Author’s response to reviews

Title: Effects of Attentional Bias Modification on Residual Symptoms in depression. A Randomized Controlled Trial.

Authors:

Rune Jonassen (rune.jonassen@psykologi.uio.no)
Catherine J. Harmer (catherine.harmer@psych.ox.ac.uk)
Eva Hilland (eva.hilland@psykologi.uio.no)
Luigi A. Maglanoc (luigi.maglanoc@psykologi.uio.no)
Brage Kraft (brage.kraft@psykologi.uio.no)
Michael Browning (michael.browning@psych.ox.ac.uk)
Tore C. Stiles (tore.stiles@svt.ntnu.no)
Vegard Ø. Haaland (Vegard.Oksendal.Haaland@sshf.no)
Torkil Berge (Torkil.Berge@diakonsyk.no)
Nils Inge Landrø (n.i.landro@psykologi.uio.no)

Version: 1 Date: 02 Feb 2019

Author’s response to reviews:

Dear Editor

The following rebuttal letter address the points raised by the reviewers in the same order as they appear in the response letter, and describe exactly what amendments have been made to the manuscript text and where these can be viewed (section, line, page).

Technical Comments from the Editorial Office:

1. Trial registration information – Please state the registry name and the registry date of this study after your abstract section. Also, please declare if it is retrospectively or prospectively registered.

Done (Abstract, line 88, page 2).
2. **CONSORT Diagram** – Please upload this as supplementary file.

Done (added as supplementary file).

3. **List of Abbreviations** - Please list all abbreviations used in your manuscript under the heading "Abbreviations" after the conclusions section. If no abbreviations are used in the manuscript, please state "Not applicable" in this section.

Done (after conclusions, line 409-414, page 10),

4. **Figure Legends** - Please provide figure titles/legends under a separate heading of 'Figure Legends' after the References. If Figure titles/legends are within the main text of the manuscript, please move them.

Done (after References, line 584-592, page 14)

Reviewer 1.

1. It was unclear whether the 37 participants who were erroneously recruited despite meeting the exclusion criteria of having a current major depressive event were included in the data analyses reported in the manuscript. They should not have been included, but the final analysis presently reported in the Results section (on page 8) suggests that they might have been, as this analysis distinguishes between participants with and without major depressive disorder (MDD). It may be that this distinction refers to presence or absence of MDD subsequent to the intervention, but this is confusing. The final analysis is confusing for other reasons also, not the least of which is that the DV is not clearly specified, which leads me to suggest that it be excluded.

Response: The study used an ITT approach, meaning that we included everyone who was recruited, including those people who were recruited incorrectly (which is a protocol violation). After formal diagnostic evaluation, thirty-seven participants also (together with earlier MDEs) fulfilled the M.I.N.I criteria for a current MDD episode at baseline. After primary outcome ITT analysis, a sensitivity test was conducted by excluding participants in the high end of HRSD scores (Results, line 264-272, page 7). We agree with the reviewer that the final post hoc analysis that also looks into depression status (current versus earlier MDE according to M.I.N.I) may be confusing and has excluded this in accordance with the reviewers’ suggestion.

The recruitment procedure and rationale for enrollment have been made clearer in the revised version of the manuscript (Methods, line 167-169, page 4). The inclusion of current MDD patients increase the generalizability of the findings but is a deviation from the pre-registered protocol, and is therefore presented as a limitation (Discussion, line 390-391, page 9).
2. The description of the task could perhaps be tightened a little. Referring to the probes as appearing "behind" faces is misleading, and potentially confusing. A probe that appeared behind a face would not be visible. Instead, I suspect that the probes actually appeared in the same screen location where a face was previously exposed, and I suggest that this be conveyed to the reader. The use of the terms "valid" and "invalid", to describe probes that respectively appeared in the locations where faces displaying positive or negative expression had just been shown, also invites misunderstanding, given that probes appearing in these two locations do not differ in terms of their validity. Instead, it would be more accurate (and informative) to distinguish these two types of probe condition by referring directly to their location, rather than to their validity (i.e. Probe Location: locus of the more positive/less negative face vs locus of the more negative/less positive face).

Response: We have changed the “valid- invalid” distinction to locus relative to stimuli and see that this is more informative. The terms “more positive/less negative” and “more negative/less positive” are also used in the revised text (Methods, line 174-175, 183-190, page 4-5).

