Dear Dr. Hugo Najberg,

Thank you again for submitting your revised Stage 2 Registered Report to PCI RR. I think one remaining major issue now is what the ***specific*** research question of the current study is, and what ***specific*** conclusion we can draw based on the data.

Thank you for your clear argument about this last issue. We have revised the manuscript accordingly, with the changes listed below. Responses to the points raised are in blue in this document, with quotes from the main manuscript in orange. Updates within the manuscript are highlighted in blue.

I appreciate the clarification on the difference between the main research question and the primary hypothesis. However, I agree with reviewer Matthias Aulbach that the research question "Can food response training modify real-world consumption behavior?" is too vague, general, and broad, and therefore needs to be made more concrete and testable in any specific studies.

It is unfortunate that such a vague research question was adopted at Stage 1, and it seems that we now have a disagreement on what the *specific* research question of the current study is. Our (the two reviewers and me) interpretation is that the specific research question is to test the difference between the 100% and 50% contingency groups. I hope you will agree that this is indeed the specific research question raised in the first place. For instance, in the introduction, you wrote (page 4) that "The effect of the intervention was contrasted with a mechanistic control group only differing in the active 'ingredient' of the training: the cue-response mapping rules will be 100% in the experimental and 50% in the control group. This contrast allowed us to control for the confounding factors developed by food cue exposure and cognitive training." For this specific research question, there is a clear answer based on the data: there is no difference between the two groups.

In the response letter, you wrote that "our approach was based on the assumption that the 50% condition would have no effect. This would have allowed us to isolate the absolute effect of the intervention, and thus could answer the main research question". If I understood it correctly, you are now saying that the research question is actually "Can food response training modify real-world consumption behavior, compared to a control condition that has no effect?" However, this is incompatible with what you wrote above, where you said that the reason for using 50% contingency in the control group was to "control for the confounding factors developed by food cue exposure and cognitive training". So you were saying that the 50% control group might still show some effects due to "food cue exposure and cognitive training", rather than assuming that there would be no effect, and your design allowed you to control for these confounding effects. Second, if the assumption was indeed that the 50% condition would have no effect, this should have been made clear at Stage 1. Furthermore, there should have been a data analysis plan at Stage 1 to explicitly test this assumption. However, this was not the case. Lastly, if the main research question was to "to isolate the absolute effect of the intervention" compared to a control condition that has no effect, I guess there are more suitable control conditions for this purpose, such as using a training that does not involve food stimuli (as in the new proposed study), or a control group that does not receive any training. The 50% control condition was not used to "to isolate the absolute effect of the intervention", so it seems unfair to criticize this control condition, after observing the results, that it "can induce a non-negligible effect of training into the control condition" and therefore could not answer the main research question.

For all these reasons, I respectfully remain unconvinced that the main research question was actually to isolate the absolute effect of the intervention. The original question, "Can food response training modify real-world consumption behavior?", is rather vague and does not explicitly say anything about the absolute effect of the training. Saying that the main research question was actually about the absolute effect of the intervention, rather than the comparison between 100% and 50% groups, sounds dangerously like changing the research question after the results are known, which is exactly one of the biases registered reports aim to guard against.

To sum the long arguments up, your current reasoning seems to be that (1) the primary hypothesis clearly shows no difference between the 100% and 50% group, (2) however, the main research question was actually about the absolute effect of the intervention, and (3) the null results do not allow us to say anything about the main research question. However, as I have tried to argue above, I am not convinced that the research question was actually about the absolute effect (the wording of the question was rather vague, and it did not say anything about absolute effects). Instead, my advice, following what both reviewers have said in their previous comments, is to say (1) there is no difference between the 100% and 50% groups, and (2) explicitly conclude that the answer to the specific research question here is that the supposed active 'ingredient' of the training does not help resist the consumption of sugary drinks. (3) You may still go on and say in applied settings, it may still be interesting to see whether such a training has any absolute effects at all, for instance by comparing it to another control condition. However, it should be clear that the question on absolute effects is different from the specific research question (i.e., the 100% vs. 50% comparison) raised at Stage 1 here. The main conclusion from this research is still that there is no difference between the 100% and 50% condition. This may seem like subtle differences in how to frame the findings. However, I believe this is crucial, because your current reasoning sounds like changing the research question at Stage 2, which is strictly forbidden.

