Advances.in
Reinventing Academic Publishing

Peer-review report of

Saleh, N., Makki, F., Van der Linden, S., & Roozenbeek, J. (2023). Inoculating against extremist persuasion techniques – Results from a randomised controlled trial in post-conflict areas in Iraq. *advances.in/psychology, 1*(1), 1-21. https://doi.org/10.56296/aip00005

Round 1

Dear Authors,

Thank you for submitting your interesting research to *advances.in/psychology.* I was fortunate to receive the evaluations of two experts in experimental social psychology, intergroup relations, and extremism. I thank them for their constructive and detailed reviews. Before taking their evaluations into account, I evaluated your manuscript myself. We all identify several key strengths of the present work. First, it focuses on a pressing social issue. Second, field experiments are scarce in psychology. Even rarer is their combination with hard-to-reach populations in non-WEIRD settings. Hence, I concur with the reviewers that these aspects make the current research unique. At the same time, both reviewers identify several issues that I invite you to address in a revision.

For instance, Reviewer 1 asks you to define the boundary conditions and unique characteristics of the inoculation method and theories more clearly. They also aimed to re-run the analyses but could not find the data in the OSF repository. Indeed, I was also unable to find the supplementary materials there. Please make sure the SOM is complete in the revision. Reviewer 2 asks you to discuss the overall theoretical approach and the specifics of the methodological approach more in light of the unique context of the investigation. They also ask for more statistical detail and a more detailed discussion of the comparability between the Salah et al. study and the current research. I concur with the idea of (qualitatively) comparing the means between the studies to make sense of the diverging results. Finally, both reviewers ask you to address more thoroughly the sample size limitations of the present research (e.g., in the methods and in the discussion).

Both reviewers highlight additional points, and I ask you to address each of them. In addition, please explain in more detail how the organization you collaborated with identified their participants (i.e., how did they determine whether someone is vulnerable to extremist recruitment?). This information is essential as it directly concerns parts of the discussion of how the context may have moderated the effectiveness of the manipulation.

I would also like you to (a) move the description of the materials from the end of the introduction to the methods section and (b) move the hypotheses from the end of the methods to the end of the introduction.

Table 1 is great, but the first column is too small for some of the text. Please adjust it accordingly. For Figure 4, please explain in the note what the different components of the boxplot and error bars (e.g., median, range, 95% CIs etc.) relate to, as this can vary from paper to paper. Please also describe the error bars for Figure 5.

On p. 13, you report a p value as “p < 0.897”. Please report it as “p = 0.897”

Finally, I was a bit unsure about the last paragraph of the discussion right before the conclusion. There is considerable research showing that factors such as threat, relative deprivation, perceived injustice, or identity fusion fuel extremism in non-WEIRD countries. At the same time, most work has indeed focused on the West. Nevertheless, some nuance would be helpful here.

In summary, we all enjoyed reading the current manuscript and see its clear potential that a revision may help you to realize fully. Therefore, I look forward to receiving a revised manuscript of the present manuscript, preferably within 30 days.

Best,

Jonas R. Kunst

Editor-in-Chief

Reviewer 1:

The goal of this paper is to use inoculation in the context of resistance to persuasion attempts to radicalization. The sample and the contexts are unique, which makes this study especially interesting. The paper is well-written and its clear that this is a group who has worked on this topic for a while. The paper also had a few major issues, here are the main ones:

1. Reading this paper raises an interesting question meta-question, it is possible to challenge the notion of inoculation and its scope with the existence of this growing literature? The paper is written with the assumption that inoculation the way its defined is a clear and well established phenomenon. But I have to admit that I am still not 100% convinced that the process and intervention are well defined enough to

be considered as one thing. This is especially important when its unclear what is the psychological mechanism that drives inoculation. A few questions that came to mind when reading the introduction and that could be addressed by the authors:

- What kind of preexposure is not considered inoculation?
- Does it work in every context?
- What is the mechanism the makes inoculation helpful?

I would love to hear some more discussion on these issues in the actual paper.

2. On the same note, I realize that papers have already been published using the game, but I also found a bit of disconnect between the theory of inoculation (an exposure to a minor dose) and the game. This did not feel at all to me as a minor dose as was discussed in the introduction. Participants are not only exposed to materials, but are also playing an active role in learning about the methods and using them. This feels a bit broader than the way inoculation was defined in the intro. I would love to further understand how the authors think of this game in relation to the inoculation idea. On the same point, it seems that participants are conducting training on how to identify manipulation tactics. It therefore makes sense that they become better at identifying these tactics. I was under the impression that the idea of inoculation is something else than mere active learning, which seems to be happening in the game.