3. I suggest that the information provided in the "measurement of attentional bias" section be slightly amended, by removing the statement that these 96 trials were "the same trial types as used in the ABM procedure", given that the mix of trial types within this AB assessment task was very different from the mix of trial types presented within the ABM procedure. In this same section, the provided equation would be easier to understand if the labels employed were more fully informative. I suspect readers may struggle to understand that "probe down" actually means "probe in lower screen position" or that "positive up" actually means "face in upper position displayed more positive emotion". Also, it's probably not helpful to refer to AB as having been measured "by a single session of the placebo training task", down in the "study outcome" section, which challenges the reader to understand how a placebo training task can also be an AB assessment task. Potential confusion would be reduced by describing the "AB assessment task" in the "measurement of attentional bias" section, and subsequently referring to AB having been measured by the "AB assessment task" in later sections also.

Response: Under “Measurement of Attentional Bias” we have now written: “The AB assessment task was identical to a single session of the placebo-training task but used novel face stimuli. Attentional biases (AB) was calculated as the difference in reaction between trials in which the probe replaced the relatively more negative face vs. the more positive face ([(SUM(more positive face in upper screen position - locus of probe in lower screen position, more positive face in lower screen position, locus of probe upper screen position) - SUM(more positive face in upper screen position - locus of probe in upper screen position, more positive face in lower screen position- locus of probe in lower screen position))/2.). Thus a more positive score reflects a greater bias towards the more positive stimuli” (Methods, line 196-205, page 5). We subsequently refer to AB having being measured by the AB assessment task (Methods, line 220-221, page 5).
4. I recommend that the authors clearly indicate within the manuscript the precise timing of the AB assessment task delivered at baseline (i.e., how long before the first ABM/placebo session?) and of the final AB assessment (i.e., how long after the final ABM/placebo session?).

Response: The AB assessment task at baseline was conducted the day before training start for all participants. The mean AB assessment time after training (days) at follow up was 2.4 (3.2) and did not differ between ABM and placebo [F (1,301) = .434, p = .51]. ("Data reduction, compliance rates, and timing of the AB assessment task”, under Supplemental information, line 483-485, page 11).

5. No analysis presently is provided that informs the reader whether the AB measure was changed by exposure to the ABM condition (i.e. increased, as intended), and whether this change was greater in the ABM condition that was the case in the placebo control condition (also as intended). This analysis should be included. I note the abstract states that "ABM induced a change of AB towards relatively more positive stimuli”, despite the fact that the analysis needed to justify this claim is not reported. Hence I anticipate that this analysis will confirm that AB was differential changed, as intended, by exposure to the ABM and placebo conditions.

Response: We apologize for not putting in the main effect of training in AB.

A repeated measure ANOVA with AB (time) as the single factor did not reveal mean differences in change between ABM and placebo [F (1,309) = .084, p = .77]. We found that AB changes follow symptom changes across groups by a statistically significant association between AB (time) and HRSD (time). When we conducted a post hoc linear regression model within groups, we found that the association between ABM and HRSD was present only within the ABM group. We thereby show a beneficial effect of ABM despite no intervention dependent change in AB. This means either than ABM is working by some other mechanism (not by change in AB) or that the measurement (based on median RTs) of AB is not very robust (Discussion, line 314-318, page 8).

The effect of training in AB has been added to the manuscript as the first sentence under the section ”Symptom change after ABM and attentional biases”(Results, line 289-290, page 7). The statement in the abstract is changed to ”ABM induced a change of AB towards relatively more positive stimuli for participants that also showed greater symptom reduction” (Abstract, line 80, page 2), and “ The current study demonstrates that ABM produces early changes in blinded clinician-rated depressive symptoms (Abstract, line 83-85, page 2).

Reviewer 2.

1. Although this is a double-blind design, the experimental condition and control condition differ remarkably in percentage of the dot probe location, which is easily observable if
the clinician simply take the time to look at 5 or more trials. So how is double-blind really achieved? Did the patients perform this with clinicians nearby, or did they do their ABM training at home?

Response: As stated by the reviewer, the blinding procedure is important and several steps were taken to ensure double blinding. All ABM training was conducted at home after an independent lab technician prepared the training computers (see details under “Randomization and blinding”). We apologize for not making this explicit in the Methods section. The revised manuscript is now clear on that the training was conducted at home (Methods, line 191, page5). The training location is also found in Discussion, line 356, page 9.

2. Following my first question above, please describe the setting in which the patients completed the baseline task, training, and post-treatment re-test.

Response: The setting in which the AB assessment task was completed is now found in Methods, line 204-205, page 5. “The ABM assessment task was conducted in the same lab environment for all participants”. The ABM training was conducted at home (Methods, line 191, page 5).

