

Field studies of psychologically targeted ads face threats to internal validity

Dean Eckles^{a,b,1,2}, Brett R. Gordon^{c,2}, Garrett A. Johnson^{c,2}

^a Sloan School of Management and ^b Institute for Data, Systems & Society, Massachusetts Institute of Technology;

^c Kellogg School of Management, Northwestern University

We applaud Matz et al. (1) for using field data on Facebook to investigate the effectiveness of psychologically-targeted messages, building on prior, largely laboratory-based research. That said, this comparison of Facebook ad campaigns does not randomly assign users to conditions, which threatens the internal validity of their findings and weakens their conclusions.

The paper uses Facebook's standard ad platform to compare how different versions of ads perform. However, this process does not create a randomized experiment: users are not randomly assigned to different ads, and individuals may even receive multiple ad types (e.g., both extroverted and introverted ads). Furthermore, ad platforms like Facebook optimize campaign performance by showing ads to users whom the platform expects are more likely to fulfill the campaign's objective (e.g., Study 1 — online purchases, Studies 2 & 3 — app installs). This optimization generates differences in the set of users exposed to each ad type, so that differences in responses across ads do not by themselves indicate a causal effect.

Accordingly, statistical tests reveal differences across ad type conditions by comparing all reported pre-existing user characteristics. Using the authors' provided data, in Study 1, we find differences in the age distribution (Fig. 1A) among the "extroverted" users exposed to congruent

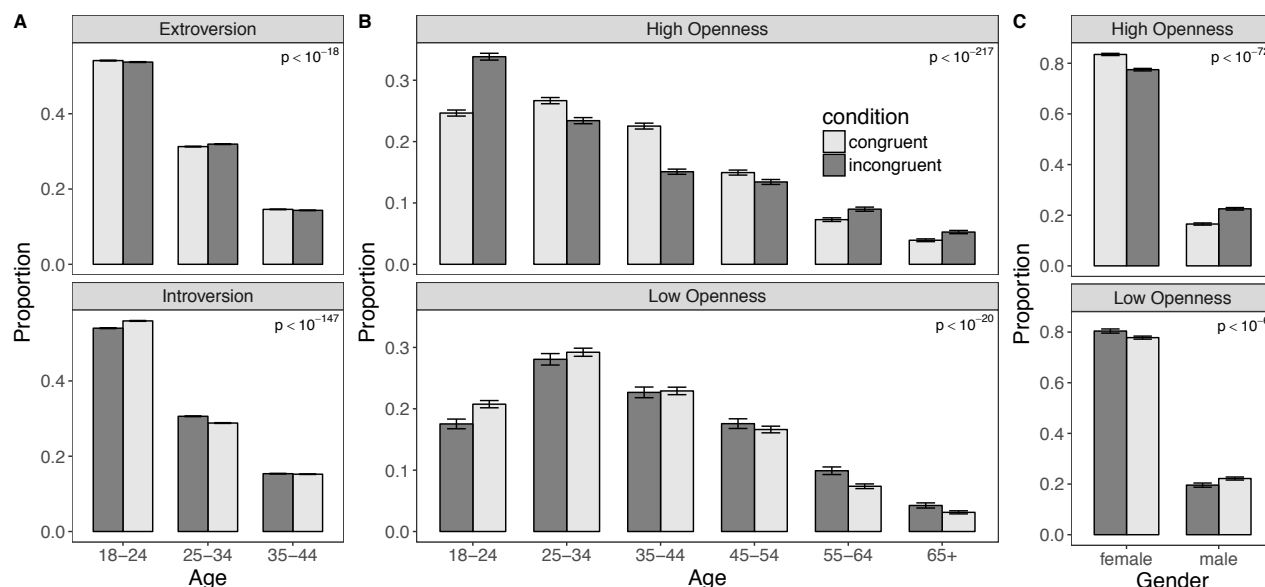


Fig. 1. Distribution of all reported user characteristics: age in Study 1 (A), and age (B) and gender (C) in Study 2. Error bars are 95% confidence intervals. Displayed p-values are from χ^2 tests of independence.

¹ To whom correspondence should be addressed. Email: eckles@mit.edu

² All authors contributed equally and are listed in alphabetical order.

versus incongruent ads, $\chi^2(2) = 84.5$, $p < 10^{-18}$, and also for the "introverted" users, $p < 10^{-147}$. In Study 2, we find similar differences in age (Fig. 1B) among "high-openness" users, $p < 10^{-217}$, and "low-openness" users, $p < 10^{-20}$, as well as differences in gender (Fig. 1C), $p < 10^{-72}$ and $p < 10^{-6}$, respectively. (Study 3 reports no user characteristics. Code for this analysis is archived at <https://osf.io/y8nuk>.) More importantly, these differences in observables indicate potential differences in unobserved characteristics between exposed groups.

Could such confounding affect the results? In Matz et al.'s (1) Studies 1, 2, and 3, the reported 95% confidence intervals for the relative increase in conversions are [25%, 90%], [9%, 58%], and [5%, 27%] respectively. Thus, biases of 25%, 9%, and 5% are sufficient to account for these results even without differences in effectiveness. As one comparison point, Gordon et al. (3, section 5.1) document a Facebook study that similarly compares two ads which produces an incorrectly signed effect estimate relative to the experimental baseline; the implied confounding bias is 6%, 95% CI [2%, 11%]. The degree of confounding in Matz et al.'s (1) studies could be larger or smaller. The intent of this comparison is to highlight that confounds in user characteristics and ad exposure might explain a non-negligible portion, if not all, of each study's effects.

While suggestive, the Matz et al. (1) studies provide limited new evidence for the efficacy of psychologically-targeted advertising. Researchers should be aware of the limitations of testing ad campaigns on platforms that optimize ad delivery and do not randomize users into conditions. Since the Matz et al. (1) studies were conducted, some ad platforms, including Facebook, introduced tools for advertisers to conduct randomized experiments (2, 4-5), which may aid future work.

References

1. Matz, SC, Kosinski, M, Nave, G, & Stillwell, DJ (2017) Psychological targeting as an effective approach to digital mass persuasion. *Proc Natl Acad Sci*, 201710966.
2. Johnson, GA, Lewis, RA, & Nubbemeyer, EI (2017) Ghost ads: Improving the economics of measuring online ad effectiveness. *J Mark Res* 54(6):867-884.
3. Gordon, BR, Zettelmeyer, F, Bhargava, N, & Chapsky, D (2016) A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. Whitepaper. Available at https://www.kellogg.northwestern.edu/faculty/gordon_b/files/kellogg_fb_whitepaper.pdf
4. Facebook. Conversion Lift. <https://www.facebook.com/business/a/conversion-lift>
5. Facebook. Split testing. <https://www.facebook.com/business/measurement/split-testing>

Competing interests statement: D.E. and G.A.J. have been previously employed by Facebook. B.R.G. is a contractor of Facebook; he donates all resulting income to charity. D.E. and B.R.G. have co-authorship relationships with Facebook employees. D.E. has a significant financial interest in Facebook.