The Microtargeting Manipulation Machine

Keywords: psychological microtargeting, personality, persuasion, AI, GPT-3

Extended Abstract

Psychological microtargeting is a marketing tactic that involves using personal data and sophisticated algorithms to create highly specific and tailored messages targeted at individual users. This technique uses data on a person's interests, behavior, and psychological characteristics to create custom-tailored messages that are designed to influence their behavior or opinions. While psychological microtargeting may be considered legitimate in domains such as consumer behavior, in political contexts, psychological microtargeting can be harmful as it can create echo chambers, reinforce existing biases, and manipulate people's emotions and beliefs without their knowledge or consent (Kozyreva et al., 2021). This can lead to the polarization of society and undermine the democratic process by manipulating public opinion and electoral outcomes (Barocas, 2012).

However, the efficacy of psychological microtargeting has been debated. Recent evidence supports the notion that political microtargeting can be effective (Tappin et al., 2022), and converging results provide robust evidence in favor of personality-congruent messages (Joyal-Desmarais et al., 2022). But the extent to which personality-based microtargeting can be effective in political domains remains unclear. The efficacy of such practices is important to understand, as warning of such attempts are often dismissed by the lack of empirical evidence. The current package of studies aimed to address the efficacy of personality-based microtargeting, as part of a larger attempt to call this practice into question. In three studies, we provide evidence that personality-congruent political microtargeting is effective, and more scalable than ever using commercial, off-the-shelf generative AI.

In Study 1, we collected over 1500 political ads from Facebook's Ad Library that targeted UK users (ads were posted in the past five years). We then applied a machine-learning model that predicted how appealing the text is to different personality types (Authors, 2023). Building on the model predictions, we selected 20 ads from the upper decile and 20 ads from lower decile of the predicted Openness to Experience and asked participants (N = 50; Prolific UK) to provide judgments of the perceived Openness of the ad. Our findings corroborated the Model's prediction, as ads in the top decile of predicted Openness received significantly higher Openness judgments than those in the bottom decile (b = 0.7, *p* < .001).

Subsequently in Study 2, we pre-registered the hypothesis that personality-congruent political messages will be more persuasive to individuals. We selected the top 10 performing ads from Study 1 (5 high and 5 low on Openness) and recruited 800 participants from Prolific UK. The participants were requested to rate the ads on their persuasiveness. In the end, participants were requested to complete an Openness questionnaire from BFI-2 (Soto & John, 2017). The results show that indeed participants' Openness moderated the persuasiveness of the ads, such that people who are high on Openness reported high-on-openness ads as more persuasive vs. people who are low on Openness, b = 0.19, *p* = .001 (see Figure 1).

Study 1 has provided evidence that people can accurately judge whether an ad was targeted at a specific personality makeup, and Study 2 showed that people who receive personality-congruent political messages find them to be more persuasive. Next, we set out to understand how scalable this practice is using an off-the-shelf, commercial AI product.

Recently, generative AI has been booming. High quality text generation models are now freely or commercially available for anyone, even without any programming expertise. While

opening such powerful AI tools may bring immense progress, it can also bring colossal harm. One of these potential harms is the scalability of manipulative tactics, such as those enabled by microtargeting.

In Study 3, we fed into GPT-3 (a commercial generative text model) the ads we collected in Study 1. We instructed GPT-3 to generate two versions of each ad, one that appeals to people who are high on Openness to Experience and one that appeals to people who are low on Openness. We did so by providing GPT-3 the definition of Openness to Experience and instructing it to rephrase the ad according to the requested personality type (high vs. low on Openness). We then ran the computational model that predicts how attractive is the generated text on the Openness scale (Authors, 2023). A paired-samples t-test revealed a significant difference such that high-on-openness generated texts received a higher predicted Openness score ($p <$.001, Cohen's d = 0.12), suggesting that GPT-3 and our computational model are aligned.

