



Exploring how feedback on memory accuracy shifts criteria

A recommendation by [Zoltan Dienes](#) ^{id} based on peer reviews by [Greg Neil](#) , [Dan Wright](#) ^{id} , [Romuald Polczyk](#) ^{id} and 1 anonymous reviewer of the STAGE 2 REPORT:

Yikang Zhang, Henry Otgaar, Robert A. Nash, Chunlin Li (2025) The effects of false feedback on state memory distrust toward commission and omission, and recognition memory errors. OSF, ver. 6, peer-reviewed and recommended by Peer Community in Registered Reports. <https://osf.io/z8mv5>

Submitted: 08 November 2024, Recommended: 27 March 2025

Cite this recommendation as:

Dienes, Z. (2025) Exploring how feedback on memory accuracy shifts criteria. *Peer Community in Registered Reports*, 100938. [10.24072/pci.rr.100938](https://doi.org/10.24072/pci.rr.100938)

Published: 27 March 2025

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We may not believe what our memory tells us: Memory may deliver a compelling recollection we believe did not happen (we know we were not there at the time); and we may know an event happened that we fail to remember. That is, there can be distrust in remembering and distrust in forgetting. Previous work by the authors has looked at this through a signal detection lens, reporting in separate studies that people who have distrust in remembering have either a high or low criterion for saying "old" (Zhang et al, 2023, 2024). A plausible explanation for these contrasting results is that the criterion can either be the means by which false memories are generated enabling the distrust (low criterion); or rather, in conditions where accuracy is at stake, the means for compensating for the distrust (high criterion). In the current study by Zhang et al (2025), participants were incentivised to be as accurate as possible, and in a memory test given feedback about commission errors or, in another group, omission errors. The manipulation check indicated that the feedback did not increase (by a meaningful amount) distrust in remembering or distrust in forgetting, respectively, compared to a no feedback control group. Nonetheless, the authors found that people adjusted the criterion to say "old" in a compensatory way in each group. The possible mechanisms underlying these criterion shifts are discussed by the authors, who grapple with the distinction between response criterion shifts versus genuine meta-memory belief changes, and for the latter case, whether any memory distrust change could be contextual versus global (the manipulation check measured the latter). The Stage 2 manuscript was evaluated over one round of in-depth review by four reviewers. Based on detailed responses to the reviewers' and recommender's comments, the recommender judged that the manuscript met the Stage 2 criteria for recommendation. **URL to the preregistered Stage 1 protocol:** <https://osf.io/x69qt> **Level of bias**

control achieved: Level 6. *No part of the data or evidence that was used to answer the research question was generated until after IPA.* **List of eligible PCI RR-friendly journals:**

- [Advances in Cognitive Psychology](#)
- [Collabra: Psychology](#)
- [Cortex](#)
- [Experimental Psychology](#)
- [Journal of Cognition](#)
- [Peer Community Journal](#)
- [PeerJ](#)
- [Psychology of Consciousness: Theory, Research, and Practice](#)

- [Royal Society Open Science](#)
- [Studia Psychologica](#)
- [Swiss Psychology Open](#)

References:

1. Zhang, Y., Qi, F., Otgaar, H., Nash, R. A., & Jelicic, M. (2023). A Tale of Two Distrusts: Memory Distrust towards Commission and Omission Errors in the Chinese Context. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1037/mac0000134>
2. Zhang, Y., Otgaar, H., Nash, R. A., & Rosar, L. (2024). Time and memory distrust shape the dynamics of recollection and belief-in-occurrence. *Memory*, 32, 484–501. <https://doi.org/10.1080/09658211.2024.2336166>
3. Zhang, Y., Otgaar, H., Nash, R. A., & Li, C. (2025). The effects of false feedback on state memory distrust toward commission and omission, and recognition memory errors [Stage 2]. Acceptance of Version 6 by Peer Community in Registered Reports. <https://osf.io/z8mv5>

Reviews

Evaluation round #1

DOI or URL of the preprint: <https://osf.io/fsxdy>

Version of the preprint: 3

Authors' reply, 27 February 2025

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Zoltan Dienes](#) , posted 18 January 2025, validated 19 January 2025

Revision invited

We have four reviews back, which indicate reviewers are largely happy you did what you said you would. There are some points for revision; here are my further points and comments:

1) Remove exploratory conclusions from abstract. I am also a little uneasy about the amount of space they take in the discussion. When analytical flexibility is unleashed, and more complex analyses are used, it may be a relief that things make sense again. But we should go largely with the conclusions that follow from the pre-registered analyses themselves. I know your writing is already concise, and you may not like this, but can you halve the space given to the exploratory analyses in the discussion?

2) p 14 "After excluding three participants who had participated in similar studies and 28 participants who questioned the validity of the performance feedback, the final sample consisted of 622 participants" State that these exclusions were preregistered.

3) The d' analyses were not pre-registered (as far as I can see) and should go in the exploratory analysis section.

