

The effect of praise from peers on empathy and political inclusion towards racial or ethnic outgroups

Corresponding Author: Dr Adeline Lo

Version 0:

Decision Letter:

30th April 2024

Dear Dr Lo,

Thank you once again for your manuscript, entitled "Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?," and for your patience during the peer review process.

Your manuscript has now been evaluated by 3 reviewers, whose comments are included at the end of this letter. Although the reviewers find your protocol to be of interest, they also raise some important concerns. We are interested in the possibility of proceeding further with your submission in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a revised manuscript before we make a decision on in principle acceptance and Stage 2 submission.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

1. Reviewer 1 finds it crucial for the research question that longitudinal evidence is provided. Please follow the reviewer's suggestion and add a follow-up to your design.
2. Our reviewers have various methodological concerns and suggestions. For example, Reviewer 3 asks that the control condition is revised to be comparable to the treatment condition and Reviewers 2 and 3 suggest revising the donation task and its instructions. We ask that you carefully consider all of these concerns and address them in full in your revised manuscript.
3. Our reviewers find the writing to not be clear and the manuscript to lack important methodological details. Please revise the writing to be clear and concise and provide all necessary methodological details. In your revised manuscript, please also explicitly motivate the choice of mediators and explain how confounds will be dealt with.
4. Reviewer 3 suggests that you also report Bayesian statistics. We agree with the reviewer and ask that you commit to a method (such as Bayes Factors) for interpreting the strength of null results.

In sum, we invite you to revise your Stage 1 Registered Report taking into account reviewer and editor comments. Please highlight all changes in the manuscript text file.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a "Response to reviewers" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* Ensure that you use our template for Stage 1 Registered Reports to prepare your revised manuscript:

https://www.nature.com/documents/NHB_Template_RR_Stage1.docx. Failure to ensure that your revised Stage 1 submission meets our requirements as specified in the template will result in your submission being returned to you, which will delay its consideration.

* In your cover letter, please include the following information:

--An anticipated timeline for completing the study if your Stage 1 submission is accepted in principle.

--A statement confirming that you agree to share your raw data, any digital study materials, computer code (if relevant), and laboratory log for all eventually published results.

--A statement confirming that, following Stage 1 in principle acceptance, you agree to register your approved protocol on the Open Science Framework (<https://osf.io/>) or other recognised repository, either publicly or under private embargo, until submission of the Stage 2 manuscript.

--A statement confirming that if you later withdraw your paper, you agree to the Journal publishing a short summary of the pre-registered study under a section Withdrawn Registrations.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Human Behaviour or published elsewhere.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Human Behaviour is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit <http://www.springernature.com/orcid>.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

[Redacted Signature]

[Redacted Name]

Nature Human Behaviour

Reviewer expertise:

Reviewer #1: interventions to increase intergroup empathy

Reviewer #2: identity

Reviewer #3: Registered Reports

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

The authors of this paper plan to test the effectiveness of "peer praise" intervention that is designed to promote empathy in reducing outgroup bias. While I see potential in the approach of the authors (testing a simple, probably scalable to reduce outgroup bias), I have several concerns that qualify my general enthusiasm for this paper.

- While the pilot studies already provided some evidence for the immediate effect of the intervention on empathy and outgroup bias, evidence for long-term effects is missing, but in my view critical to back up the claim of the authors that this intervention is a simple and promising one. So in my view, the authors should add a follow-up (e.g., assessing empathy and indicators of outgroup bias some time after the intervention) in their design. This also helps to reduce concerns that the immediate effects might be mainly a result of priming, inducing positive mood (one of the proposed mediators!), and/or demand characteristics.

- I also missed more arguments that back up the claim of the authors that this kind of intervention is more easy to scale up compared to alternative interventions. For instance, it is not clear to me in which way the authors plan to implement this intervention in different contexts (i.e., instructions for those who implement the intervention; instructions for participants), how they make sure to reach individuals that mainly profit from this intervention (i.e., highly prejudiced individuals), whether a single “treatment” is sufficient, etc.

- I also missed a more critical review of the literature on empathy interventions, especially those who mainly target general empathy, since there is mixed evidence that inducing empathy helps to overcome outgroup bias (see Behler, & Berry, 20022). I also wondered why the authors target general empathy, not directly empathic concern toward outgroups.

- I’m also a bit sceptical about the proposed mediators. Positive emotions might be partly responsible for the documented effects of the intervention in the pilot studies. But this means that long-term effects of the intervention are less likely. Because of the design of the intervention (manipulation), some sort of norm information is given to the participants. However, there are already simple and easily scalable interventions tested that directly target social norms (Murrar, Campbell, & Brauer, 2020).

- The authors provide many details about their planned study, but important information is missing that would enable other researchers to replicate the experimental procedure (e.g, instructions given to the authors). Although they plan to make all data and materials openly available, I think it is important to include all study materials (instructions, items, etc.) in this preregistered report.

Reviewer #2:

Remarks to the Author:

“Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?” is an interesting manuscript that describes seven pilot studies testing components of a “peer praise” intervention, and proposes well-powered tests of that intervention in a registered report proposal. Overall, the manuscript is well-written, the proposed intervention has been well-conceptualized and vetted in pilot tests, and the proposed registered report studies are both feasible and reasonable tests of the central research question. I just have a few methodological suggestions for the authors to consider before conducting these studies, and then a more minor theoretical point to consider incorporating either in the introduction or eventual discussion section.

My biggest methodological suggestion (if financially feasible) is for the authors to take a stratified sampling approach in their proposed studies—specifically to recruit equally sized samples of Democrats and Republicans so that they can test for intervention moderation by partisanship. The reason for this recommendation is due to the results the authors found in Pilot 7. The authors ruled out social norms as the mechanism through which their intervention might operate, but in my view, their Pilot 7 results suggest *political norms* might in fact matter greatly for whether the praise intervention would yield empathy and political inclusion. Specifically, on page 11, the authors noted that Democrat co-partisan praise had a positive effect on empathizing and choosing to learn about BLM, while Republican co-partisan praise had a non-significant or negative effects. That suggests to me that for a peer praise intervention to be effective, peer praise must be aligned with what is normative for the group. In contemporary American politics, Democrats are the party of racial inclusion whereas Republicans have abandoned inclusion as a priority (Levitsky & Ziblatt, 2019, 2023). It makes sense, then, that when Democrats are praised for inclusion efforts, that has a positive effect on their behavior, whereas the same intervention has a different effect on Republicans. That kind of heterogeneity is important to consider (Bryan, Tipton, & Yeager, 2021), and I think the authors’ own data suggests it should be factored into the research design—perhaps as a 2 (Intervention, Control) X 2 (Democrat, Republican) factorial—again, if that is financially feasible (they would need to redo their power calculations to figure out how such a design change would affect sample size).

A second methodological suggestion is to consider how the use of racialized faces from the Chicago Face Database might ultimately affect inferences. Recent research from face perception researchers suggests that this practice can distort our inferences, especially when trying to make claims about race and racism (see Martinez, 2023 for a more thorough description of the issue at hand).

A third methodological suggestion is to consider a broader set of donation options for the donation DV. I know that the authors already pilot tested BLM and UnidosUS, but I am concerned about how recent shifts in attitudes about BLM might affect responses. Specifically, BLM has become such a charged movement, especially in the backlash against the 2020 racial justice movement, that it is even less popular now than it used to be (see Chudy & Jefferson, 2021). Therefore, I wonder whether lack of willingness to donate to BLM means the same thing as it did when the authors ran their pilots.

My (more minor) theoretical point is related to the discussion of empathy and perspective taking interventions in the introduction. Recent research suggests perspective-taking is not the most effective pathway (in part because perspective taking amplifies egocentric biases); instead perspective-*getting* is more fruitful (see Eyal, Steffel, & Epley, 2018). This distinction does not change the design of the proposed study, but I do think it is worth discussing either in the introduction or discussion.

References

Bryan, C. J. Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 5(8), 980-989

Chudy, J., & Jefferson, H. (2021, May 22). Support for Black Lives Matter Surged Last Year. Did it Last? New York Times.
Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, 114(4), 547-571.
Levitsky, S., & Ziblatt, D. (2019). *How Democracies Die*. Crown
Levitsky, S., & Ziblatt, D. (2023). *Tyranny of the Minority*. Crown
Martinez, J. E. (2023). Facecraft: Race reification in psychological research with faces. *Perspectives on Psychological Science*.

Reviewer #3:

Remarks to the Author:

Thank you for submitting this Registered Report Stage 1 manuscript. I read the work with great interest.

I believe that the general topic area of the study and that the research question are valuable. The overall methodological approach is coherent, although there are inherent limitations of the design, which will naturally constrain the knowledge gained from the study. These are that it is still an online study with mostly self-report outcomes, and those outcomes that are behavioural are relatively low-effort/lost-cost and may not reflect typical behavioural situations for the participants. It would be out of the scope of the study to test whether the intervention has a longer-lasting impact, although such a goal (and the scientific work that would test this) would be more valuable and informative for the research question (and this may be something that the researchers could consider for future work, i.e., even testing a gap of just a few minutes between the intervention and the outcome measures would be a good start).

That being said, my general assessment is that this work would make an important contribution to the literature and that it would be valuable for us to have this work done and disseminated, than not done at all. Nonetheless, I have some concerns and questions about the manuscript and the study design that I would like to see addressed. I have separated these into "major" and "minor" categories.

Major concerns and questions:

1. In general, I found much of the manuscript difficult to read. This is not a comment about the writing style per se, but more about the density of the information and how that information has been structured. I have a nagging feeling that some of my questions/concerns below may already be addressed in the manuscript, but I have missed these because of my difficulty in gaining a confident understanding of some portions. In the next few comments, I have outlined a few places where a clear understanding seemed particularly difficult, but I would advise the authors to cast a critical eye over the whole manuscript and improve the clarity wherever possible. Indeed, reading the manuscript felt too effortful, and I believe this will lead to disengagement by future readers.

2. Following on from (1): When the Introduction moves to the description of peer praise interventions and previous work, I found it very difficult to follow (page 4 and some of page 5). This section is very dense and I found it hard to keep track of what is being proposed in the current work, what had come through pilot work, and what had come from previous publications. I think the clarity of this portion of the Introduction needs to be improved so that it is easier to grasp the key points with just one read-through. For example, one change could be that the Introduction needs to be broken down into sub-sections with informative sub-headings.

3. The Introduction seems to jump around a lot. For example, near the top of page 5, the first full paragraph is general information about empathy, but such information had been introduced on page 3. It's not clear why this type of information is then returned to on page 5. After this, there is general literature review information about empathy-based interventions and, again, this feels misplaced. The authors should think about restructuring the Introduction so that all the background information from previously published work comes first, before moving on to information about the pilot work and their plans for the current work. Page 7 then describes the planned work in more detail, but much of this overlaps with information on page 4, which adds to the difficulty of following the Introduction.

4. I would prefer to see Hypotheses 7-12 explicitly listed in the manuscript. Since they are very similar to H1-6, it might help to integrate them, particularly if the Introduction is re-written more generally as suggested. For example, they could be written separately but linked as H1a and H1b etc., or they could be written in a more combined way like "H1: Peer praise (Study 1: Neutral Peer Praise; Study 2: Co-partisan Peer Praise) increases the likelihood..."