3. On page 7 the authors state that "Means and standard deviations at baseline were 7.5 (3.9) for ABM and 7.2 (3.9) for placebo and changed to 7.3 (5.1) and 8.1 (5.4) at two weeks follow-up [F (1,272) = 4.48, η²=.02, p = .03] indicating that the observed changes in clinician-rated symptoms was not confounded by random variation at baseline." But these numbers do not match Figure 1?

Response: Here, we understand that the reviewer refers to the post-hoc sensitivity analysis. The sensitivity analysis does not refer to Figure 1. We understand that the reference to HRSD differences at baseline (Table 1) in the next section might be what causes this confusion and has removed the reference to “(Table 1)” from the section (Results, line 264-272, page 7).

4. Also, from Figure 1 it seems that the so-called treatment effect is solely coming from increased scores in the placebo group, rather than a decrease in the ABM group. This can be potentially concerning and need to be addressed.

Response: As the reviewer addresses, in contrast to general improvement in self-reported symptoms, increased depression was found within the placebo group for clinician rated depression. Clinical assessment might drive the participants attention towards depression symptoms and both positive- and negative assessment effects could therefore influence the change patterns in ABM and placebo. Previously studies have found control conditions in ABM to be beneficial so it is generally a very tight control. An assessment only group will help in distinguishing assessment effects- from ABM in future studies. Studies that represent the trajectory of symptoms in this high-risk group could also better explain symptom changes that may affect ABM and AB (ie. a lot of them will relapse). The manuscript addressed the potential
placebo effect, and we thank the reviewer for also reminding us on addressing the negative change in the clinician-rated placebo group (Discussion, line 341-346, page 8).

5. Can the authors elaborate on how AB was computed? Participants performed the placebo task at baseline, then the same placebo task again 2 weeks later? So the units (in milliseconds) on the X axis of Figure 3 is the improvement in reaction time?

Response: Reviewer 1. also addressed this issue and gave a concrete suggestion on how this might be rewritten. We also made the meaning “differences in RT in milliseconds” explicit in accordance with this reviewers comment (Methods, line 196-205, page 5).

6. Lastly, since the task is similar to the ones used by the Browning et al. (2012) study, I think it would be worthwhile if the authors can describe the novelty and importance of the current study beyond the Browning 2012 study (in Discussion, perhaps).

Response: As written in the Conclusions section: "Previous studies have reported a mixture of positive and negative findings (Conclusions, line 400-402, page 10 ) leading to a degree of scepticism about the impact of ABM on symptoms of depression. Our study is by far the largest randomized controlled clinical trial of ABM in a depressed population and gives a broader and more transparent picture of the true advantage of the intervention in this population” (Conclusions, line 402-404, page 10 ), and in the Background section: While some studies have reported an effect of ABM in depression a number of meta-analyses have suggested a small effect size, although definitive conclusions have been limited by small sample sizes and poor trial methodology employed in many studies (Background, line 124-127, page 3).

The Browning et al. study described an interesting effect in an initial small study. The purpose of this study was to assess the reliability of these results using a larger sample and preregistered design. While we also found a positive effect on symptoms there are important differences in both sample characteristics and results. The initial study was based on a small sample (n= 30 for the face condition used in the current study), and had less residual symptoms (HRSD 3.1 and 1.8 for the ABM and placebo respectively). The latter is important, as more symptoms will increase the potential for change. The study found an effect on self-reported symptoms after 4 weeks, and no effect at 2-weeks follow up. The novelty and importance of the current study is now further elaborated on in the revised manuscript (Discussion, line 350-353, page 8-9).

Reviewer 3.

1. My main point relates to the interpretation of the analysis relating change in AB in the ABM group to change in depressive symptoms. The authors conclude that change in AB towards positive stimuli is associated with a decrease in symptoms. However in the Fig 3 note on page 7 the authors state for AB that: "positive values = change towards more positive bias" and for HRSD that changes in HRSD = "two weeks follow up minus
baseline" so for HRSD this would mean a negative HRSD change score is = decrease in symptoms. But Figure 3 shows a positive association between change AB and change HRSD. So increase in positive AB is related to increase in symptoms from BL to FU? This seems inconsistent with the authors' interpretation of this finding. Please clarify.

Response: We thank the reviewer for noticing (and we apologize for) this writing error. HRSD = "two weeks follow up minus baseline" is now changed to HRSD = " baseline minus two up weeks follow ". The Figure 3. caption is changed to: “Relationship between changes in HRSD (baseline minus two weeks follow-up) and changes in AB (two weeks follow-up minus baseline). Positive values = changes towards more positive biases (AB change) and symptom improvement (HRSD change)”(line 590-592, page 14).

2. The intervention targeted residual depressive symptoms in patients with a previous episode of depression. Do the authors have information about the time since the last episode? Residual symptoms of a recent episode are probably more malleable than persistent residual symptoms from an episode many years or decades ago?

