Reviewer's report

Title: Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review

Version: 0 Date: 20 Nov 2018

Reviewer: Angela Ballantyne

Reviewer's report:

Thanks for the opportunity to review this paper. It is an important topic that will be of interest to a wide range of readers in different disciplines. As the authors point out, there is only emerging literature on this important topic and I think this paper makes a valuable contribution, both to summarising the existing research and making conceptual contributions to how we understand bias. I think that overall the paper is well argued, but the arguments could be clarified and strengthen in a number of ways.

1. Please include an explicit and clear definition of both implicit bias, implicit stereotype, and implicit prejudice. The use of these terms caused some confusion throughout the paper.
   * You define implicit bias solely in the negative "prejudicial for vulnerable groups"; but in my reading of the literature, implicit biases can be both positive and negative (are any preference that we have that we are not consciously aware of). Can you please provide explicit definitions of these critical terms?
   * P7L8 "implicit prejudice or implicit stereotype" these are not the same, can you comment on the difference? I understand a stereotype or bias can be positive (e.g. "older white men make authoritative leaders"). Prejudice can mean 'bias' or 'bigotry' and often has a negative connotation. So you really need to clarify for readers how you are using the term "implicit prejudice".
   * P7L52 You say you excluded "interventions undertaken with the aim of increasing prejudice." Can you describe what these interventions aimed to do? I am surprised that such interventions exist (given to me prejudice is negative) and wonder if this reflects something to do with how you are interpreting "prejudice" and if this could therefore be clarified.
   * P12L29 "The distinction between implicit prejudice and implicit stereotype, which is important in the psychological literature, was also often blurred so that stereotype was cited in the title when the method described using an IAT to test implicit prejudice." This comment is hard to follow because you have not yet clearly explained to the reader how 'stereotype' differs from 'bias' and differs from 'prejudice'.
   * P12L29 You say you excluded "interventions undertaken with the aim of increasing prejudice." Can you describe what these interventions aimed to do? I am surprised that such interventions exist (given to me prejudice is negative) and wonder if this reflects something to do with how you are interpreting "prejudice" and if this could therefore be clarified.
   * On P16 you provide more of an explanation of the relationship between prejudice and stereotype, but this should come earlier. And I think even here, you talk around the issue rather than providing an explicit definition of each term.
   * Could you please also define "out-group"

You do a commendable job of explaining concerns about using IAT to measure bias. I think you could make the general point about the basically difficulty in measuring anything that is "hidden". Most of
your studies measured IAT relative to control group, what do the producers of IAT say about this specifically (i.e. the ability of the IAT to measure relative levels of implicit bias in two different populations)?

P4L5 "The latter is doubly problematic because" I don't understand the relationship between this sentence and the previous sentence. What is the latter and how is it "doubly problematic"?

P4L7 What is egalitarian behaviour - is it the opposite of bias?

P5L37 You use the phrase "morally problematic psychological associations that professionals with integrity would wish to avoid." Here is feels like you are talking around the issues - can you please be more explicit?

P8 What is the difference between "screening" papers for eligibility and "rating" papers - this suggest two different systems of evaluation and it is not clear to the reader what these involved or the purpose of each?

P12L1 "The lack of statistical description suggests a risk of bias in the studies." I understand what you mean, but given you are talking about 'bias' throughout the paper in a slightly different context, it might be useful to clarify what sort of bias you are referring to here, or use a different term.

P12L4-5 "bias towards publishing positive results" again, I understood your point here but I think other readers may appreciate a description of this phenomenon and/or a reference.

P12L21 "variety of fields from which articles proceeded" it would be helpful to list some of these fields to give the reader an idea of the scope of disciplines working on this topic

P12 "Other important information was sometimes omitted from the abstract, such as that the study was performed on under-18s." it would be worth explaining to the reader why you excluded studies on minors in your review…maybe because some stereotypes like gender operate differently in adults versus children?

P13L52 "Finally, there is a risk of publication bias in this field: negative or neutral results of interventional studies may not be published, or to a much lesser extent than efficient interventions. However, on this count it is reassuring that many studies in our review did in fact publish their negative results." I found this comment confusing. It would perhaps be better to reframe this as "while there are well-establish general publication biases in favour of positive publications, we did not find this in our study."

P13L30 "high involvement of the participants where they linked themselves to their attitude also aided success" I did not understand this, can you explain more clearly what this involved?

P14L58 I found this hard to follow ". . . especially because of the difficulties caused by variations in terminology in the field. This variation is reflected in the fact our keywords in search criteria differed, above and beyond other differences in our research questions and inclusion criteria." Maybe you could reframe this as ". . . especially given that we used different search terms/keywords, research questions, and inclusion criteria."

P15 "Any counteractions, even if effective immediately, would then themselves be rapidly countered since participants remain part of their culture from which they receive constant inputs." This point is
important and well made.

P16 "If change is really to be produced, a commitment to more in-depth training is necessary." I think the point here is broader. In my view the structure of organisations and institutions needs to change, so that the external cultural "inputs" change - this can be achieved for example by affirmative action programs that seek to promote women and minorities to visible leadership positions. So it not just a question of training, or culture, but also of structures and policies (and how these influence culture). I would be interested in your views on this issue; and think you could make a valuable contribution to this emerging literature by commenting on these structural issues.

Given all the methodological errors and challenges the authors highlight in the papers reviewed (see Limitations section), I think a table of advice for future studies in this area would be really useful - this could just be dot points.

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

**Quality of written English**
Please indicate the quality of language in the manuscript:

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.