5. Like the Introduction, the section describing the Pilot Studies is very dense (and repetitious with some aspects of the Introduction). There is detail about the pilot studies already in the supplemental information, but can much of the narrative about the pilot work be moved from the manuscript into supplemental material as well? The manuscript could then contain a more concise section summarising the key take-aways from the pilot studies.

6. In Figures 2 and 3, many of the words are cut off – is this how they actually will appear in the study? It will be effortful for the participants to read the word clouds because of this (which will already be difficult for some neurodiverse participants). Where participants persist with the word clouds, you will get different outcomes where some will put in the effort to guess the words (and there may be different options for some, like "excel" -- is this "excel" or "excellent"?) and some will not bother. Thus, participants will have different experiences of the word clouds because of this effortfulness. It would be best if all the words and phrases on the clouds are clearly displayed in full.

7. My understanding of the experimental design is that those in the praise conditions will see a word cloud and that those in the control condition will see nothing at all. Is this correct? Assuming it is, from an experimental control perspective, it would be better for the control participants to have something to look at, for example, a word cloud of neutrally rated, concrete words. Technically speaking, any differences could be attributed to other features of the manipulation and not necessarily to the praise words themselves (e.g., maybe the bright, happy colours of the word cloud lifted the participants' mood).

8. The self-reported empathy item is double-barrelled – participants are asked about “how they are treated by police and fellow citizens” and “the hardships they face due to discrimination by race and ethnicity” – these are two different things and participants may be inclined to give different responses (one could even argue that “treatment by police,” and “treatment by fellow citizens,” are also separate things). Can participants be asked to respond to these as separate items? (even if the responses are then averaged for the purposes of the analysis). I believe that doing so would increase validity.

9. I wonder if the White House letter-writing task would be affected by political identity. I understand that the analysis will check for balance between groups in terms of covariates but will identity be entered into the analysis as a statistical covariate at all? Similarly, this kind of task may also be impacted by education level. Will this be controlled for statistically?

10. In the donation task, the text of the item simply says “a portion.” This is really unclear and how this is interpreted may massively impact the responses. This is because such lexical interpretation differences cannot be measured as a covariate and this could differ between groups. Can “a portion” be specified? Will it be specified? Or can participants choose how much of a portion they want to donate? (The final suggestion may even be more informative than just yes/no.)

11. The measures are described but oftentimes this seems to leave out the response scales (Likert scales). Can these be clearly given or described in the manuscript for all measures?

12. I would like to see what information is given to participants before they begin, perhaps as part of the supplemental information. The manuscript alludes to a lack of deception, but it is difficult for me to then understand what participants are told about the purpose of the study and the manipulation. This may impact how they respond so I would like to see this.

13. The power analysis indicates 400 participants. If my understanding is correct, this means 200 per group (i.e., not 400 per group). It would be helpful to know more about any inclusion/exclusion criteria in relation to this sample size. For example, my understanding is that anyone can participate but then the analysis will focus on white participants. Thus, does the “400” refer to white participants who passed the attention checks etc. (i.e., after other datasets are excluded), OR does the “400” refer to everyone who starts the study and is randomised, in which case the analysis will end up with less after excluding datasets from people of colour, from those who failed the attention checks, etc.? (i.e., will you replace excluded datasets to ensure that there are 400 useable datasets? And are you ensuring that there are 200 useable datasets per group, or is it possible that the praise vs control group sizes might differ?)

14. What platform is the survey held on? Is it directly on MTurk or on something else, like Qualtrics?

15. I was glad to see checks for the effects of attrition built in. I am assuming that participants will be able to skip items...? If so, will patterns of missing data be explored? (i.e., for skipping items rather than attrition overall)

16. Will participants be excluded only if they fail both attention checks, or will they be excluded if they fail just one?

17. Even if a particular null hypothesis can be rejected, why not use Bayes Factors for assessing the evidence in favour of the alternative hypothesis?

18. It is not clear to me why the two types of praise will not be compared against each other. This was briefly discussed in the context of one of the pilot studies, but it would seem valuable to test this here. Knowing that one works better than the other, or that they work as equally well as each other, seems worthwhile to analyse. The effect is likely to be smaller, of course, but it could be included even as an exploratory analysis as I think this is the obvious question that most readers will be asking.

19. While it's clear that the study is focusing on “Americans,” how is this being defined? Do participants have to be natural-born or naturalised citizens? What if they are living in the U.S. but are not citizens? What if they are temporary immigrants? Furthermore, there may be American citizens who have lived abroad for a long time. It may be helpful to gain information on these aspects for use in the analysis and in determining the final sample.

Minor comments:

1. In the first paragraph of the introduction, I would suggest having an example of the health consequences of outgroup bias that is instead of or in addition to Medicaid, since that is very US-specific, and the “US context” has not been introduced at this point. A more general example will better show that this is a widespread phenomenon.

2. There is a missing closing parenthesis in line 6, after “towards outgroup members”

3. In the second paragraph there are some missing words here: “referred to as “perspective taking” and)” – unless the “and” is meant to join the citation numbers; it's unclear to me.

4. In the Introduction, it would be better to avoid the word “you” in this sentence: “(14,704 observations once you account for multiple trials)” perhaps by changing it to something like “14,704 observations after accounting for multiple trials”
5. In most places, the authors refer to “black Americans” but there is the odd place where it is “Black Americans.” This should be double checked for consistency.
6. Near the bottom of page 7, I believe there is a typo: “latino” – elsewhere it is written as “Latino/a”
7. There is a lot of word-for-word repetition between the Analysis plan and Contingency plans sections, which seems unnecessary.

Version 1:

Decision Letter:

6th November 2024

Dear Dr Lo,

Thank you once again for your manuscript, entitled "Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?," and for your patience during the peer review process.

Your manuscript has now been evaluated by 3 reviewers, whose comments are included at the end of this letter. Although the reviewers find your protocol to be of interest, they also raise some important concerns. We are very interested in the possibility of proceeding further with your submission in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a revised manuscript before we make a decision on in principle acceptance and Stage 2 submission.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. Please see the attached marked-up version of your manuscript for all editorial requests.

In sum, we invite you to revise your Stage 1 Registered Report taking into account reviewer and editor comments. Please highlight all changes in the manuscript text file.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a “Response to reviewers” document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* Ensure that you use our template for Stage 1 Registered Reports to prepare your revised manuscript: https://www.nature.com/documents/NHB_Template_RR_Stage1.docx. Failure to ensure that your revised Stage 1 submission meets our requirements as specified in the template will result in your submission being returned to you, which will delay its consideration.

* In your cover letter, please include the following information:

- An anticipated timeline for completing the study if your Stage 1 submission is accepted in principle.
- A statement confirming that you agree to share your raw data, any digital study materials, computer code (if relevant), and laboratory log for all eventually published results.
- A statement confirming that, following Stage 1 in principle acceptance, you agree to register your approved protocol on the Open Science Framework (<https://osf.io/>) or other recognised repository, either publicly or under private embargo, until submission of the Stage 2 manuscript.
- A statement confirming that if you later withdraw your paper, you agree to the Journal publishing a short summary of the pre-registered study under a section Withdrawn Registrations.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Human Behaviour or published elsewhere.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Human Behaviour is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit <http://www.springernature.com/orcid>.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

[REDACTED]

[REDACTED]

Nature Human Behaviour

Reviewer expertise:

Reviewer #2: identity (original reviewer)

Reviewer #3: Registered Reports (original reviewer)

Reviewer #4: interventions to increase intergroup empathy (newly recruited reviewer)

Reviewers' Comments:

Reviewer #2 (Remarks to the Author):

I have now read the authors' response to suggestions and am satisfied with the changes they made to address my (and other reviewers') concerns.

Reviewer #3 (Remarks to the Author):

Thank you for the revised version of your manuscript, and for the detailed response letter.

Overall, my previous comments and questions have been addressed well. In particular, the writing and flow of the manuscript is vastly improved, and so the background and the study plan are much clearer.

There are just a couple of points that require clarification in the manuscript main text:

1. Previously, I queried if participants would be excluded on the basis of failing one or both attention checks. The letter says that participants will be excluded if they fail "one of the two attention checks" and refers to page 10 of the manuscript. However, page 10 does not seem to be completely clear about this as it says "Analyses will be conducted on all respondents who complete the survey and pass both attention checks. Participants will be selectively excluded from specific analyses if they... (b) fail one of the pre-treatment attention checks."

2. I appreciate the detailed response to my questions about the words appearing to be "cut off" in the word clouds, i.e., the clarification that these are word stems, as mentioned in the letter: "As a reminder, we used word stems to summarize our peer praise for empathy, and as is common in word clouds, used stems to group similar words. For example, "empathetic," "empathy," "empathize" and "empathizer," "empathizes" and "empathizes" would all be summarized by the stem, empath—".

However, in going back over the manuscript, I cannot find any reference to "word stems" -- even doing a search through the document for "stem" does not produce any hits for this (only for "systems"). I can accept the explanation that grouping similar words according to stems is "common" in word clouds, but that is not something I was aware of. It is certainly possible that participants may not be aware of this either; I was trying to imagine looking at each word cloud as a participant would, and I felt distracted by the oddness of reading the word stems and wondered if it had just been poorly formatted. But, I will accept the authors' argument that this is "uncontroversial" given their expertise and prior use of word clouds. However, for the

manuscript, given that NHB has a broad audience, some of whom may also not be aware that it is common for word clouds to use word stems, it would be helpful if the manuscript explained that the word cloud uses stems instead of (entirely) whole words.

I thank the authors again for the time and effort they have put in to taking my previous comments and questions seriously and for the well-considered ways they have addressed them.

Reviewer #4 (Remarks to the Author):

Review for NHB: Registered Report

NATHUMBEHAV-23113998A

Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?

Research on psychological intergroup interventions is gaining momentum and, in my view, constitutes one of the most relevant and engaging research areas within the social sciences. This is one reason why I find this work both valuable and well-aligned with current trends in the field. However, I did wonder why the authors chose not to situate their work within the broader context of psychological intergroup interventions (e.g., Halperin et al., 2023

<https://www.routledge.com/Psychological-Intergroup-Interventions-Evidence-based-Approaches-to-Improve-Intergroup-Relations/Halperin-Hameiri-Littman/p/book/9781032264257?srsId=AfmBOopvsUPyosAYs21Z-g6uVx5WIRiuTDI805g9IOzjNTiUIU5B3Q16>, focusing solely on empathy-related interventions. Given the research question and primary outcomes, I would recommend placing this work within the larger field of intergroup interventions to enhance its relevance and applicability.

Related to the above, recent review on intergroup interventions (Čehajić-Clancy & Halperin, 2024, Nature Reviews <https://www.nature.com/articles/s44159-024-00330-z>

has identified key steps and existing limitations for advancing research on intergroup interventions. Some of those limitations (such as addressing durability of intervention effectiveness and identifying mechanisms) are directly addressed in this work. This is good, timely and necessary. I encourage the authors to emphasize these contributions in light of the recent review on advancing research on psychological intergroup interventions.

Besides relevance and important contributions that this work might yield, there is also much to appreciate about the design. In particular, the longitudinal approach, combination of attitudinal and behavioural measures as well as inclusion of control/placebo group strengthen the design. Having said this, I do have several concerns and/or suggestions as to how this work could be improved.

