

Letter to Reviewer

Reliable Sources?

Correcting Misinformation in Polarized Media Environments

We thank the editor and reviewer for engaging with our work and providing us with such helpful feedback. We have made a number of changes in response to the comments, which in our view, have considerably strengthened our manuscript. Below, we respond to the reviewer's comments and summarize our corresponding revisions in the manuscript.

1. My main concern is with the methodology and thus the conclusions of the paper. Specifically, I failed to understand how the authors arrive at such conclusions without having a clear and identifiable set of pretreatment questions that map precisely to the post-treatment outcomes. As I understand, the logic behind the experiment is to implement some pretest-posttest to manipulate whether participants are allowed to chose Fox or MSNBC; however, there are no pretreatment questions that could serve as a baseline. If, however, the objective of the paper is to test the manipulation without a baseline, then it needs to refine and clarify the language of the research design that is highly confusing at the very least. If the objective of the experiment is to test the manipulation (freely to choose or assigned to a media outlet), the distribution between treatment and control groups are everything but balanced (I assume that the table in Appendix A describes the sample after randomization), which makes me question if the differences that exist between treatment conditions will on average disappear after randomization, the distributions are all over the place. Simply by looking at table I and without more information, it is impossible to ascertain if the randomization worked or not as it is presented; the data suggests that there are significant problems. For example, given that there seem to be substantial differences between groups, how can we be sure that there is about a 20 to 30 pp difference?

We agree that the discussion of our research design and analytical strategy needed clarification and made substantial additions to that effect on p. 11 in the manuscript. We explain that while our analysis is based on between-subject comparisons, we incorporate elements of a "quasi-pretest-posttest design" (Clifford et al. 2021) by including a set of pretreatment questions that are highly predictive of our outcome measures (albeit not exactly the same questions, since that would jeopardize the experimental treatment itself as we now discuss on p. 11). More specifically, our pretreatment measures capture whether participants view immigration as an

important problem facing the country and whether they hold explicit racial stereotypes against Hispanic-Americans. The full regression results in Appendix C confirm that these pretreatment attitudes are highly predictive of post-treatment attitudes towards immigration and therefore improve balance across treatment conditions and substantially increase statistical power.

Regarding the question of balance across experimental conditions, we agree that our presentation of the sample composition in Appendix A.I was confusing and may have created the impression that our treatment groups are not balanced. Thus, we replaced the table with a graphical comparison of pretreatment covariate distributions across experimental conditions, which allows for an easier visual examination to ensure that our randomization was indeed successful. Of course, there can always be slight variations in pretreatment covariates due to chance, so we additionally supplement this graphical presentation with a multinomial logistic regression predicting assignment to each of the treatment groups to provide a statistical test whether our randomization ensured balanced experimental conditions (the control group serves as the reference category). This analysis reveals gender as the only (barely) significant predictor of assignment to the "forced exposure" condition (the difference between the proportion of men in the "forced exposure" and "free choice" condition, on the other hand, is not statistically significant). Notwithstanding, the analyses reported in the manuscript control for sociodemographics as well as other pretreatment covariates that are predictive of our outcome measures, thereby controlling for any remaining imbalance on these observed confounders due to chance during randomization.

2. The authors also need to discuss the flaws inherent with MTurk data and the use of VPNs outside the U.S. that may significantly bias the quality of the data. The authors need to present a more balanced view of the data and address the concerns that many researchers have with such data.

This is a very important point. We elaborated on the data quality of MTurk samples as well as our measures to block VPNs in the beginning of our results section (p. 12).

3. The authors also need to discuss why they decided to implement OLS when the outcome variable is dichotomous, as indicated on page 11.

We added a discussion of our modeling choices in the results section and explained why we implemented a linear probability model instead of a logit/probit when the outcome variable is dichotomous (now on p. 13). In addition, we replicated all analyses with dichotomous outcome measures in the paper using logistic

regression (see Appendix B.III). The substantive results are consistent across model specifications.

4. My last concern has to do with how credible it is for Fox News to publish such a positive article on immigration. Are there any potential issues with such a treatment that may not be believable and/or trustworthy?

This is another important point and we agree with the reviewer that the credibility of attributing the article to Fox News is a potential concern. Our goal was to frame the article as business reporting that could conceivably be published by either network since it emphasized a pro-business perspective highlighting the economic benefits of immigration through entrepreneurship. While we didn't ask respondents directly whether they consider the article to be authentic (in order to avoid related priming effects), we asked them to evaluate it on various dimensions (whether they viewed the article as balanced, accurate, fair, good, friendly, and American; see the relevant part of the questionnaire in Appendix D.II).

We now added a summary of these evaluations conditional on the source and pretreatment media preference in Appendix A.IV. The response patterns suggest that they viewed the source attribution as believable since the article was evaluated more favorably if it was published their preferred news network. In other words, the fact that media preferences (and presumably differences in perceived source credibility etc.) result in diverging evaluations of the same content suggests that the source attribution itself was believable. We added a brief discussion of this point in footnote 1 (p. 9). This conclusion is further corroborated by the comparison of Average Choice-Specific Treatment Effects discussed in Appendix B.I.