

I'm Haiyang Jin and I sign my review.

Review of "The importance of consolidating perceptual experience and contextual knowledge in face recognition" (PCI-RR#313\_Stage1).

The authors proposed an interesting study to explore whether the video sequence (original vs. scrambled) will influence the performance of recognizing faces with varied contexts ("in show" vs. "out show") both immediately after watching the videos and after 4 weeks. The results will potentially bring new insights into how the context, manipulated via video sequences, affects recognition performance/familiarization.

The authors have included many experimental design and analysis details in the report. It would be better if the authors could clarify some parts further.

For the power analysis reported in P6, the authors need to clarify the employed power analysis is intended for which of the four hypotheses. Specifically, the employed effect size seemed to come from one simple effect (P14), while no simple effects were employed for testing any of Hypothesis 2-4. Also, when GPower was used, whether the employed effect size was submitted to the (one-sided or two-sided) t-test or ANOVA?

For the stimuli, there are obvious differences in luminance between "In Show" and "Out of show" images. I'm not sure if the authors could do anything about this, but these luminance differences could be the main source of the differences in performance (thus a potential confounding to the effects of within-person variability). Moreover, it is not clear how the foil faces differ from the target stimuli (e.g., in luminance), it would be helpful to additionally display some examples of foil faces in Figure 2.

For the planned analysis for Hypothesis 2, the proposed analysis is essentially the interaction between Test time (0hours vs. 4weeks) and Image condition (original vs. scrambled) with a particular direction. But the interaction itself is unlikely sufficient. For example, there is a possible case where the deduction in the scrambled condition is 0 while the deduction in the original condition is -0.5 (i.e., an increase due to certain reasons, e.g., measurement errors). Then the deduction will be larger in the scrambled relative to the original condition, which matches the proposed supporting results (Table 1). However, probably the authors should not claim the hypothesis was supported by the case above. Instead, in addition to the proposed analysis (the interaction with a particular direction), the authors may also need to clarify at least one of the simple effects (for more please see the example in Jin, 2022). A similar comment applies to Hypothesis 4 as well.

In addition to the list of exclusion criteria, the authors may also need to consider some other criteria. For example, since it is an online study, authors may also like to consider including the criteria to include/exclude participants with behavioral performance (e.g., exclude participants with too fast or too slow responses). Also, authors may also need to consider excluding anyone who is already familiar with the actors (both the actors for the 10 characters and potentially

also the foil faces).

Also, I'm not sure whether the "consolidation" in the introduction and in Hypothesis 3 and 4 have the same meaning. Specifically, "consolidation" in the introduction seems to refer to the enhancement of the memory of the identity/face potentially with some manipulations (e.g., watch the videos/images again) while "consolidation" in the hypothesis only refers to what will happen during the delay time in general and no explicit enhancement manipulation will be applied. Therefore, it remains elusive whether there is "consolidation" (or enhancement in this study). Probably it is more appropriate to argue the effect as "how much participants will forget after the delay of 4 weeks".

#### **Minor points:**

1. It would be helpful if authors could summarize the whole experimental design, which seems to be  $2 \times 2 \times 2$ ?
2. P.4. It is stated that "However, if recognition memory for faces is greater in the Original condition, then this would suggest the importance of contextual knowledge." Probably authors intended to argue that "if the contextual knowledge is important, the recognition memory for faces is greater in the Original condition."
3. P.5. Please clarify what is "a normal range" to be used for CFMT.
4. P.9. How many raters will be?
5. Table 1. "the null effect" does not seem to be applied appropriately. The authors do not seem to apply Bayesian methods or Equivalence tests to test the potential support for the null hypothesis. If the authors do not obtain significant results with the NHST they specified, it remain unclear whether the evidence is inconclusive or it supports the null hypothesis. Therefore, the authors cannot claim "A null result would suggest that..." or "A null result for In Show images...", etc., especially when they used a more stringent alpha (i.e., 0.02).

#### **Reference**

Jin, H. (2022). *Pre-specifying effects of interest: Necessary practice to clarify conclusion*

*credibility in individual research*. PsyArXiv. <https://doi.org/10.31234/osf.io/t5hv3>