1. The main argument in this work is that peer-praise might be a relevant psychological approach to changing/influencing intergroup outcomes (e.g., biases). I wondered why the authors have not provided a more comprehensive theoretical review on social learning and the mechanisms through which learning from others impact intergroup (e.g., racial) attitudes. Work by Olsson and colleagues on social learning and racial attitudes might be relevant here and it will strengthen and provide theoretical framework for authors' argumentation on how/why peer-praise might influence intergroup outcomes.

2. My main concern relates to the selection and proposed mechanisms. I think that the authors should motivate these choices more rigorously. First, they should provide theoretical reasoning for selecting positive emotions and social norms as the predicted mechanisms. Second, they should review existing research that directly supports these choices. For example, recent work on the role of positive emotions (more general) and moral elevation (more specifically) has identified the relevance of positive emotions for improving intergroup relations in contexts of contested and conflictual intergroup relations (see the recent JPSP issue <https://psycnet.apa.org/record/2025-15509-001?doi=1>

Related to this, I wondered why the authors did not consider social learning mechanisms as possible explanations for the intervention's effectiveness. This could, at least, be discussed.

3. Finally, a minor issue relates to the problem statement behind this work. It is argued that outgroup biases (racial/ethnic) in the US context are on the rise, particularly as expressed by the dominant group. I would strongly recommend supporting this claim with more concrete evidence to better justify the problem choice. Additionally, it would be helpful to clarify why the study focused exclusively on the dominant group and not on the other racial/ethnic groups in the US context.

Version 2:

Decision Letter:

27th January 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Lo,

Thank you once again for submitting your revised Stage 1 Registered Report, entitled "Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?". Everything is in order and I am delighted to say that we can offer acceptance in principle. You may progress to Stage 2 and complete the study as approved.

As you know, a condition of in-principle-acceptance is that the authors agree to deposit their Stage 1 accepted protocol in a repository, either publicly or under embargo until Stage 2 acceptance and publication. We are very keen to showcase our in-principle accepted protocols, so that our readers, reviewers, and potential authors can gain insight into the requirements of the format as well as an idea of the types of projects that are suitable for publication in Nature Human Behaviour. We have set up a space on figshare (https://springernature.figshare.com/registered-reports_NHB) to host all of our in-principle accepted protocols, which can either be made public or kept under embargo until Stage 2 acceptance (depending on author preference). This gives you the opportunity to have your work publicly associated with Nature Human Behaviour, and of course we will be very pleased to showcase your report if you agree to share it publicly.

Depositing the work on our figshare space does not preclude deposition of your Stage 1 protocol on other depositories – your protocol can also be posted on OSF, Dataverse, Dryad or any other public repository of your choice. You also do not need to do anything – if you agree with posting your protocol on our figshare space, we will upload your protocol on your behalf and either set it public or place it under embargo, depending on your choice. Your protocol will be licensed under a CC BY license (Creative Commons Attribution 4.0 International License). The CC BY license allows for maximum dissemination and re-use of open access materials and is preferred by many research funding bodies. Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor (read full legal code: <http://creativecommons.org/licenses/by/4.0/legalcode>) Please note that any use of <https://springernature.figshare.com> will be subject to the Figshare terms of use. Figshare has the right to enforce these terms and conditions where applicable. Use of third party services and sites will be subject to the relevant terms of use and will apply if we act on your behalf in this regard. Do let me know if you would like to take up this option or if you have any questions regarding the protocol deposition requirement.

Following completion of your study, we invite you to resubmit your paper for peer review as a Stage 2 Registered Report. Please note that your manuscript can still be rejected for publication at Stage 2 if the Editors consider any of the following to hold:

- The results were unable to test the authors' proposed hypotheses by failing to meet the approved outcome-neutral criteria
- The authors altered the Introduction, rationale, or hypotheses, as approved in the Stage 1 submission
- The authors failed to adhere closely to the registered experimental procedures
- Any post hoc (unregistered) analyses were either unjustified, insufficiently caveated, or overly dominant in shaping the authors' conclusions
- The authors' conclusions were not justified given the data obtained

We encourage you to read the complete guidelines for authors concerning Stage 2 submissions at <https://www.nature.com/nathumbehav/registeredreports>. Please especially note the requirements for protocol deposition, data sharing, and that withdrawing your manuscript will result in publication of a Retracted Registration.

In recognition of the time and expertise our reviewers provide to Nature Human Behaviour's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?". For those reviewers who give their assent, we will be publishing their names alongside the published article.

When you are ready, please use the following link to access your home page and submit your Stage 2 Registered Report:

Link Redacted

*This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

We expect your Stage 2 Registered Report to be submitted by the date specified in your latest cover letter. If unforeseen circumstances prevent submission by that date, please contact us as soon as possible to discuss any changes to the submission time-frame.

Thank you again for offering us this work and we look forward to receiving your Stage 2 Registered Report.

Yours sincerely,

[Redacted Signature]

[Redacted Name]

[Redacted Title]

Version 4:

Decision Letter:

10th December 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Lo,

Thank you once again for submitting your Stage 2 Registered Report, entitled "Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?," and for your patience during the re-review process.

Your manuscript has now been evaluated by the reviewers from the previous rounds of review, whose comments are included at the end of this letter. In the light of our reviewers' advice, we are pleased to inform you that we will be able to accept your Stage 2 manuscript, pending revisions to address reviewer comments and editorial requests.

One of the main reasons for delays in eventual acceptance is failure to fully comply with editorial policies and formatting requirements. To assist you with finalizing your manuscript for publication, I attach a checklist that lists all of our editorial policies and formatting requirements.

Please attend to *every item* in the checklist and upload a copy of the completed checklist with your submission.

Nature Human Behaviour offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't.** Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our [FAQ page](https://www.nature.com/documents/nr-transparent-peer-review.pdf).

We hope to hear from you within three weeks; please let us know if the revision process is likely to take longer.

To submit your revised manuscript, you will need to provide the following:

- Cover letter
- Point-by-point response to the reviewers (if applicable)
- Manuscript text (not including the figures) in .docx or .tex format
- Individual figure files (one figure per file)
- Extended Data & Supplementary Information, as instructed
- Reporting summary
- Third-party rights table (if applicable)
- Suggestions for cover illustrations (if desired)

Consortia authorship:

For papers containing one or more consortia, all members of the consortium who contributed to the paper must be listed in the paper (i.e., print/online PDF). If necessary, individual authors can be listed in both the main author list and as a member of a consortium listed at the end of the paper. When submitting your revised manuscript via the online submission system, the consortium name should be entered as an author, together with the contact details of a nominated consortium representative. See <https://www.nature.com/authors/policies/authorship.html> for our authorship policy and <https://www.nature.com/documents/nr-consortia-formatting.pdf> for further consortia formatting guidelines, which should be adhered to prior to acceptance.

Forms:

Nature Human Behaviour has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Once your paper is accepted, you will receive an email in approximately 10 business days providing you with a link to complete the grant of rights. If you choose to publish Open Access, our Author Services team will also be in touch at that time regarding any additional information that may be required to arrange payment for your article.

For information regarding our different publishing models please see our <https://www.springernature.com/gp/open-research/transformative-journals> Transformative Journals page. If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

Please use the following link for uploading these materials:
Link Redacted

If you have any further questions, please feel free to contact me.

With best regards,

[Redacted]

[Redacted]

Nature Human Behaviour

Reviewer #2 (Remarks to the Author):

I was one of the original reviewers of “Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?” and am pleased to see how this project evolved over time, as well as the final outcomes from the authors’ extraordinary work. I want to begin by commending the authors on conducting a rigorous set of studies, reporting their findings as transparently as they did, testing for robustness and heterogeneity of effects, and explaining what all of it means very clearly. This paper is an example of social science at its best.

I only have one remaining comment for the authors, which they can choose to ignore if they wish; it is something I would like to see in the discussion, but it does not *have to be* included, as it is a broader issue beyond just this paper. Despite the largely null effects they found, the authors continue to (implicitly) tout the value of “low-cost, scalable interventions” for encouraging empathy toward racial and ethnic outgroups. I wonder whether the authors think this is still the path the field should be taking, given their own null results. While they critique the deep-canvassing studies for their high cost and thus inherent lack of scalability, at the end of the day, those high-touch interventions produce very long-lasting effects. There seems to be an inherent tradeoff here: the light-touch studies are easier to do, but they don’t change very much or last very long; the high-touch studies are expensive, but they do produce lasting and meaningful changes. One of the big debates in the behavior change literature at this moment is how much of our time we should be spending on light-touch vs high-touch approaches, given their relative benefits vs. costs (for review, see Chater & Loewenstein, 2023). Given the findings of this manuscript, I am curious about the authors’ thoughts on this larger issue.

References

Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, 46, e147.

Reviewer #3 (Remarks to the Author):

Thank you for submitting your Stage 2 manuscript -- I have been looking forward to seeing the results of the study. In line with reviewing a Stage 2 manuscript, I would first like to confirm that my assessment is that the authors have followed the approved study plan.

In terms of the Results, these are clearly reported and structured well for the reader. There are a number of analyses, but I did not feel lost at any point about what was being tested and communicated, so I commend the authors for writing a complicated Results section so clearly.

Where I would like to see some changes is with the Discussion, which seems very brief. A good portion of it is spent reminding the reader about what was done and what the strengths of the study were, which is fine, but then the portion spent explaining the results and discussing their implications was limited (especially given how much the Introduction covers). I would strongly encourage the authors to delve into explaining the results more and contextualising them more within relevant theory in order for the manuscript to reach its full potential. For example, while I'm aware that the authors weren't testing an increase in empathy as a potential mechanism for changing political behaviours, I was really left wondering about what may have driven the change in political behaviour observed, if there was no simultaneous change in empathy. Why would participants bother to donate or write letters, if there was no meaningful change in empathy? Are there theoretical, or possibly even methodological explanations for this? Furthermore, if the intervention works best with those who already have relatively high levels of empathy towards the outgroups, then what should researchers in the area do next for intervening to change empathy and behaviours in those who have lower baseline levels of empathy? (As an aside, the finding that the intervention seemed more effective for those with a higher baseline level of empathy made me wonder if this could be explained by the Matthew effect, so I was a little surprised that the authors do not mention it in any way.)

Thus, my overall assessment is that I very much enjoyed reading the paper and feel the rigour of the study and the set of

findings have a lot to contribute. However, I was disappointed that the Discussion was not as meaty as I hoped it would be, and so I strongly encourage the authors to expand this section. The paper has potentially a wide audience, with many future readers likely not having direct expertise in this research area, and so a more expansive and in-depth Discussion would really help the future audience know what the implications are both for theory and for potential application of the findings.

Minor points:

1. In the abstract, I think the phrasing around "can change attitudes" is a little vague, even for an abstract. Different phrasing, or perhaps a little more detail, would strengthen the summary of the study here.
2. I noticed a typo in the Introduction: the word "beliefs" is repeated, this occurs in the second full paragraph on page 3, towards the end of the third line.

Reviewer #4 (Remarks to the Author):

The authors have succeeded in addressing my earlier suggestions and comments and (I believe) that this has strengthened the overall quality of the current MS. While I do recommend the current version for publication, and after having read the new revised version in its entirety, I still believe that two issues remain to be addressed (elaborated). The introduction is still missing an argumentative explanation for the intervention selection. While the authors did argue why empathy is relevant for reducing outgroup bias (e.g., prejudice), there still remains a question why this particular intervention was selected and not other established (and even more effective) strategies (such as norms messaging, contact etc.). A more systematic rationale for selecting empathy and why it was expected to be the most appropriate intervention would strengthen the introduction part of this MS.