To address the recommender's concerns, we propose the following changes to the manuscript:

- The discussed section has been splinted into two: "The Choice of Comparator Group" and "The Effects of the Training Might Have Been the Same for Both Groups".
- In this first section, we stress that there is no difference between 100% and 50% mapping. It reads p.19: "Our results demonstrate that there is no difference between the 100% and 50% mapping consistency conditions, indicating that this difference in the 'dose' of the supposed active ingredient of the training does not lead to observable change in dieting behavior in this population".
- In this second section, we discuss a possible interpretation of the null effect through the lens of similar devaluation in both groups and conclude on the interest of looking at another non-food comparator group, p.19 and 20. We do not challenge the primary hypothesis in this section.
- The section "Individual Differences in Baseline Dieting Capacities Influence Training Outcomes" has been moved up to mirror the structure of the Abstract and for better coherency.
- The Abstract does not mention a main research question anymore, it now reads: "We propose to conduct another study that includes a control training focused on non-food, i.e. without any mapping with food cue".

- The Conclusions does not mention a main research question anymore. It now lists the two highlighted possible interpretations for the primary hypothesis null results, like in the abstract. It now reads: "Two possible interpretations for the null results of the primary hypothesis are: i) food response training might have induced the same effect in both experimental and control groups, and ii) the recruited population might not have been suited to observe a change in our measure of dieting behavior induced by response training. As it remains important to identify whether food response training can impact dieting behavior from an application-oriented perspective, we suggest conducting a similar study but with a control group focused on non-food items".

We hope that these changes satisfactorily address the recommender's comments.

Reviewer Matthias Aulbach reiterated an important point concerning the relationship between devaluation and successful days of dieting. The equivalent devaluation effect between the experimental and control group is now emphasized strongly in the discussion. However, if devaluation is not related to successful days of dieting, to what extent can the null finding on successful days of dieting be explained by equivalent devaluation in both groups? This issue needs to be addressed more carefully.

We now address the point of reviewer Matthias Aulbach in the manuscript. As said in our response to their comment, the reviewer is correct in pointing out that the absence of correlation between the devaluation and the number of days of diet reduces the strength of the argument based on our finding for equivalent devaluation. We rephrased the paragraph to reduce emphasis on the devaluation effect. We think that evaluating the direct link between devaluation and dieting behavior is not the same as evaluating the effect of training through devaluation. The training can affect dieting behavior through multiple mechanisms, and measuring devaluation helps us understand if the groups experienced the training differently, regardless of whether devaluation itself is associated with dieting behavior.

We have added the following sentence to the discussion, p.19: "While changes in explicit liking did not directly predict dieting success (H2: BF01 = 3.87), valuation remains a useful measure to understand how the training influenced both groups".

Lastly, concerning the proposal to conduct a follow-up study, as both reviewers pointed out, the new proposed study will address a different specific question. Regardless of what results you may get from the follow-up study, this will not change the specific conclusion we can draw based on the current data. As such, at least for now, I do not see why this follow-up study should be added as an incremental registration. The proposed study differs from the current study in so many aspects, including (1) the study design, (2) the population, (3) the type of trained items, (4) the behavioral outcome, and (5) most importantly, the specific question being addressed. For these reasons, I think it makes more sense to submit the follow-up study as an independent Stage 1 manuscript (if you plan to do this also as a registered report), rather than as an incremental registration added to the current Stage 2 manuscript. Independent from whether you eventually decide to go for a new or an incremental submission, I think the issues with the current Stage 2 paper remain, and will need to be carefully addressed first before considering any potential follow-up studies.

After acceptance of this Stage 2, we will attempt to design a second study to complement this paper, the main point being to draw another more applied conclusion using a new comparator. All the reviewers' comments are well noted and will be taken into account if such a second study is to be conducted.