3. The second main problem of the paper is the obvious lack of power. The study is clearly underpowered and 2 of the 3 hypotheses were not supported. Its unclear whether this is because of some unique aspect of the context (especially true for H2), or merely because of lack of power. Can we really make any conclusions, either positive or negative, from these underpowered studies? I realize that getting this sample was extremely challenging, but I am just not sure what we can take from these results.

4. I could not find the analysis of this data. The specified OSF repository that was suggested only had a script for plots, but not the main analysis. Therefore I could not replicate the results or test their robustness. The Qualtrics folder was empty and I wasn't able to see the survey. Also, it would be great to know more about the methods used in the study. For example: what were the selection criteria for the studies and how many people were removed?

Minor comments:

1. It would be helpful to divide the method section to participants, procedure and measures as what is manipulated and what is measured can be sometimes confusing.

Reviewer 2:

The work described in this manuscript, Inoculating Against Extremist Techniques – Results from a Rollout in Post-Conflict Areas in Iraq – is potentially valuable scientifically and practically. The research tests a psychological inoculation intervention, which is administered in the form of a digital game, that has the objective of increasing participants' resistance to extremist manipulation techniques. The intervention is well-grounded in psychological research on persuasion, particularly with regard to extremism. Practically, the work is timely because, as the authors explain, terrorist groups continue to be active in Iraq (5 years after victory against IS was declared). And, of course, terrorism in many different forms is a major issue of concern internationally.

While the grounding of the rationale in psychological theory and previous empirical work and the use of pre-post, treatment-control design are strengths of the current work, there are also some weaknesses. One limitation of the research in its formulation and methodology is that it essentially represents the application of a previous study by Saleh et al. (2021) with a different subject population; it is described as a "conceptual replication." Thus, particularly as the research is currently framed, it does not directly advance psychological theory. Of course, the application is in a relevant context with a hard-to-reach population. Also, given recent issues of replicability and reproducibility of findings in psychology, this approach certainly has other scientific merits. But, as a describe below, I think that if it included some additional measures, it could have been designed to advance theory more. A second weakness of the current work is the inconsistent results and, particularly, the lack of definitive explanations reconciling inconsistent findings. (Overall, though, the results are not too bad. One of the three main measures, perceptions of manipulativeness, which is probably the most important one, is significant, and the confidence measure had a p-value of .051.) Still, despite these weaknesses, a revised version of the manuscript could make this an appropriate candidate for publication in the journal. I believe that the work is scientifically rigorous, the context is unusual and relevant, and it uses a rare, hard-to-reach sample, and the results are stimulating. I offer some suggestions for re-thinking, reframing, and minor re-analysis.

Currently, the manuscript begins with a brief, informative description of the context in which the present research was conducted – Iraq, "one of the world's foremost countries where terrorism poses a significant problem." The introductory paragraphs also explain the collaboration with the Spirit of Soccer to test the researchers' gamified inoculation intervention. Following that, in the presentation of the

psychological grounding of that intervention, the rest of the introduction essentially parallels the introduction of the recent Saleh et al. (2021) study. I do not see that as a problem per se – it would be hard not to do that given the same reliance on the *Radicalise/MindFort* game using essentially the same procedure and measures as Saleh et al. (2021).

My reservation is that the current work is described mainly in terms of the culturally-sensitive refinements that were made to the game and other materials to make them appropriate for participants in the Mosul and Duhok regions of Iraq. What I believe would make this work more effective is a fuller description of what the present work is and what it is not. The authors do state that the "purpose of this study was to assess the impact of the same game in real life settings as opposed to online environments." However, I am a bit confused by that statement because participants in the current research also participated online, on "tablets and mobile phones." My suggestion is to emphasize more clearly that this is not simply meant as a replication of the Saleh et al. (2021) study but it might also be considered more as a test of the robustness of the intervention under the particular conditions in Iraq. (I suspect that is what the authors were implying when they contrasted "real life" with "online" environments, but I think Iraq represents a unique environment that is particular form of real life.) Then, I recommend that the authors link some of the characteristics of the context in Iraq directly to psychological theory and research that might limit (or promote) the impact of the intervention in Iraq relative to the context in Saleh et al. (2021). For instance, would the recent experiences there sensitize participants to extremist persuasive attempts or, because it relates to more common discourse, make such attempts less obvious? I do not feel that the authors need to have a firm position about this; a discussion that is largely exploratory is fine. My main point is that it would enhance the value of the work to more explicitly consider what it might be about the context in Iraq that might make the current work a test of the robustness of the effectiveness of the intervention. The nature and structure of the issues considered here could foreshadow some of the explanations that appear later, in the Discussion section. Some insights into these potential dynamics might have been gained from additional measures added after the main dependent variables were assessed in the actual study (which, because they are measured later, would not have affected the main responses of interest); this is something that might be cited as a limitation with some guidance offered about what kinds of measures reflecting the potential dynamics could be included in future work testing in the intervention in various contexts. These changes, which I do not feel misrepresents the work, will require some reframing of the research.