Second, the study did not systematically account for contextual and individual differences that might moderate intervention effects. Even though the authors do acknowledge this throughout the MS, I believe a more thorough discussion regarding this limitation deserves to be included in the discussion section. Variability in participants' contextual background (where they live), or prior intergroup experiences could have influenced both receptivity to and outcomes of the intervention. Authors should, at least, argue that future research would benefit from explicitly and systematically aligning key (relevant) factors to better understand for whom and under what conditions empathy-based (but also all other intergroup) interventions are most effective.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Dear [REDACTED]

Thank you for providing the opportunity to revise and resubmit our pre-registered report, “Can praise from peers promote empathy and inclusive behavior towards racial or ethnic outgroups?”

We are grateful for the thoughtful and supportive comments provided by you and the reviewers. As we hope can be seen from the memo below, we took all suggestions extremely seriously. Below, we detail how we have revised the manuscript in response to suggestions and questions. We hope you find these responses and the revisions to the manuscript to your satisfaction. If further revisions would help make the report more acceptable to NHB, we are happy to attempt them.

We anticipate a timeline of 3 months for completing the study if our Stage 1 submission is accepted in principle. We agree to share our raw data, any digital study materials and all computer code and the eventually published results. We confirm that, following Stage 1 in principle acceptance, we agree to register our approved protocol on the Open Science Framework (<https://osf.io/>) or other recognised repository, either publicly or under private embargo, until submission of the Stage 2 manuscript. We further confirm that if we later withdraw our paper, we agree to the Journal publishing a short summary of the pre-registered study under the section “Withdrawn Registrations.”

Sincerely,

The authors of Ms. No. 23113998

1 Summary of reviewer comments

R1 summarized their views by noting that “they see potential in the approach of the authors (testing a simple, probably scalable [intervention] to reduce outgroup bias).” R2 summarized it as an “interesting manuscript” and noted that, “the manuscript is well-written, the proposed intervention has been well-conceptualized and vetted in pilot tests, and the proposed registered report studies are both feasible and reasonable tests of the central research question.” R3 noted that “the general topic area of the study and that the research question are valuable” and that “this work would make an important contribution to the literature and that it would be valuable for us to have this work done and disseminated.”

Nonetheless, the reviewers and editors made a number of helpful suggestions in order to improve the manuscript. The most significant of these comments (as highlighted by the NHB editor) were to revise the design to include longitudinal evidence, better motivate the choice of mediators, report Bayesian statistics and endeavor to make the treatment/control more comparable. The other high priority item flagged by the editor and reviewers was to improve the clarity of the writing throughout the manuscript.

2 Editor

1. Reviewer 1 finds it crucial for the research question that longitudinal evidence is provided. Please follow the reviewer’s suggestion and add a follow-up to your design.

We have revised the design to include a longitudinal panel element. See our response to R1 on

page 3 of this memo.

2. Our reviewers have various methodological concerns and suggestions. For example, Reviewer 3 asks that the control condition is revised to be comparable to the treatment condition and Reviewers 2 and 3 suggest revising the donation task and its instructions. We ask that you carefully consider all of these concerns and address them in full in your revised manuscript.

We have endeavored to answer all reviewer questions regarding methodology, and taken every suggestion that we feasibly could to aid in identification and transparency. Regarding the specific issue of the control condition, we now include a placebo treatment arm designed to preempt concerns that the treatment works through extraneous elements such as color of text (discussed on page 11 of this memo). Our response to the question regarding the donation task is discussed on page 7 of this memo.

3. Our reviewers find the writing to not be clear and the manuscript to lack important methodological details. Please revise the writing to be clear and concise and provide all necessary methodological details.

We have heavily revised the writing and structure of the paper throughout to aid in clarity and eliminate redundancy. We have also included our full draft survey instrument as well as recruitment materials, and re-organized our SI to aid in readability. Our responses to more specific questions about the writing can be found on pages 4 and 5 (transparency and replication). And while many of the responses in the memo implicate revisions we've made in the writing of the paper, the majority of them can be found in response to R3 beginning on page 5 of this memo.

In your revised manuscript, please also explicitly motivate the choice of mediators and explain how confounds will be dealt with.

We have improved our discussion of mediators as well as revising our design to better to investigate the plausibility of different potential causal paths. Our discussion of mechanisms/mediators can be found on page 3 of this memo.

4. Reviewer 3 suggests that you also report Bayesian statistics. We agree with the reviewer and ask that you commit to a method (such as Bayes Factors) for interpreting the strength of null results.

We have taken R3's suggestion and outline on page 5 of this memo our plan to report Bayes Factors.

3 R1

While the pilot studies already provided some evidence for the immediate effect of the intervention on empathy and outgroup bias, evidence for long-term effects is missing, but in my view critical to back up the claim of the authors that this intervention is a simple and promising one. So in my view, the authors should add a follow-up (e.g., assessing empathy and indicators of outgroup bias some time after the intervention) in their design. This also helps to reduce concerns that the immediate effects might be mainly a result of priming, inducing positive mood (one of the proposed mediators!), and/or demand characteristics.

By far the most significant revision requested by the reviewers and Editor relates to the design itself. R1 wrote that, “evidence for long-term effects is missing, but in my view critical to back up the claim of the authors. . .” R3 suggested the same, though they argued that such a revision would be “out of the scope” of this study. The matter was settled by the Editor, who requested that we take R1’s suggestion to add a longitudinal element to the design.

In general, we see our diverse group of outcomes—including choices to engage in costly, real (behavioral) empathy, self-reported empathy, and measures of attitudinal and political inclusion—as a significant contribution to the literature, which has relied largely on self-report measures (only 7% of all outcomes in a recent meta-analysis were explicitly behavioral; see Paluck et al., 2021, 539). However, we agree wholeheartedly that a longitudinal element in the research design would be beneficial and greatly improve the contribution of the paper.

We see the panel design as accomplishing several things simultaneously. First, long-term effects are a tougher test of our intervention; while changing behaviors immediately following treatment is (we argue) an important task, doing so after a delay is an even taller order. Second, as R1 points out, a panel design can help disentangle mechanisms; to the extent that evidence for long-term effects is found, it’s unlikely that those result from either priming or “immediate positive mood,” which presumably would decay after a short period of time. Finally, the longitudinal design takes up the challenge set out by recent reviews (Paluck et al., 2021, 555) to investigate whether effects from empathy-related interventions persist over time.

What has changed: We have re-designed our experiment so that the main study (“peer praise”) has a panel design in which respondents from Wave 1 are re-contacted 1 week following treatment and outcomes are re-measured (see Figure 5 Main Study Consort Diagram in revised manuscript). To aid slightly in power, we create two indices for political inclusion and attitudinal inclusion, whereby the former is composed of two measures on donation and letter writing, while the latter is composed of two measures on social distance and thermometer rating. For longitudinal hypotheses, we explicitly register and power our design for the empathetic *behavior* and *attitudinal index* outcomes (Hypotheses 5-6 in Table 1 in the manuscript). The construction of the indices is detailed in the Methods section.

I also missed more arguments that back up the claim of the authors that this kind of intervention is more easy to scale up compared to alternative interventions. For instance, it is not clear to me in which way the authors plan to implement this intervention in different contexts (i.e., instructions for those who implement the intervention; instructions for participants), how they make sure to reach individuals that mainly profit from this intervention (i.e., highly prejudiced individuals), whether a single “treatment” is sufficient, etc.

We see these comments as being about two related issues: calibrating the relative cost of out

intervention compared to others and providing a blueprint (and sufficient detail) for other researchers to adapt our intervention.

What's changed: To address the transparency aspects, we now include the following in a new section called “Scale up of Peer Praise” (SI Section C.3):

1. a figure (Figure C.41 in Appendix C.3) that visually depicts the process of scaling up for researchers and readers for either an online or in-person setting.
2. a set of minimal working example (MWE) survey questions to help facilitate replication and scaling up for others.
3. MWE qsf files (from item 2 in this list) for fielding a peer praise generating survey in Qualtrics (to be placed in the open-source repository of materials with all other study materials). These are discussed as well on p.3 of the introduction.

To address the former issue, our Introduction now features a brief discussion of the relative cost of our peer praise intervention compared to existing options. The main cost of our intervention is the initial survey to generate real peer praise among whatever group one is interested in treating, since the actual stimulus materials are of negligible expense (the word cloud and ratings). The cost of the initial survey to generate praise is small, and especially so when compared to in-person interventions which require between ten and twenty minutes per person of contact with trained enumerators, multiplied by thousands of people (Kalla and Broockman, 2020, 2023).

Finally, we note that nothing in our design requires the intervention to be targeted at “highly prejudiced individuals.” However, should one wish to target specific subgroups, doing so would only require using data from baseline outcome surveys to target the individuals that are theoretically or normatively most relevant in a given context. In exploratory analyses from our pilot studies 3 and 5, we measure a self-reported empathy score and find that the peer praise intervention encourages choices to engage in empathy among respondents in all terciles of baseline empathy scores, with statistically significant results for high and medium empathy (see Appendix Figure B.33). Our pre-registered study also includes pre-treatment measures of social distance, thermometer and self-reported empathy so in order to conduct (exploratory) investigations of the effect of our intervention among different subgroups.

I also missed a more critical review of the literature on empathy interventions, especially those who mainly target general empathy, since there is mixed evidence that inducing empathy helps to overcome outgroup bias (see Behler, & Berry, 20022). I also wondered why the authors target general empathy, not directly empathic concern toward outgroups.

R1 noted that they would like us to see a more structured and critical review of the empathy intervention literature, incorporating Behler and Berry (2022) and the distinction between “general” and “parochial” empathy. They had a related question about our intervention, asking why we target “general empathy” rather than empathy for outgroups specifically. Our answer to that question is that Behler and Berry (2022) make the point that while increased empathy “is a reliable promoter of prosocial behavior,” bias towards ingroups is more likely to occur “when the group membership of an empathy target is made salient (p. 4)” and that researchers should leverage “external social motivations” to reduce parochial empathy (p. 10). In order to avoid making anything about group identity salient (and inadvertently triggering parochial empathy), our treatment mentions only empathy (not any social groups or divisions) and the group

identity of the targets is never mentioned as it is only treated via the photograph of the face. And following from the exhortation to “leverage external social motivations,” our treatment uses praise from peers in our attempt to promote empathy towards racial and ethnic outgroups.

We now cite and discuss Behler and Berry (2022) in the Introduction on p.2 of our manuscript as part of our discussion of the empathy gap. Additionally, in our our discussion of the treatment itself (on p.3 of the manuscript), we discuss the rationale for our “general empathy” treatment that leverages social motivations while avoiding making any group divisions salient.

I’m also a bit sceptical about the proposed mediators. Positive emotions might be partly responsible for the documented effects of the intervention in the pilot studies. But this means that long-term effects of the intervention are less likely. Because of the design of the intervention (manipulation), some sort of norm information is given to the participants. However, there are already simple and easily scalable interventions tested that directly target social norms (Murrar, Campbell, & Brauer, 2020).

Our response is two-fold. First, we agree with R1 that to the extent that positive emotions account for our findings in the pilot, our peer praise treatment would be less likely to promote empathy towards outgroups in a longitudinal setting. Because of our significant shift towards a panel experimental design, we are now able to assess this empirical implication directly.

