité de la distribution des publicités en ligne à des éléments que les annonceurs ne peuvent anticiper, notamment le traitement du contenu des publicités par les algorithmes. Par ce travail, nous souhaitons susciter l’intérêt des détenteurs des réseaux sociaux et des décideurs politiques sur les dérives que peuvent avoir les algorithmes de ces plateformes, réduisant l’efficacité du point de vue de l’annonceur et l’inaccès à l’information du côté de l’utilisateur.

4 Annexes

4. Pour des raisons techniques nous avons dû conduire séparément les campagnes par genre

5. Ratio du nombre de swipes ou clics divisés par le nombre d’impressions

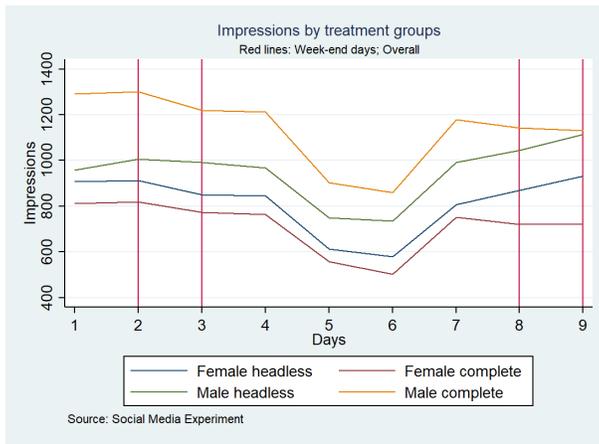


FIGURE 1: Nombre global d'impressions par traitement

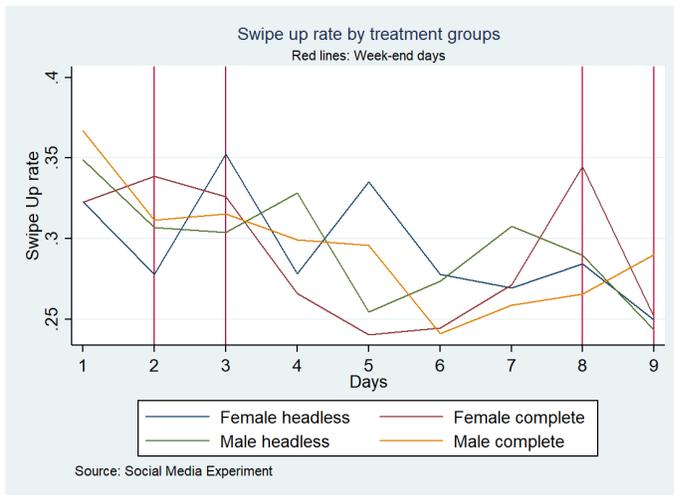


FIGURE 2: Swipe up ratio par traitement

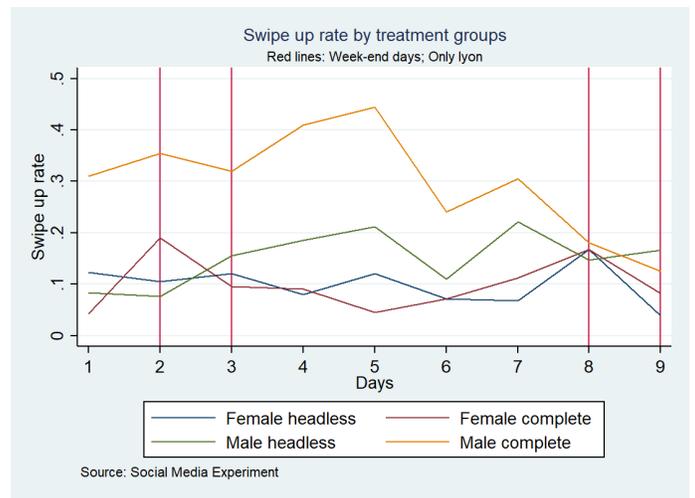


FIGURE 3: Swipe up ratio par traitement pour **Lyon**

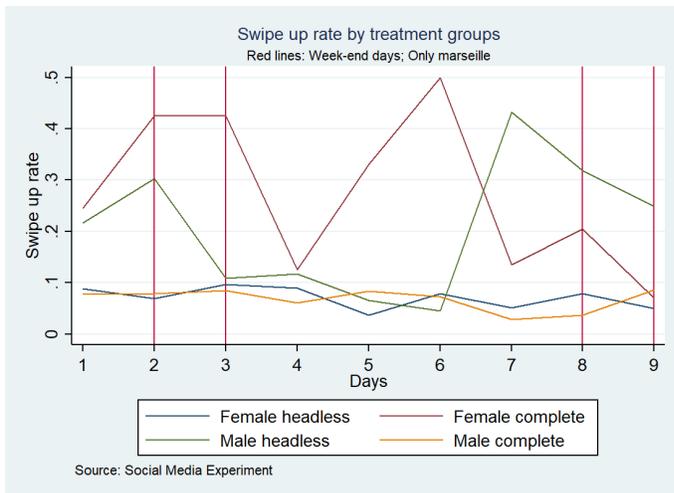


FIGURE 4: Swipe up ratio par traitement pour **Marseille**

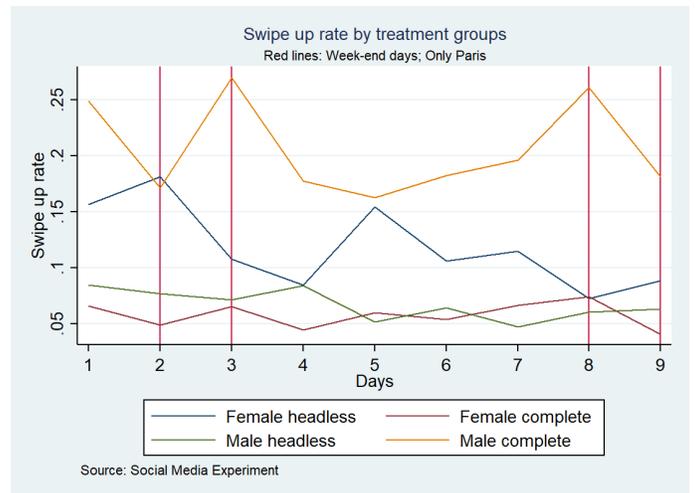


FIGURE 5: Swipe up ratio par traitement pour **Paris**