Response: We agree with the reviewers comment and have added this topic in the limitations section under “Residual depression was widely defined…” (Discussion, line 392-394, page 9).

3. The ABM condition had an 87% to 13% ration of positive/negative probe location - I was wondering on what grounds this was based and why the authors did not use 100% positive probe location trials? Also why were 13% of trials in the active treatment condition probe replaced by negative stimuli rather than neutral? Related to this, the control condition had a 50/50 ratio of positive/negative probe locations. Why did the authors choose for this control condition rather than e.g. 100% of trials probe replacing neutral stimuli? It seems that with the scheme that was used in the study both conditions contain some elements of the active ABM towards positive stimuli intervention.

Response: The rationale for not using 100% positive probe location trials in ABM was that the participants also should be blinded to the intervention. The paradigm combines faces that are relatively more or less positive/negative, further making it difficult for the participant to explicitly notice any systematic combinations of face stimuli. We decided not to tell the participants about the purpose of the training-, or ask the participants to guess their allocation to ABM or placebo after two weeks as the study includes several secondary longitudinal outcomes that might be confounded. The weighting of probe location (87%/13%) will also allow measures of AB during training (not among the predictors/outcomes in the current study). A clearer description of the ABM procedure (and purpose) and AB assessment is provided in the revised version of the manuscript (Methods, line 186-197, page 4-5).

4. The AB computation on page 5 is a bit hard to follow and could benefit from further explanation or a reference.
Response: See answer and revision from comment above. Reviewer 1. did also ask for this clarification (Methods, line 196-205, page 5).

5. The AB measurement used novel stimuli. Do the authors have any ratings of how positive the positive faces and how negative the negative faces were actually perceived?

Response: We chose to use an ABM task that had been used in previous research studies (e.g. Browning, 2010, Browning, 2012). The stimuli come from four different stimulus sets, and this combination of stimuli has never been cross-balanced or validated based on ratings of valence and arousal. However, in previous studies the task has shown to produce significant effects of ABM, and has also shown to produce neural effects off ABM (Browning et al, 2010; Li et al, 2015). We are sorry that we cannot report normative data for the ABM task. We have commented on this as a limitation (Discussion, line 395-397, page 9).

6. Means (SD) in text do not always match those reported in Table 1. Rounding errors?

Response: We thank the reviewer for making us aware of the rounding errors in Table 1. The means and SD now match those reported in the text.

7. In general the authors could be more specific in their description of the data/results e.g. Table 1 state what the numbers in the Table represent (M, SD, n, % etc.). What does "Medication (SSRI) represent? Number of participants currently receiving SSRI treatment? or percentage?

Response: We apologize for not being clearer on what the numbers in Table 1. represent and have added and explanations in the figure text and table (Results, line 249-256, page 6).

8. Primary outcomes: "… changed to 8.3 (5.9) and 8.8 (5.7) at two weeks follow up". To which conditions are these numbers referring to?

Response: Now changed to "Means and standard deviations at baseline were 9.2 (5.9) for ABM and 8.3 (5.0) for placebo and changed to 8.3 (5.9) for ABM and 8.8 (5.7) for placebo at two weeks follow-up" (Results, line 260-261, page 6-7).

9. Page 7 line 35-44. Change in AB was associated with change in HRSD in the ABM group. From the text the direction of this effect remains unclear. More change towards positive AB related to stronger decrease in symptoms on HRSD?

Response: From the Reviewers first comment: ". The Figure 3. caption is changed to: “Relationship between changes in HRSD (baseline minus two weeks follow-up) and changes in AB (two weeks follow-up minus baseline). Positive values = changes towards more positive
biases (AB change) and symptom improvement (HRSD change)”.

The text is also now clear on the meaning in that “A post hoc linear regression model showed that there was a statistically significant positive association between HRSD improvement and positive changes and AB within the ABM group (Results, line 295-267, page 7).

10. P9. Line 19. The authors state that the finding that ABM had an effect on clinician rated symptoms cannot be considered clinically significant? Why not? What would be a clinically significant effect?

Response: We see the lack of context and references for this statement and chose change the sentence to: “The small effect sizes could be considered clinically non significant in treatment trials”(Discussion, line 365, page 9). The literature on clinical significance and conventions linked to the different inventories is outside the scope of this work.

We want to thank the Reviewers and Editor for their time, comments and suggestions that we think further improved the quality of the paper and clarified important aspects of the study design, methods, results and interpretations of findings.

Kind regards

Dr. Rune Jonassen & Professor Nils Inge Landrø