Conceptually, we agree with R1 that norms are a likely mediator, but note that existing work suggested by R1 (Murrar, Campbell and Brauer, 2020)—while simple and easily scalable—did not measure or address behavioral outcomes towards marginalized groups (though it’s a great paper with a fantastic and rare longitudinal design). Their core finding was that their norms intervention affected “positive attitudes” towards outgroups” and “greater appreciation of diversity” among non-marginalized groups. In that sense, we see our work as building upon theirs by by incorporating direct behavioral measures of empathy towards outgroups.

What has changed: We thank the editor and R1 for the suggestion to more explicitly motivate the mediator choices, and there is now a brief discussion of this in the Introduction where we note that our two mediators (positive emotions and norms) derive from the two components of our intervention (praise and peers, respectively). Discussion of our two potential mechanisms has been revised for clarity, to incorporate more recent literature suggested by R1 and to more explicitly motivate our choices as per Editor’s suggestion. It is now in the restructured Introduction section on p.3. We have also revised our design by adding a direct measure of social norms, enabling us to better test that potential mechanism. Both measures for our mechanisms — norms and emotions—are now discussed on in the Methods section on p.9.

The authors provide many details about their planned study, but important information is missing that would enable other researchers to replicate the experimental procedure (e.g, instructions given to the authors). Although they plan to make all data and materials openly available, I think it is important to include all study materials (instructions, items, etc.) in this preregistered report.

In response to the request for more information, we have attached the full Qualtrics survey instrument—including recruitment materials—in SI section C.4. We have also added a new section called “Scale up of Peer Praise” (SI Section C.3), which includes directions and materials required for other researchers to generate and field peer praise experiments based on our studies.

4 R2

My biggest methodological suggestion (if financially feasible) is for the authors to take a stratified sampling approach in their proposed studies—specifically to recruit equally sized samples of Democrats and Republicans so that they can test for intervention moderation by partisanship. The reason for this recommendation is due to the results the authors found in Pilot 7. The authors ruled out social norms as the mechanism through which their intervention might operate, but in my view, their Pilot 7 results suggest **political norms** might in fact matter greatly for whether the praise intervention would yield empathy and political inclusion. Specifically, on page 11, the authors noted that Democrat co-partisan praise had a positive effect on empathizing and choosing to learn about BLM, while Republican co-partisan praise had a non-significant or negative effects. That suggests to me that for a peer praise intervention to be effective, peer praise must be aligned with what is normative for the group. In contemporary American politics, Democrats are the party of racial inclusion whereas Republicans have abandoned inclusion as a priority (Levitsky & Ziblatt, 2019, 2023). It makes sense, then, that when Democrats are praised for inclusion efforts, that has a positive effect on their behavior, whereas the same intervention has a different effect on Republicans. That kind of heterogeneity is important to consider (Bryan, Tipton, & Yeager, 2021), and I think the authors' own data suggests it should be factored into the research design—perhaps as a 2 (Intervention, Control) X 2 (Democrat, Republican) factorial—again, if that is financially feasible (they would need to redo their power calculations to figure out how such a design change would affect sample size).

R2 suggests that “political norms might...matter greatly for whether the praise intervention would yield empathy and political inclusion.” This follows from some of our findings in Pilot 7, and R2 goes on to suggest that perhaps “peer praise must be aligned with what is normative for the group.” Given that Democrats have been described as the party of inclusivity (Levitsky and Ziblatt, 2019), one might expect Democratic respondents to respond more to praise for empathy than Republican respondents, though conversely if the correlation between party ID and empathy towards racial outgroups is high enough, it's possible that “ceiling effects” among Democrats might suggest the opposite prediction.

R2 suggested that this should be incorporated into the research design to the extent that it is “financially feasible.” Unfortunately, it is not financially feasible to add both a panel component to our design *and* power to detect heterogeneous treatment effects by party. However, we mention the possibility of these heterogeneous effects by party and discuss why they may be present on p.6 of the revised manuscript and commit to an exploratory investigation of these effects in the paper, following the pre-registered analyses.

A second methodological suggestion is to consider how the use of racialized faces from the Chicago Face Database might ultimately affect inferences. Recent research from face perception researchers suggests that this practice can distort our inferences, especially when trying to make claims about race and racism (see Martinez, 2023 for a more thorough description of the issue at hand).

R2 asked us to consider how the use of “racialized faces” from the Chicago Faces Database might affect our inferences, referencing research (Martinez, 2022) suggesting that this practice

can distort inferences “especially when trying to make claims about race and racism.” The crux of the issue is whether the use of faces classified as different “races” (according to the Faces Database) neglects the perspective that “race is illusory” and (intentionally or not) “essentializes race (Martinez, 2022, 2).”

In order to address these inferential and normative issues, we revise our manuscript in two ways. First, we take the advice of Martinez (2022, 3) to clarify in our revised manuscript that the faces represent “racialized groups” rather than races, and cite the appropriate literature. Second, since Martinez (2022, 7) points out that using standardized databases can induce inferential slippage—since one’s sample might perceive the faces differently than those who aided in the construction of the dataset—we propose using data from our actual participants to help verify the perception of racial groups. Specifically, all participants complete a “describe” task as part of their introduction to the empathy task, so we propose to use that text data to verify that subjects’ perceptions align with the coding of racial groups in the Chicago Faces Database by conducting text analyses for correlations of text mentions of race and ethnicity for each image with the actual image race/ethnicity.

Ultimately, and as Martinez (2022) makes clear, there is no clear and easy “fix” for these issues: for instance, turning the racialized groupings into our DV as Martinez and colleagues suggest, is not an option in our design. We believe our response addresses the inferential issues at stake and hope that it also sparks a thoughtful debate about the broader normative issues as well.

A third methodological suggestion is to consider a broader set of donation options for the donation DV. I know that the authors already pilot tested BLM and UnidosUS, but I am concerned about how recent shifts in attitudes about BLM might affect responses. Specifically, BLM has become such a charged movement, especially in the backlash against the 2020 racial justice movement, that it is even less popular now than it used to be (see Chudy & Jefferson, 2021). Therefore, I wonder whether lack of willingness to donate to BLM means the same thing as it did when the authors ran their pilots.

R3 asked about our donation outcome measures, noting that though we had piloted BLM and UnidosUS, recent shifts in attitudes—particularly declining support for BLM¹—might affect responses. While R3 is certainly right to point out that BLM experienced a significant drop in its popularity in the summer and fall of 2020, Braga (2023) shows that by 2021 and 2022, BLM support had stabilized (with a little more than 50% Americans in support and under 50% opposing). Because our BLM outcomes were piloted in 2021, *after* the significant dip in approval, we believe that our pilot data already incorporate the lowered BLM approval, making comparisons to our pre-registered study more plausible. To the extent a further shift in attitudes has occurred since the fielding of our pilots, however, it might imply that any treatment effects on those outcomes would be more difficult to achieve (and thus figure into our interpretation should we find such effects).

My (more minor) theoretical point is related to the discussion of empathy and perspective taking interventions in the introduction. Recent research suggests perspective-taking is

¹Documented in “Support for Black Lives Matter Surged Last Year. Did it Last?” *New York Times*, May 2022 2021.

not the most effective pathway (in part because perspective taking amplifies egocentric biases); instead perspective-*getting* is more fruitful (see Eyal, Steffell, & Epley, 2018). This distinction does not change the design of the proposed study, but I do think it is worth discussing either in the introduction or discussion.

R2 asks us to incorporate the distinction between “perspective-taking” and “perspective-getting” into our discussion of the literature. While we agree with R2 that this distinction does not imply any changes in our design, we have now incorporated this conceptual dichotomy into our discussion in the paper. Specifically, we now cite and discuss Eyal, Steffell and Epley (2018) in our discussion of extant perspective-taking (and getting) interventions in the Introduction.

5 R3

In general, I found much of the manuscript difficult to read. This is not a comment about the writing style per se, but more about the density of the information and how that information has been structured. I have a nagging feeling that some of my questions/concerns below may already be addressed in the manuscript, but I have missed these because of my difficulty in gaining a confident understanding of some portions. In the next few comments, I have outlined a few places where a clear understanding seemed particularly difficult, but I would advise the authors to cast a critical eye over the whole manuscript and improve the clarity wherever possible. Indeed, reading the manuscript felt too effortful, and I believe this will lead to disengagement by future readers.

We have taken all the advice we have received and re-structured and re-written a great deal of the manuscript with an eye towards improving clarity and cohesion. See more specific answers below. We feel as though the manuscript has been improved significantly as a result and hope the editors agree.

Following on from (1): When the Introduction moves to the description of peer praise interventions and previous work, I found it very difficult to follow (page 4 and some of page 5). This section is very dense and I found it hard to keep track of what is being proposed in the current work, what had come through pilot work, and what had come from previous publications. I think the clarity of this portion of the Introduction needs to be improved so that it is easier to grasp the key points with just one read-through. For example, one change could be that the Introduction needs to be broken down into sub-sections with informative sub-headings.

We have taken all the advice we have received and re-structured and re-written a great deal of the manuscript with an eye towards improving clarity and cohesion. While we could not take all the specific suggestions—for example, NHB does not allow subheadings in the introduction as far as we know—we strove to take the spirit of the suggestions to improve the overall readability of the manuscript. In any case that we could take the specific suggestion—such as R3’s suggestion for a table with main hypotheses listed in a clear and intuitive way—we did (e.g., Table 1 in the paper). Overall, a great deal of attention and care went into re-writing the Introduction to address R3’s concerns and improve the paper.

The Introduction seems to jump around a lot. For example, near the top of page 5, the first full paragraph is general information about empathy, but such information had been introduced on page 3. It's not clear why this type of information is then returned to on page 5. After this, there is general literature review information about empathy-based interventions and, again, this feels misplaced. The authors should think about restructuring the Introduction so that all the background information from previously published work comes first, before moving on to information about the pilot work and their plans for the current work. Page 7 then describes the planned work in more detail, but much of this overlaps with information on page 4, which adds to the difficulty of following the Introduction.

See answer above, in which we describe our revisions. Broadly, we have taken R3's advice to revise the entire introduction.

I would prefer to see Hypotheses 7-12 explicitly listed in the manuscript. Since they are very similar to H1-6, it might help to integrate them, particularly if the Introduction is rewritten more generally as suggested. For example, they could be written separately but linked as H1a and H1b etc., or they could be written in a more combined way like "H1: Peer praise (Study 1: Neutral Peer Praise; Study 2: Co-partisan Peer Praise) increases the likelihood. . ."

We have taken R3's specific suggestion for a table with main hypotheses listed in a clear and intuitive way, which can be found in Table 1 in the paper.

Like the Introduction, the section describing the Pilot Studies is very dense (and repetitive with some aspects of the Introduction). There is detail about the pilot studies already in the supplemental information, but can much of the narrative about the pilot work be moved from the manuscript into supplemental material as well? The manuscript could then contain a more concise section summarising the key take-aways from the pilot studies.

We have taken R3's suggestions, condensing the discussion of pilot studies in the Introduction to one paragraph and a list of key empirical findings, and moved nearly all of the rest of the discussion of details from the Methods section to a (heavily revised) summary of the studies in Section A of the supporting materials. The Supplementary Appendices have also been simplified into three and only three sections: ethical considerations, pilot studies (including a brief narrative summary as well as all materials and analyses) and pre-registered study.