I also would like to see some more analysis-related information added to main the text to help readers with the interpretation of what was not, as well as a what was, found. Some of these data may be available in Supplementary Materials, but having them in the text would help readers (like me) get some answers as questions arise about the findings and interpretations, without the effort and distraction of having to

seek answers elsewhere. One important piece of information would be the statistical power of the tests for the present study compared to that for Saleh et al. The participants in the present research are a hard-to-reach population, so falling short of the target of 291 participants is not a “fatal flaw” of the current research. Also, the authors allude to limited power in the present research and less power than Saleh et al. in their explanations. Thus, to better quantify these points, I recommend that the authors report a sensitivity power analysis for the present sample not where it is in a footnote but rather more prominently when the sample is described in the main text. If possible, a power analysis for the Saleh et al. research might be noted there, as well. This explicit information would be helpful as the authors interpret the results of the present research in comparison to those obtained by Saleh et al. (2021). Another bit of information that I would find useful in the main text is a correlation matrix for the primary variables of interest. In addition, I suggest a brief mention (but not necessarily a full table) about whether any of the demographic variables (1) differ by condition (which presumably is not the case if randomization is successful) and (2) are related to any of the outcome variables of interest.

Because of the many things that vary between the present study and Saleh et al. (2021), it is not possible to come up with a fully persuasive explanation for the different findings between the two studies. There were substantial differences recruitment strategies (Prolific vs. a more target recruitment in the present research), in the culture and experiences of participants, and in demographic distributions (e.g., 57% of the sample identifying as female in Saleh et al. compared to 30% in the present research). As the authors note in their Discussion, even the seemingly minor and culturally sensitive differences in the wording of materials could have significant impact on responses. Also, there is the issue of measurement invariance when interpreting results across cultures. Or, it could be that naming the game Radicalise in the UK sensitizes participants to manipulativeness or creates a sensitivity to manipulativeness or produces demand characteristics that drive the effects observed in Salah et al. Still, I believe that the Discussion would be more effective if the issues were foreshadowed in the Introduction and organized in a more structured way. Currently, the interpretations have a list-like quality rather than appearing like a systematic critique.

I strongly believe that the Discussion would be stronger if it was more forward-looking, identifying how future investigations could be conducted in ways that could test theoretically-relevant factors that could provide insight in the present work, the Saleh et al. study, and the comparison of results between the studies. One question that might be considered more fully involves how the different participants and contexts of the two studies might affect (and potentially moderate) the impact of the intervention. This consideration might drawn on some clues in the current data, and then conclude by suggesting concrete ways these ideas could be tested in future research.

Let me offer one example, building on an observation that the authors currently make. The authors note, “the complexity of a post-conflict region makes it more difficult for one to pick up on traits that could be seen ‘out of the ordinary’ in normal circumstances.” I think there are data in the present study that are at least consistent with that point. While any comparisons across study must be made cautiously, it could be useful to call attention to the mean differences across studies (but not with any formal significance tests). With respect to the perceived manipulativeness (related to H1), the mean values for treatment vs. control condition (4.07 vs. 3.78) were much lower in the present study than in Saleh et al. (6.22 vs. 5.64). Regarding the perceived vulnerability measure (see H2), the inoculation and control conditions in the current study had basically the same means (4.00), which were similar to the control condition (4.28) and lower than the inoculation condition (5.11) in Saleh et al. For the confidence measure (related to H3), the means for both the inoculation condition (5.24) and the control condition (4.93) were again lower than the comparable conditions in Saleh et al. (6.12 and 5.83). While a number of the differences between studies noted by me and/or the authors might help explain these mean differences between the studies, these mean differences between the studies could be considered as suggesting that the generally lower level of perceived manipulativeness in the present research than in Saleh et al. could be because such manipulative attempts may be more a part of normalized discourse based on people’s experiences in this area of Iraq, which reduces responsiveness to these efforts and blunts perceptions of manipulative attempts. Future research relating to this explanation might be suggested in the current region by including an independent variable that does or does not prime perceptions of manipulativeness; for a more general sample, future research might vary the clarity/ambiguity of manipulation attempts in vignettes to test if that moderates the impact of the intervention. I think such considerations and specific guidance would enrich the Discussion.

In conclusion, I found this manuscript to be very stimulating. The goals of the work are important, the design is appropriate, the procedure was thoughtfully adapted, the analyses skillfully conducted, and the results are interesting. The context of the research is very relevant. However, I do feel that some revision – mainly some re-thinking, re-framing, and re-writing – is needed before I can make a firm recommendation. Mainly I am recommending that the authors prepare readers more about why the intervention might be more or less effective than in the non-conflict context of Saleh et al. in the Introduction. Then, the Discussion might return to these theoretically-relevant issues and draw on elements of the current findings to suggest specific directions, designs, and measures for future research. All studies have loose ends. In my view, studies are more valuable if these loose ends can be used to suggest concrete, productive avenues for future work. I strongly encourage the authors to make these revisions and continue to pursue publication in this journal.