Kind regards,

Zhang Chen

by ***Zhang Chen****, 28 Oct 2024 07:42*
Manuscript: **https://osf.io/jckxr?view_only=4934c0215f2943cfb42e019792a30b53**
version: 3

Review by Matthias Aulbach, 23 Oct 2024 11:14

Again, I think the authors have done a good job at addressing my comments. There are, however, still a few points where I am not entirely convinced.

Regarding the distinction between hypothesis and research question, I see the authors' point and agree that it is a good idea to transparently discuss this. In the light of the authors' proposal to run another study, I would, however, recommend caution against this very generally worded research question: I think with any intervention, we need to ask, "is the intervention effective *compared to what*?" (Think: "Does Aspirin reduce headaches compared to doing nothing/taking a placebo/taking Ibuprofen/drinking a glass of water?"). One strength of the current study was to be very specific about this and the new study will be very specific about it, too. That also means that the proposed new study will not be able to answer the more general question because it is a question without an answer per se. Of course, that does not mean that running that other study is a bad idea as it will provide more data relating to the research question.

Thank you for your comment and clear discussion about a potential second study. We will take these points into account if/when conducting a study 2.

I might have missed this, but I think one of my points has not been properly addressed. In my earlier comments I wrote "[...] if we assume that devaluation is the (only) mechanism of action that would drive behavioral differences between groups. However, the analyses on hypothesis 2 revealed that changes in liking did not relate to successful days of dieting, indicating that this is not the mechanism by which training would change behavior. This indicates that processes other than devaluation would be driving behavioral effects in both groups." I invite the authors to discuss this issue – devaluation and behavior did not correlate, so why place so much emphasis on devaluation in the interpretation of (null) effects?

We apologize for missing out on this relevant comment.

The reviewer is correct in pointing out that the absence of correlation between the devaluation and the number of days of diet reduces the strength of the argument related to the equivalent devaluation. We rephrased the paragraph to reduce emphasis on the devaluation effect. We think that evaluating the direct link between devaluation and dieting behavior is not the same as evaluating the effect of training through devaluation. The training can affect dieting behavior through multiple mechanisms, and measuring devaluation helps us understand if the groups experienced the training differently, regardless of whether devaluation itself is associated with dieting behavior.

We have added the following sentence to the discussion, p.19: "While changes in explicit liking did not directly predict dieting success (H2: BF01 = 3.87), valuation remains a useful measure to understand how the training influenced both groups".

Figure 5 and 6: I apologize for spotting this only now, but I think it would be worthwhile to switch the x- and y-axis. I'm aware that the authors computed correlations (which are "non-directional") but I would argue the implicit assumption is that changes in liking/days of training predict the days of successful dieting. Thus, the "dependent variable" successful days of dieting should be on the y-axis, as is common. Of course, the presented information remains the same, it might just be more intuitive to read.

We agree and now switched the x- and y-axis. Figure 5 and 6 have been modified in the manuscript.

Review by Pieter Van Dessel, 24 Oct 2024 11:08

The authors have done a good job in revising their paper, clarifying the distinction between their main research question and the primary hypothesis.

I also support their proposal to conduct an additional study as an incremental follow-up to the current research. This new study represents a logical extension as it is well-suited to examine the effects of food response training on consumption behavior.

That said, it is important to emphasize that the proposed follow-up study will not provide insights into the mechanisms underlying how the training produces its effects. The proposed design, by recruiting participants who have already shown difficulties with diet adherence and focusing on consumption frequency, may indeed show changes in behavior. However, as the authors themselves suggest, this does not necessarily speak to how or why these effects occur. It is entirely possible that any type of task involving target stimuli could lead to changes in consumption behavior, regardless of whether those tasks involve specific contingencies or cognitive processes. This is not inherently problematic if the authors do not intend to make mechanistic claims in the follow-up. The first study can then be noted as a study that allows more specific conclusions in that respect whereas the second study looks at the overall potential of the training in light of the study 1 results.

Thank you for the positive evaluation of our revision. We will definitely keep this comment in mind if/when conducting a second study. We indeed want to focus on an applied approach, not focusing on the underlying mechanisms of action.