4) p 32 "This result suggests the effect of feedback on old-new recognition judgments was through its effect on metacognitive judgments of belief." There is no power calculation for this analysis so this conclusion should be deleted.

5) Dan Wright suggests he should have asked for further manipulation testing if the manipulation check failed. IPA couldn't be given under these conditions because the Stage 1 should set in stone exactly what you will do next. You also did not say you would not do the remaining analyses if the manipulation check failed, so proceeding with the other analyses is right. However, you cannot interpret any effect of the manipulation as due to a manipulation of memory distrust. So you have done the right thing. He also asks about other analyses you did but did not report; as these are not part of the Stage 1, it does not matter whether or not you did them. His suggestion that your abstract more closely reflect the Stage 1 I agree with, including reporting the failed manipulation check (and as I mentioned above, not the exploratory analysis).

6) You don't need to do anything about the following, just a comment. Given what seemed what the clear conclusion would be before the data were in was that if the manipulation check failed according to your criteria, that is likely because the manipulation did not manipulate memory distrust, this is the most straightforward conclusion. The manipulation involved an easy to program deception. I sometimes wonder if participants believe anything a psychologist tells them given how often and freely we try to deceive them. I suspect we pay the price for this: They treat us as seriously as we treat them. That makes it hard to answer one of Wright's main points: How to make a better manipulation. (Is it possible to do without deception?)

best
Zoltan

Reviewed by [Dan Wright](#) , 12 November 2024

Note: I was a reviewer on the Stage 1 manuscript.

This is well written and the methods follow what I was expecting from the Stage 1 report.

I have two main comments. The first is of more importance.

1. The manipulations had little (if any) affect on people's responses to the manipulation check (or the memory distrust scale, but I'll focus here on the manipulation check). This was surprising, particularly for the manipulation check as those questions seem to be asking about the exact psychological constructs that the two non-control conditions are meant to affect. On lines about 411 of the Stage 1 manuscript it states:

Since the manipulation needs to reach a certain level of strength, only if the lower bound of the 80% CI on the effect size is above the minimal effect of interest (raw score difference of 1.6 with a SD of 2), will we consider the manipulation adequate. If the 80% CI is within the equivalence bounds [-1.6, 1.6], we will conclude that the manipulation did not reach an adequate strength.

A summary of the finding in the Stage 2 manuscript and the decision to still go ahead were:

We therefore concluded that the manipulation did not reach an adequate strength in manipulating state memory distrust. Despite the manipulation not producing the intended effects, we proceeded with the main analyses to examine potential effects that may still inform our research question.

I had to go back and re-read the relevant section of the Stage 1 manuscript. I had interpreted the phrase above as meaning if they found, according to the criteria they state above, that the manipulation had not worked, they would work on making a manipulation that, according to their criteria, did work. I understand that having the data they went onto further analyses, but it seems this is now different. Of course it may be that the manipulation checks and the memory scale do not work as they were expected to (e.g., people aren't aware of the influences of the manipulation), but that seems a question worth both addressing and trying to see if there is a manipulation check that does. The alternative is that the manipulation checks do work but the manipulation did not. If the manipulation had no (or little) effect, then the significant findings appear suspect. I went back to my original review to see if I raised issues with the manipulation check. I did:

"My concern at this point is whether the manipulation will have the desired effect on memory distrust that the authors believe. If I read their power analysis correctly they believe it will almost completely account for memory distrust because they base their analysis on the memory distrust to response bias effect, if the causal chain above how they believe that this works. This makes the manipulation check critical.

As such, details of what counts for the manipulation check working is important for this. "

While the authors did state what they meant by failing the manipulation check, they did not examine, presumably through further data collection, why this occurred. I had assumed they would, but as a Stage 1 reviewer, I guess I should have asked them explicitly what they would do or requested a smaller manipulation check study before okaying the Stage 1 report. I am relatively new (as I guess we all are!) to reviewing these. My apologies for not requesting these. I had assumed failing the manipulation check would have prompted actions other than just continue, and that is my fault making that assumption. The editor may have views about what failing a manipulation check means for the rest of these analyses.

On a minor note, in the stage 1 review they describe why it might be of interest to compare when the manipulation check was administered. In their code they list:

```
MC_Distrust_Commission <-
```

```
afex::aov_car(Manipulation_Check_1~ Condition*Manipulation_order + Error(participantId),data = dat_S2)
```

This should be reported. I did find myself when reading the results going back to the code. The code page has a lot more analyses than appear reported.

One other note. In the original footnote 3 (but not included in the new version for some reason) discusses some issues with the manipulation. Thus, the authors were aware that this could be an issue.

2). If it is assumed that the manipulation worked (and therefore there was some unknown problem with the manipulation check), then the findings work as expected. I have two points. First, many of the outcome measures are dependent. For example, the c and beta. Similarly, when mixed models are done (and they aren't all linear, they are all generalized linear) statistics like d' and C are just coefficients in the probit regression (and really close in logit regression) if care is taken in coding the variables. Thus, these do not provide independent evidence and I assume the editor would prefer p-values only for one of these.