Finally, we selected 25 ad-pairs from the top 40 pairs that maximized the difference in predicted Openness. We recruited 50 participants through Prolific (UK sample). Each participant was given the same explanation as GPT-3 about what is Openness to Experience and was asked to rate 25 ads (one ad from each ad-pair) on a 1-6 Likert scale ranging from low to high on Openness. A Linear mixed model, predicting the rating given the original prompt classification (high or low on Openness) resulted in a significant effect such that high Openness ads received an average rating higher by 1.45 ($p < $.001).

In sum, the current package of studies aimed to address the efficacy of personality-based microtargeting, showing that personality-congruent political microtargeting is effective and scalable using commercial, off-the-shelf generative AI. The studies demonstrate that people can accurately judge whether an ad was targeted at a specific personality makeup and that individuals who receive personality-congruent political messages are more persuaded by them. The scalability of such practices, particularly with the use of powerful AI tools, can bring immense harm, and thus it is crucial to understand the efficacy of microtargeting to call its practice into question. Our results shed light on these questions and emphasize the critical need for algorithmic transparency in our online environment.

References

Barocas, S. (2012). The price of precision: voter microtargeting and its potential harms to the democratic process. *Proceedings of the First Edition Workshop on Politics, Elections and Data*, 31–36.

Joyal-Desmarais, K., Scharmer, A. K., Madzelan, M. K., See, J. V., Rothman, A. J., & Snyder, M. (2022). Appealing to motivation to change attitudes, intentions, and behavior: A systematic review and meta-analysis of 702 experimental tests of the effects of motivational message matching on persuasion. *Psychological Bulletin*, *148*(7–8), 465–517.

Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications*, *8*(1), 1–11.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143.

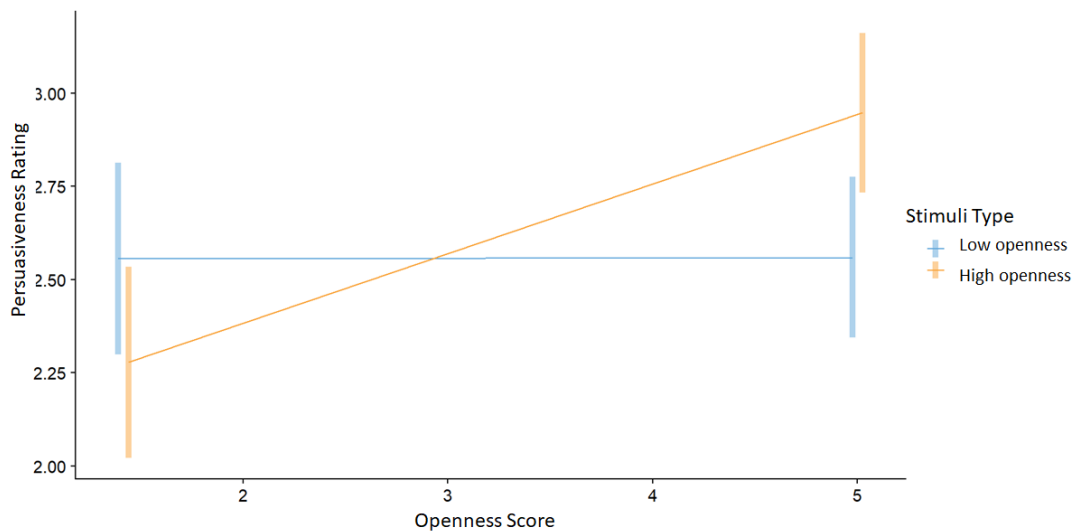Tappin, B. M., Wittenberg, C., Hewitt, L., Berinsky, A., & Rand, D. G. (2022). *Quantifying the Persuasive Returns to Political Microtargeting*. https://doi.org/10.31234/osf.io/dhg6k

Figure 1. The effect of persoanlity-congruent messgaes on persuasion. User's personaliy is plotted on the X axis, persuasivness ratings is plotted on the Y axis. Color denotes the type of stimuli: in blue ads that appeal to low Openness; in orange ads that appeal to high Openness. Error bars debote 95% confidence intervals.