In Figures 2 and 3, many of the words are cut off – is this how they actually will appear in the study? It will be effortful for the participants to read the word clouds because of this (which will already be difficult for some neurodiverse participants). Where participants persist with the word clouds, you will get different outcomes where some will put in the effort to guess the words (and there may be different options for some, like "excel" – is this "excel" or "excellent"?) and some will not bother. Thus, participants will have different experiences of the word clouds because of this effortfulness. It would be best if all the words and phrases on the clouds are clearly displayed in full.

R3 asked a series of questions about the word cloud treatment, notably whether it would be effortful for respondents to read the word (because they are “cut off”) and whether this would result in different outcomes for different participants. As a reminder, we used word *stems* to summarize our peer praise for empathy, and as is common in word clouds, used stems to group similar words. For example, “empathetic,” “empathy,” “empathize” and “empathizer,” “empathizes” and “empathizes” would all be summarized by the stem, *empath-*.

Briefly, our response is that effort should be balanced across treatment conditions (meaning it would not result in challenges to our inferences), and while overall higher effort could result in attrition, that is something that we already investigate carefully. And while it’s possible that respondents may interpret words differently, even in the example provided by R3, “excel” and “excellent” do have similar meanings (and the same Latin root, *excellere*). While we see the use of word stems (as a way of grouping similar words) as rather straightforward and uncontroversial, we have incorporated an end-of-survey manipulation check that presents respondents with the treatment once more (if they had been randomized to received treatment previously) and asks respondents to ruminate and write about what they thought of the treatment (including word cloud). We expect text answers to this question to reflect understanding that the words in the word cloud are of praise and positive support towards empathy, with instances of respondents directly using words that have roots shown in the cloud as further evidence of such understanding.

My understanding of the experimental design is that those in the praise conditions will see a word cloud and that those in the control condition will see nothing at all. Is this correct? Assuming it is, from an experimental control perspective, it would be better for the control participants to have something to look at, for example, a word cloud of neutrally rated, concrete words. Technically speaking, any differences could be attributed to other features of the manipulation and not necessarily to the praise words themselves (e.g., maybe the bright, happy colours of the word cloud lifted the participants’ mood).

This question is fundamentally about *how* the treatment works; if it’s a bundle of different things (e.g., praise from peers + “bright happy colours of the word cloud”), then we may be less sure of how much causal work is being done by the peer praise alone.

First, we note that interventions that are bundled in some way and compared against simply a “no treatment” control condition are commonplace in contexts where treatments are visual or in field experiments. For example, consider Murrar, Campbell and Brauer (2020, 895), Experiment 1 in *Nature Human Behavior*, where a poster about the value of diversity (including statistics about valuing diversity, headshots of students, diagrams of the state in which they lived, etc.) was compared with a very different control with text about a flu vaccine (and no headshots, statistics, etc.). In Experiment 2, that same poster was compared to a control condition of “no treatment,” rather than a different poster. The same basic point could be made about many studies: interventions are often compared against control arms in which more than one thing is changed.

However, we take R3’s point that from an experimental control perspective, such a design could be improved. In fact, this is something we tried to address in an early pilot. Pilot 3 described in Appendix C found that peer praise for empathy promoted decisions to engage in empathy at a higher rate than our placebo condition (though this difference was not statistically

significant in our underpowered pilot).

What has changed: To improve upon our early pilot and take R3's suggestion, we have added a third arm in our main experimental design for peer praise that includes a placebo (Arm C in Figure 5) in which respondents see peer praise for something other than empathy. That information appears similar to peer praise for empathy—information in the form of a word cloud and descriptions by peers, similar colors, etc.— but provides praise for *objective description*. It is presented without deception as coming from their peers, and was collected from Pilot 2 (described in Appendix B). If the elements of the treatment that matter for promoting empathy are only presentational, then “praise for description” should also cause a higher likelihood of choosing to engage in empathy in our choice task. While our main pre-registered hypotheses still features the contrast between the peer praise intervention and a neutral control, our exploratory analysis and robustness check for our treatment are discussed in the Introduction on p. 5, the Design section on p.6-7, and Methods section on p.11. This placebo group is compared against both our control arm as well as our peer praise arm as a robustness check for the peer praise effect.

The self-reported empathy item is double-barrelled – participants are asked about “how they are treated by police and fellow citizens” and “the hardships they face due to discrimination by race and ethnicity” – these are two different things and participants may be inclined to give different responses (one could even argue that “treatment by police,” and “treatment by fellow citizens,” are also separate things). Can participants be asked to respond to these as separate items? (even if the responses are then averaged for the purposes of the analysis). I believe that doing so would increase validity.

R3 asked if the self-reported empathy items were double-barreled and asked if they could be made into separate items and then summed into an index. We thank R3 for this suggestion and have adjusted our measure accordingly (see p.8), asking respondents how much they agree with the statements (1) “I feel empathy for the grievances expressed by Black and Hispanic Americans about how they are treated by police”, (2) “I feel empathy for the grievances expressed by Black and Hispanic Americans about how they are treated by fellow citizens”, (3) “I feel empathy for the hardships they face due to discrimination by race and ethnicity”. We combine the responses to the statements into a single “Self-reported empathy” outcome measure (normalizing each and adding to an index taking values from 0 to 3).

I wonder if the White House letter-writing task would be affected by political identity. I understand that the analysis will check for balance between groups in terms of covariates but will identity be entered into the analysis as a statistical covariate at all? Similarly, this kind of task may also be impacted by education level. Will this be controlled for statistically?

R3 noted that the White House letter-writing task “might be affected by political identity (and education)” and queried if these variables would be controlled. We address this in two ways. First, we will control for education and Party ID if imbalanced, as detailed in our Analysis Plan (p.10). The question also implicates the issue of heterogeneous treatment effects, and we plan to explore HTEs by partisan ID, as mentioned earlier (p.5).

In the donation task, the text of the item simply says "a portion." This is really unclear and how this is interpreted may massively impact the responses. This is because such lexical interpretation differences cannot be measured as a covariate and this could differ between groups. Can "a portion" be specified? Will it be specified? Or can participants choose how much of a portion they want to donate? (The final suggestion may even be more informative than just yes/no.)

R3 asked us to clarify the options for the donation task, which we now do in SI section C.4 which provides the full survey instrument. Our donation measure consists of two questions, each paired with a slider from 0 to \$0.5 for respondents to allocate any portion—inclusive of none or all—of their bonus to BLM or UnidosUS.

The measures are described but oftentimes this seems to leave out the response scales (Likert scales). Can these be clearly given or described in the manuscript for all measures?

In response to the request for more information about the response scales from R3 and the editor, as well as what information is given to respondents, we have attached the full Qualtrics survey instrument—along with recruitment materials—in SI section C.4.

I would like to see what information is given to participants before they begin, perhaps as part of the supplemental information. The manuscript alludes to a lack of deception, but it is difficult for me to then understand what participants are told about the purpose of the study and the manipulation. This may impact how they respond so I would like to see this.

See answer directly above.

The power analysis indicates 400 participants. If my understanding is correct, this means 200 per group (i.e., not 400 per group). It would be helpful to know more about any inclusion/exclusion criteria in relation to this sample size. For example, my understanding is that anyone can participate but then the analysis will focus on white participants. Thus, does the "400" refer to white participants who passed the attention checks etc. (i.e., after other datasets are excluded), OR does the "400" refer to everyone who starts the study and is randomised, in which case the analysis will end up with less after excluding datasets from people of colour, from those who failed the attention checks, etc.? (i.e., will you replace excluded datasets to ensure that there are 400 useable datasets? And are you ensuring that there are 200 useable datasets per group, or is it possible that the praise vs control group sizes might differ?)

R3 asked a question about our power analysis, specifically whether the numbers we refer to are the total number recruited or the number who fit all criteria and remain in the study. The answer to that is that the "N required" for our studies take into account that only a subset of those recruited will be usable for our analysis; i.e., our power calculation plots depict the *total number of recruited subjects required*. The code that we include in Appendix Section C.2 takes

into account that only a % of total respondents will (1) be white subjects (2) pass both attention checks (3) reside in the U.S. (4) be Republican/Democrat (for Co-Partisan Study only). We draw on past pilot studies to estimate those percentages. For example, pilot data suggest only 66% of respondents identify as white, so our power analysis code takes that proportion into account when calculating how many subjects will need to be recruited in order to end up with the required sample size for each hypothesis.

What platform is the survey held on? Is it directly on MTurk or on something else, like Qualtrics?

We clarify on p.9 that the survey is built using Qualtrics but fielded on MTurk.

I was glad to see checks for the effects of attrition built in. I am assuming that participants will be able to skip items...? If so, will patterns of missing data be explored? (i.e., for skipping items rather than attrition overall)

We will explore both types of missingness in our final collected sample. Indeed, we will use the same set of visual and statistical assessments that we used in our pilot studies (Appendix B Pilot) where, among other visualizations, we show when and where respondents skip questions.

Will participants be excluded only if they fail both attention checks, or will they be excluded if they fail just one?

We clarify on p.10 that respondents are excluded if they fail *one* of the two attention checks.

Even if a particular null hypothesis can be rejected, why not use Bayes Factors for assessing the evidence in favour of the alternative hypothesis?

R3 suggested that we revise our analyses to use Bayes Factors for “assessing the evidence in favor of the alternative hypothesis” and the Editor concurred. We agree with this line of thinking and in response, include explicit Bayes Factor assessment plans in both the contingency plans section and design tables. Specifically, we note that should the treatment effect size(s) be statistically indistinguishable from zero, we interpret this as a lack of evidence for rejecting the null; how statistically significant the evidence is against the null will be calculated through Bayes Factors and interpreted via Jeffreys rules.

It is not clear to me why the two types of praise will not be compared against each other. This was briefly discussed in the context of one of the pilot studies, but it would seem valuable to test this here. Knowing that one works better than the other, or that they work as equally well as each other, seems worthwhile to analyse. The effect is likely to be smaller, of course, but it could be included even as an exploratory analysis as I think this is the obvious question that most readers will be asking.

R3 suggested that we compare the efficacy of peer praise that comes from co-partisans to peer praise that comes from MTurk workers in an exploratory analysis. We have committed to exactly that in the revised manuscript on p.5.

While it's clear that the study is focusing on "Americans," how is this being defined? Do participants have to be natural-born or naturalised citizens? What if they are living in the U.S. but are not citizens? What if they are temporary immigrants? Furthermore, there may be American citizens who have lived abroad for a long time. It may be helpful to gain information on these aspects for use in the analysis and in determining the final sample.

We now clarify in the Sampling Plan section we sample 18+ year old respondents on the MTurk platform who are physically registered to MTurk in the U.S. as part of the inclusion criteria. We further ask respondents to select their country/countries of citizenship in our sociodemographic questions to adjust for and consider treatment heterogeneity via citizenship in exploratory analyses.

Minor comments:

In the first paragraph of the introduction, I would suggest having an example of the health consequences of outgroup bias that is instead of or in addition to Medicaid, since that is very US-specific, and the "US context" has not been introduced at this point. A more general example will better show that this is a widespread phenomenon.

We've included a reference to a health consequence of outgroup bias that is more broad in our introductory paragraph (Karlsen and Nazroo, 2002).

There is a missing closing parenthesis in line 6, after "towards outgroup members"

This has been fixed.

In the second paragraph there are some missing words here: "referred to as "perspective taking" and)" – unless the "and" is meant to join the citation numbers; it's unclear to me.