The second point is that the author note that the scores on the first recognition test (before the manipulation) could be used to predict the scores during recognition 2, and then measuring the additional impact of the manipulations would likely be more powerful. I skimmed the code file, but this is a long file. Were these done? This seems a better use of additional predictors than, for example, predicting the old/new response from belief ratings and other variables.

In summary, the authors appear to have done what they said that they were going to do, just I had assumed (wrongly) what would occur if they failed their manipulation check. The topic they purport to be examining has changed. Their original title was:

The effects of memory distrust toward commission and omission on recollection-belief correspondence and memory errors

But since their manipulation did not seem to affect distrust as expected, the new title focuses on the manipulation. This seems a largely substantial shift between a Stage 1 and Stage 2 document, but I am new to these registered reports. Maybe this is common to change what the focus is on.

With that, and recognizing I am part to blame for not asking the authors to be explicit in what they would do if the manipulation check failed, I recommend acceptance with three caveats. First, I think it is important, assuming that this is accepted, that this aspect is made clear to readers at the start (e.g., in the abstract say the manipulation check failed, and that possible reasons for this are discussed). Second, I believe the paper would be much stronger if the authors address this issue (likely with new data) as part of this paper prior to publishing this manuscript. Third, given their original intent (agreed upon from the title in the Stage 1 manuscript) was exploring the causal chain: manipulation -> distrust -> memory errors, they should stress that this was not shown.

Reviewed by **Romuald Polczyk** , 06 December 2024

The impression of the great competence of the Authors which I gained after reading Stage 1, only deepened after reading Stage 2. I do not see any weak points in the analyses (in fact, in several places they went beyond my own analytical competence).

The manipulation check did not show the effectiveness of the manipulation. Despite this, the Authors made the difficult decision to continue the analyses. I agree with this decision. Nonsignificant p-values do not imply nonexistence of the effect. It is therefore possible that in reality the manipulation worked, but it was not possible to show this in the check. Still, the effects of the manipulation can be seen in the main analyses. After all, the method used for the manipulation check - individual sentences assessed on a Likert scale - is rather weak, and much weaker than the tests used in the main analyses. The Authors commented in the Discussion that it may be that "state memory distrust measure did not adequately capture the change in state memory distrust levels" (p. 31, l. 4-5; also p. 32, l. 20ff., and 'Limitations'). It may be added that a single question is a less accurately measure than a questionnaire. This does not mean that other interpretations of the lack of an effect in the manipulation check are invalid. But the explanation may also be the low reliability of the measurement in the manipulation check. Currently, the Authors mainly discuss the issue of the validity of the manipulation check questions. It may also be worth mentioning their possible low reliability. After all, the p-values were not far from the significance level.

Minor points

I suggest not reporting confidence intervals as '.00'. Actually, if it was indeed exactly zero, it would mean that the effect is significant. Perhaps increase the decimal places to three in such cases. Writing '< .01' would be problematic unfortunately, as this does not exclude the the lower bound is negative, which was not the case.

Reviewed by anonymous reviewer 1, 10 December 2024

[Download the review](#)

Reviewed by Greg Neil , 09 January 2025

As this is a stage 2 review, I have reviewed this manuscript according to the stage 2 criteria, and categorised my comments within each of the criteria requirements.

Summary:

This study looks at whether giving false feedback about commission or omission errors influences participants' tendency to respond either more, or less, liberally to a memory test. The intended manipulation of state mistrust did not appear to be effective, although changes in the criterion were detected across feedback conditions anyway. Overall, the paper is well constructed and well written, and I have no concerns relating to the stage 2 criteria.

1) Can the data test the hypothesis, by passing the outcome neutral criteria?

Yes, The authors had previously shown a very thorough approach to their planning, and their approach to the data has been carried out, with all criteria being accounted for. The approach to excluding participants was well justified, and sampling sizes were achieved.

2) Are the introduction, rationale and hypotheses the same as the stage 1 manuscript?

Yes, I could see no substantial changes from the stage 1 manuscript.

3) Were the registered study procedures adhered to?

Yes, the stage 1 study procedures were adhered to.

4) Where there any significant deviations from the analytical approach set out in stage 1?

The initial plan was followed, but additional exploratory analyses were conducted to investigate the failure of the manipulation to change experimental condition state mistrust ratings when compared to control state mistrust ratings. Personally I found the additional analyses to be well-justified and appropriate to investigating the potential mechanisms at work, so in terms of a stage 2 review, I have no changes to suggest.

5) Are the conclusions justified on the basis of the data collected and the analysis used?

I believe they are. I agree with the authors that, on the basis of their analyses, the issue here may be that the state mistrust ratings are either not measuring what the authors' wanted them to measure, or else the criterion shifts did not occur because of changes in explicit state mistrust. However, overall, I believe that this manuscript fulfills the stage 2 criteria.