Round 2

Dear Authors,

I appreciate your hard work in refining your manuscript and responding to the reviewers' feedback. There is a clear consensus among us that the manuscript has substantially improved, thanks to your more nuanced theorizing, self-reflective approach, and conscientious testing of robustness.

While one reviewer is ready to accept the paper, the other suggests a few minor revisions which I agree would further improve the manuscript. One key area identified is the interpretation of the effects, specifically those with p-values just above the 0.05 cutoff. This has been a subject of ongoing debate among researchers - whether to refer to these as 'marginally significant' or to avoid an arbitrary cut-off entirely. I recommend in this instance you could consider substituting phrases such as 'marginally significant' with something akin to 'falling just above the traditional 0.05 significance level' to avoid any confusion.

I would also appreciate it if you could address a few additional minor issues:

1. Table 2's title and caption seem misplaced. Moreover, in line with the suggestions from the previous round, could you display the mean values from both studies in it to illustrate the different baselines resulting from the study contexts? You have mentioned these in the discussion, but perhaps including them in the table and referencing them in the text could provide greater clarity.
2. Please ensure all quotes include page numbers, for instance, p.17, quote by Compton (2021).
3. You have conducted a sensitivity analysis, but the effect size you have selected (0.15) is considerably smaller than that used by Saleh et al. (.26). Could you clarify why you have not used the same f an in Saleh et al., given that you have followed their pre-registration? What is the power if you set the effect size to .26?
4. Please ensure statistical abbreviations, other than Greek letters, are italicized (e.g., p. 26 and within tables). Additionally, please always present p-values with three decimal places in both the text and tables.
5. Could you provide a direct link to the Supplementary Information (SI)? I assume it is identical to the link to the OSF repository?
6. There is a small typo on p. 28: “After being presented with each of the WhatsApp conversations, participants were then asked to the following questions”.
7. In the results section, please ensure you consistently use past tense.
8. Lastly, please include a conflict-of-interest statement, indicating whether any exist.

In sum, I am very pleased with the revisions you have made thus far, and I believe that, after these minor changes, the manuscript will be ready for publication. The small sample size naturally limits the conclusions that can be drawn from this study.

However, it is of significant importance to conduct field studies like this, especially in non-WEIRD settings. I am confident that your manuscript is very close to striking an effective balance between these considerations.

I look forward to receiving your revised manuscript.

Best,

Jonas R. Kunst

Editor-in-Chief

Reviewer 1:

Generally speaking, most of my comments have been addressed. I have a few minor issues primarily related to the interpretation and summary of findings in the discussion:

1. Figure 7 is difficult to read. I recommend placing both conditions next to each other (rather than in different facets), and separating the confidence and manipulation aspects into facets. Hence, there would be a manipulation facet and a confidence facet, and within each one, we will have pre and post for both conditions differentiated by color. Additionally, adding a grid in the background would make the comparison of these graphs easier.

2. I feel that the description of the results in the discussion still leans towards optimism, and the language is somewhat vague. This could potentially lead the reader to believe that the results are more significant than they truly are. For instance, I'm unsure if the general discussion should include the statement: “We report encouraging results.” To me, most of the hypotheses were not supported, so I am not entirely convinced that these results are encouraging. I do not believe they are discouraging, I simply want to highlight that a casual reader of this paragraph might misconstrue the message. Furthermore, I don’t regard p = 0.051 as "not quite significant," but rather as not significant. I also believe that the authors should exercise caution when claiming that these results align with previous findings, and would ask the authors to drop this sentence. Currently, perhaps due to the lack of power, we cannot definitively say whether they align or not. I would ask the authors to remove this sentence. It would also be helpful if the authors could report the results of H3 in the first paragraph, indicating that it was also non-significant. I think that presenting to the reader a realistic picture of what was found in the paper is important.

Reviewer 2:

I have reviewed the latest version of the manuscript, both in terms of how it addresses the specific points raised in the last round of comments and a more

holistic assessment of the work as it stands. I believe that the authors have carefully considered and responded to the issues raised about the previous version of the manuscript. Also, overall, the authors do a nice job in the current version of the manuscript of describing the rationale, context, method, and findings more fully, and they include informative additional analyses. Moreover, the authors alert readers to some shortcomings of the research (e.g., sample size and interpretation of null findings). I found their discussion of the “loose ends” to be thoughtful, plausible, and potentially generative. I believe that the work is much stronger than the previous version that the work makes a valuable contribution to the literature and should attract the attention of other scholars, practitioners, and the general public. I am pleased to support it for publication.