This has been fixed.

In the Introduction, it would be better to avoid the word "you" in this sentence: "(14,704 observations once you account for multiple trials)" perhaps by changing it to something like "14,704 observations after accounting for multiple trials"

This has been fixed.

In most places, the authors refer to "black Americans" but there is the odd place where it is "Black Americans." This should be double checked for consistency.

We have edited the manuscript for consistency in capitalization of “Black.”

6. Near the bottom of page 7, I believe there is a typo: “latino” – elsewhere it is written as “Latino/a”

This has been fixed.

7. There is a lot of word-for-word repetition between the Analysis plan and Contingency plans sections, which seems unnecessary.

The Sampling and Contingency Plan sections have been revised to eliminate redundancies in language.

References

- Behler, Anna Maria C and Daniel R Berry. 2022. "Closing the empathy gap: A narrative review of the measurement and reduction of parochial empathy." *Social and Personality Psychology Compass* 16(9):e12701.
- Braga, Kiley Hurst and Dana, Juliana Menasce Horowitz. 2023. "Support for the Black Lives Matter Movement Has Dropped Considerably From Its Peak in 2020."
URL: <https://www.pewresearch.org/social-trends/2023/06/14/support-for-the-black-lives-matter-movement-has-dropped-considerably-from-its-peak-in-2020/>
- Eyal, Tal, Mary Steffel and Nicholas Epley. 2018. "Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective." *Journal of personality and social psychology* 114(4):547.
- Kalla, Joshua L. and David E. Broockman. 2020. "Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments." *American Political Science Review* 114(2):410–425.
URL: <https://www.cambridge.org/core/journals/american-political-science-review/article/abs/reducing-exclusionary-attitudes-through-interpersonal-conversation-evidence-from-three-field-experiments/4AA5B97806A4CAFBAB0651F5DAD8F223>
- Kalla, Joshua L and David E Broockman. 2023. "Which narrative strategies durably reduce prejudice? Evidence from field and survey experiments supporting the efficacy of perspective-getting." *American Journal of Political Science* 67(1):185–204.
- Karlsen, Saffron and James Y Nazroo. 2002. "Relation between racial discrimination, social class, and health among ethnic minority groups." *American journal of public health* 92(4):624–631.
- Levitsky, Steven and Daniel Ziblatt. 2019. *How democracies die*. Crown.
- Martinez, Joel E. 2022. "Facecraft: race reification in psychological research with faces." *Perspectives on Psychological Science* p. 17456916231194953.
- Murrar, Sohad, Mitchell R Campbell and Markus Brauer. 2020. "Exposure to peers' pro-diversity attitudes increases inclusion and reduces the achievement gap." *Nature Human Behaviour* 4(9):889–897.
- Paluck, Elizabeth Levy, Roni Porat, Chelsey S Clark and Donald P Green. 2021. "Prejudice reduction: Progress and challenges." *Annual review of psychology* 72:533–560.

Dear [REDACTED]

Thank you for providing the opportunity to revise and resubmit our pre-registered report, “Can praise from peers promote empathy and inclusive behavior towards racial or ethnic outgroups?”

We are grateful for the thoughtful and supportive comments provided by you and the reviewers. As we hope can be seen from the memo below, we took all suggestions extremely seriously. Below, we detail how we have revised the manuscript in response to suggestions and questions in a point by point response to each. We hope you find these responses and the revisions to the manuscript to your satisfaction. If further revisions would help make the report more acceptable to NHB, we are happy to attempt them.

We anticipate a timeline of 3 months for completing the study if our Stage 1 submission is accepted in principle. We agree to share our raw data, any digital study materials and all computer code and the eventually published results. We confirm that, following Stage 1 in principle acceptance, we agree to register our approved protocol on the Open Science Framework (<https://osf.io/>) or other recognised repository, either publicly or under private embargo, until submission of the Stage 2 manuscript. We further confirm that if we later withdraw our paper, we agree to the Journal publishing a short summary of the pre-registered study under the section “Withdrawn Registrations.”

Sincerely,

The authors of Ms. No. 23113998

1 Summary of reviewer comments

R2—a reviewer on the previous version of the manuscript—wrote only, “I have now read the authors’ response to suggestions and am satisfied with the changes they made to address my (and other reviewers’) concerns.” R3, also a previous reviewer, concluded that “Overall, my previous comments and questions have been addressed well. In particular, the writing and flow of the manuscript is vastly improved, and so the background and the study plan are much clearer.” They also wrote, “I thank the authors again for the time and effort they have put in to taking my previous comments and questions seriously and for the well-considered ways they have addressed them.”

R4, a new reviewer, noted that “Research on psychological intergroup interventions is gaining momentum and, in my view, constitutes one of the most relevant and engaging research areas within the social sciences. This is one reason why I find this work both valuable and well-aligned with current trends in the field.” Further, they argue that “besides relevance and important contributions that this work might yield, there is also much to appreciate about the design. In particular, the longitudinal approach, combination of attitudinal and behavioural measures as well as inclusion of control/placebo group strengthen the design.” Finally, they note that some limitations of existing research are “directly addressed in this work” and that the manuscript is “good, timely and necessary.”

Nonetheless, the reviewers and editors made a number of helpful suggestions in order to improve the manuscript. The most significant of these comments were to revise our reporting of statistics, revise our interpretation of possible outcomes and take R4’s suggestions about more closely embedding the study in the broader literature.

2 Editor

SI, page 8, footnote 4: Please correct the typo in the reporting of the statistics.

Fixed.

Please revise the reporting of Frequentist statistics and include CIs around effect sizes.

Done.

Please note that we cannot accommodate italicized or bold-faced font in the main text.

Fixed.

Please follow Reviewer 4's feedback and clearly embed your research in the existing literature.

We have taken all of R4's suggestions regarding the literature, and the revised manuscript now reflects a deeper engagement with literature on social learning, intergroup interventions and a more in-depth motivation for our mechanisms.

The description of hypotheses must commit to interpretation of all potential data patterns (those that are predicted and those that would run counter to predictions). You cannot interpret lack of evidence (e.g. a $p > 0.05$ in a t-test) for the existence of an effect in null hypothesis significance testing as evidence for the absence of an effect. To be able to interpret data patterns other than the predicted effect or a significant difference in the opposite direction, you must commit to using Bayesian inferential methods or frequentist equivalence testing

We have revised the section, "Contingency Plans" (p. 13) to reflect this guidance. In the revised version, we commit to using Bayes Factors for interpretation given hypothesis tests for which we cannot reject the null hypothesis.

The Methods section should appear at the end of your manuscript (after the Discussion section).

Done.

Please use "white" (in lowercase letters) here.

Fixed.

Footnotes need to be removed and incorporated in the main text.

Done.

Please include the details of the power analysis in the main text.

We have added several paragraphs in the “Sampling Plan” section (pp. 11-12) with details of our power calculations.

Please provide the highlighted part of this statement in the Code Availability statement instead of in the Data Availability statement.

Done.

We ask authors to declare both financial and non-financial competing interests. For more details, see <https://www.nature.com/authors/policies/competing.html>. If you have no financial or non-financial competing interests, please state so: “The authors declare no competing interests.”

We have added the phrase, “the authors declare no competing interests” to the Methods section.

In your Design table, please provide the exact results of your power analysis, as well as provide the parameters that you used for the analysis.

We have edited the Design tables so they include power analysis results and parameters.

Your sub-section “Interpretation given to different outcomes” needs to be revised. Currently, this section simply repeats your expectations. This should be revised to be clear about statistical cut-offs. For Bayesian statistics, please specify how Bayes Factors will be interpreted.

We have edited the Design Tables’ “Interpretation given to different outcomes” column so that they provide statistical cut-offs and interpretation ranges of Bayes Factors.

Please remove Table 1 and convert Figures 9 and 10 into a Design *Table* (not a Figure).

Done.

3 R3

Previously, I queried if participants would be excluded on the basis of failing one or both attention checks. The letter says that participants will be excluded if they fail "one of the two attention checks" and refers to page 10 of the manuscript. However, page 10 does not seem to be completely clear about this as it says "Analyses will be conducted on all respondents who complete the survey and pass both attention checks. Participants will be selectively excluded from specific analyses if they... (b) fail one of the pre-treatment attention checks."

We now explain (on p. 13 of main text and in SI) that: Analyses will be conducted on all respondents who complete the survey and pass *both* attention checks. Participants will be excluded from analyses if they (a) do not provide a response and are thus missing variables involved in the analysis or (b) fail either of the pre-treatment attention checks.

I appreciate the detailed response to my questions about the words appearing to be "cut off" in the word clouds, i.e., the clarification that these are word stems, as mentioned in the letter: "As a reminder, we used word stems to summarize our peer praise for empathy, and as is common in word clouds, used stems to group similar words. For example, "empathetic," "empathy," "empathize" and "empathizer," "empathizes" and "empathizes" would all be summarized by the stem, empath-". However, in going back over the manuscript, I cannot find any reference to "word stems" – even doing a search through the document for "stem" does not produce any hits for this (only for "systems"). I can accept the explanation that grouping similar words according to stems is "common" in word clouds, but that is not something I was aware of. It is certainly possible that participants may not be aware of this either; I was trying to imagine looking at each word cloud as a participant would, and I felt distracted by the oddness of reading the word stems and wondered if it had just been poorly formatted. But, I will accept the authors' argument that this is "uncontroversial" given their expertise and prior use of word clouds. However, for the manuscript, given that NHB has a broad audience, some of whom may also not be aware that it is common for word clouds to use word stems, it would be helpful if the manuscript explained that the word cloud uses stems instead of (entirely) whole words.

We have added further explanation in the Design section (p. 8) of what stemming does, how this method enhances interpretability (by reducing redundancy) and what other works use it (including citations to relevant examples). We have also reinforced these points in "The Development of our Peer Praise Intervention" section of the SI.

4 R4

Research on psychological intergroup interventions is gaining momentum and, in my view, constitutes one of the most relevant and engaging research areas within the social sciences. This is one reason why I find this work both valuable and well-aligned with current trends in the field. However, I did wonder why the authors chose not to situate their work within the broader context of psychological intergroup interventions (Halperin, Hameiri and Littman, 2024), focusing solely on empathy-related interventions. Given the research question and

primary outcomes, I would recommend placing this work within the larger field of intergroup interventions to enhance its relevance and applicability.

We have taken R4's helpful advice to incorporate the work of Halperin, Hameiri and Littman (2024) into our Introduction in several places.

Related to the above, recent review on intergroup interventions, Čehajić-Clancy and Halperin (2024) has identified key steps and existing limitations for advancing research on intergroup interventions. Some of those limitations (such as addressing durability of intervention effectiveness and identifying mechanisms) are directly addressed in this work. This is good, timely and necessary. I encourage the authors to emphasize these contributions in light of the recent review on advancing research on psychological intergroup interventions.

We thank R4 for the recommendation; we agree that the discussion in Čehajić-Clancy and Halperin (2024) helps to motivate our work and have incorporated it into the front-end of our paper in several places.

The main argument in this work is that peer-praise might be a relevant psychological approach to changing/influencing intergroup outcomes (e.g., biases). I wondered why the authors have not provided a more comprehensive theoretical review on social learning and the mechanisms through which learning from others impact intergroup (e.g., racial) attitudes. Work by Olsson and colleagues on social learning and racial attitudes might be relevant here and it will strengthen and provide theoretical framework for authors' argumentation on how/why peer-praise might influence intergroup outcomes.

While we do not necessarily engage directly with the core of social learning research, since its key prerequisite is "the presence of at least two individuals" and the most common method of analysis is the neuroscientific measurement of "brain-to-brain coupling (BtBC)" (Pan, Novembre and Olsson, 2022, 682), we have added a brief discussion of how this line of research dovetails with and reinforces our focus on peers in the introduction.

My main concern relates to the selection and proposed mechanisms. I think that the authors should motivate these choices more rigorously. First, they should provide theoretical reasoning for selecting positive emotions and social norms as the predicted mechanisms. Second, they should review existing research that directly supports these choices. For example, recent work on the role of positive emotions (more general) and moral elevation (more specifically) has identified the relevance of positive emotions for improving intergroup relations in contexts of contested and conflictual intergroup relations (see the recent JPSP issue psycnet.apa.org/record/2025-15509-001?doi=1). Related to this, I wondered why the authors did not consider social learning mechanisms as possible explanations for the intervention's effectiveness. This could, at least, be discussed.

In response to the helpful suggestion from R4, we have expanded and revised our three long paragraphs on our proposed mechanisms (on page 3 of the manuscript), including adding citations as suggested. Regarding the last point, on social learning, please see answer directly above (it is now discussed).

Finally, a minor issue relates to the problem statement behind this work. It is argued that outgroup biases (racial/ethnic) in the US context are on the rise, particularly as expressed by the dominant group. I would strongly recommend supporting this claim with more concrete evidence to better justify the problem choice. Additionally, it would be helpful to clarify why the study focused exclusively on the dominant group and not on the other racial/ethnic groups in the US context.

To clarify, we do not claim anywhere in the paper that outgroup biases are “on the rise,” only that they are “among the most well documented and widely observed phenomenon in the social sciences,” and that racial biases towards Black and Latino/a citizens held by the dominant group have deep historical and institutional roots and are reflected in a number of policy debates. While our first paragraph already contains a number of citations to sources documenting these biases, we have revised the text to slightly expand this section.

References

- Čehajić-Clancy, Sabina and Eran Halperin. 2024. “Advancing research and practice of psychological intergroup interventions.” *Nature Reviews Psychology* 3(9):574–588.
- Halperin, Eran, Boaz Hameiri and Rebecca Littman. 2024. *Psychological Intergroup Interventions: Evidence-based Approaches to Improve Intergroup Relations*. Routledge.
- Pan, Yafeng, Giacomo Novembre and Andreas Olsson. 2022. “The interpersonal neuroscience of social learning.” *Perspectives on Psychological Science* 17(3):680–695.

Point-by-point response

We thank the Editor and Reviewers for their careful evaluation of the Stage 2 submission and for the constructive suggestions provided. We appreciate the reviewers' overall assessment that the registered report is a strong contribution, and we believe that addressing the points raised has further strengthened the manuscript. We provide detailed, point-by-point responses to each comment below.

Editor

*"Please attend to **every item** in the checklist and upload a copy of the completed checklist with your submission."*

Please find attached in our submission the completed checklist.

"Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't."

We have included in the cover letter the opt-into transparent peer review language.

"In the Discussion section, please include a discussion of the limitations of your work."

We've included a discussion of limitations of our work in the manuscript.

"Please supply a brief (maximum 250 characters, including spaces) summary of the main findings of the paper to be used on our website and in our e-alerts. The summary should be written in the third person in language suitable for a broad audience. The summary may be edited by the editors prior to publication. Please provide this summary in your cover letter."

We have included a summary (<=250 characters) in the cover letter.

"Tables must be editable and prepared using the table menu in Word, the table environment of LaTeX or the table functionality of ChemDraw where relevant."

Our tables are in the table environment of LaTeX.

Reviewer 2

"am pleased to see how this project evolved over time, as well as the final outcomes from the authors' extraordinary work. I want to begin by commending the authors on conducting a rigorous set of studies, reporting their findings as transparently as they did, testing for robustness and heterogeneity of effects, and explaining what all of it means very clearly. This paper is an example of social science at its best."

We are very excited R2 finds the work of high caliber, and are grateful for their comments throughout.

*"I only have one remaining comment for the authors, which they can choose to ignore if they wish; it is something I would like to see in the discussion, but it does not *have to be* included, as it is a broader issue beyond just this paper. Despite the largely null effects they found, the authors continue to (implicitly) tout the value of "low-cost, scalable interventions" for encouraging empathy toward racial and ethnic outgroups. I wonder whether the authors think this is still the path the field should be taking, given their own null results. While they critique the deep-canvassing studies for their high cost and thus inherent lack of scalability, at the end of the day, those high-touch interventions produce very long-lasting effects. There seems to be an inherent tradeoff here: the light-touch studies are easier to do, but they don't change very much or last very long; the high-touch studies are expensive, but they do produce lasting and meaningful changes. One of the big debates in the behavior change literature at this moment is how much of our time we should be spending on light-touch vs high-touch approaches, given their relative benefits vs. costs (for review, see Chater & Loewenstein, 2023). Given the findings of this manuscript, I am curious about the authors' thoughts on this larger issue."*

We are grateful for R2 for spurring us to consider these excellent points on the relative merits of different types of interventions and the implications that our findings have for this broader debate in the field. As we see it, there are two issues: the more narrow issue of "high cost" vs. "light touch" interventions and the broader issue of "i" (individual) vs. "s" (systemic) interventions. We now discuss these tradeoffs in several places throughout the Discussion, notably in the 4th, 5th and 9th paragraphs of that section.

Reviewer 3

"Thank you for submitting your Stage 2 manuscript -- I have been looking forward to seeing the results of the study. In line with reviewing a Stage 2 manuscript, I would first like to confirm that my assessment is that the authors have followed the approved study plan. In terms of the Results, these are clearly reported and structured well for the reader. There are a number of analyses, but I did not feel lost at any point about what was being tested and communicated, so I commend the authors for writing a complicated Results section so clearly."

We are grateful for the comments R3 has offered throughout, and are glad to receive confirmation that R3 assesses the work as having followed the approved study plan and that the results are clearly reported for the reader.

"Where I would like to see some changes is with the Discussion, which seems very brief. A good portion of it is spent reminding the reader about what was done and what the strengths of the study were, which is fine, but then the portion spent explaining the results and discussing their implications was limited (especially given how much the Introduction covers). I would strongly encourage the authors to delve into explaining the results more and contextualising them more within relevant theory in order for the manuscript to reach its full potential. For example, while I'm aware that the authors weren't testing an increase in empathy as a potential mechanism for changing political behaviours, I was really left wondering about what may have driven the change in political behaviour observed, if there was no simultaneous change in empathy. Why would participants bother to donate or write letters, if there was no meaningful change in empathy? Are there theoretical, or possibly even methodological explanations for this? Furthermore, if the intervention works best with those who already have relatively high levels of empathy towards the outgroups, then what should researchers in the area do next for intervening to change empathy and behaviours in those who have lower baseline levels of empathy?"

We thank the reviewer for pushing us to fully discuss the mechanisms behind our observed behavioral effects. In the revised Discussion, we highlight that peer praise increased low-cost symbolic political behavior (letter writing) while leaving attitudes and more costly behaviors (donations) unchanged. One possibility might be moral licensing – where engaging in positive behavior gives respondents license to avoid moral effort – but this framework would predict negative effects on subsequent prosocial behavior, which we do not observe. Instead, our exploratory analyses suggest that affective mechanisms, primarily anxiety and anger, may have motivated participants to engage in low cost behaviors, consistent with cognitive dissonance and self-regulation frameworks. We incorporated a suggestive discussion of these mechanisms in the revised manuscript and noted that future work could explore whether such affect driven behavioral changes eventually lead to shifts in underlying attitudes.

“As an aside, the finding that the intervention seemed more effective for those with a higher baseline level of empathy made me wonder if this could be explained by the Matthew effect, so I was a little surprised that the authors do not mention it in any way.)”

Thank you for this suggestion in framing this result; we have accordingly incorporated language in the Discussion that brings in discussion of the Matthew effect, and other possible explanations: “This is consistent with prior work, for instance, that find perspective-taking nudges work better among higher-than-average empathy respondents \cite{adida2018perspective}, and shares observational similarities to a Matthew effect at play \cite{merton1968matthew} and/or elements of learning/habit-forming \cite{shamay2022adaptive}.”

“In the abstract, I think the phrasing around “can change attitudes” is a little vague, even for an abstract. Different phrasing, or perhaps a little more detail, would strengthen the summary of the study here.”

We added specific details to the indices of attitudinal and political inclusion in the abstract so that readers can quickly retain a sense of what measures are used for each index: “We measure costly choices to engage in empathy, test whether peer praise promotes political (an index of letter writing and donations) and attitudinal (an index of social distance and thermometer ratings) inclusion, and, in a separate experiment (N=4,404), test if praise specifically from co-partisans can also promote inclusion.”

“I noticed a typo in the Introduction: the word “beliefs” is repeated, this occurs in the second full paragraph on page 3, towards the end of the third line.”

We have edited out this typo.

Reviewer 4

“The authors have succeeded in addressing my earlier suggestions and comments and (I believe) that this has strengthen the overall quality of the current MS. While I do recommend the current version for publication, and after having read the new revised version in its entirety, I still believe that two issues remain to addressed (elaborated”

“The introduction is still missing an argumentative explanation for the intervention selection. While the authors did argue why empathy is relevant for reducing outgroup bias (e.g., prejudice), there still remains a question why this particular intervention was selected and not other established (and even more effective) strategies (such as norms messaging, contact etc.). A more systematic rationale for selecting empathy and why it was expected to be the most appropriate intervention would strengthen the introduction part of this MS.”

We thank the reviewer for raising this point. As noted by the editor, we have not revised the Introduction, which was approved in its current form. However, the Discussion now clarifies how our intervention differs from prior work and why it represents a theoretically and practically meaningful approach. Specifically, we emphasize that the peer-praise-for-empathy intervention was designed to be light-touch, low-cost, and scalable, in contrast to many established interventions that require substantial time, resources, or repeated contact. We also highlight that our intervention targets a particularly challenging domain—encouraging empathy toward racial and ethnic outgroups—and focuses on promoting an “approach to empathy” rather than assigning empathic behavior. In addition, the Discussion explicitly notes that our design does not allow us to compare the efficacy of peer praise for empathy to more intensive interventions, and we treat this as an important limitation.

“Second, the study did not systematically account for contextual and individual differences that might moderate intervention effects. Even though the authors do acknowledge this throughout the MS, I believe a more thorough discussion regarding this limitation deserves to be included in the discussion section. Variability in participants’ contextual background (where they live), or prior intergroup experiences could have influenced both receptivity to and outcomes of the intervention. Authors should, at least, argue that future research would benefit from explicitly and systematically aligning key (relevant) factors to better understand for whom and under what conditions empathy-based (but also all other intergroup) interventions are most effective.”

We thank the reviewer for this thoughtful suggestion. We agree that understanding contextual and individual factors that may moderate the effects we identified is important. We thus expanded our discussion of moderating effects (see section “How and Why the intervention Works, cont’d (exploratory)”). Specifically, in this section we systematically discuss heterogeneous treatment effects by key demographics (age, income, gender, education, and state), finding largely null moderation across most dimensions, scattered and non-systematic state differences, and one substantively meaningful moderator (education) in study 2.