**Peer Review and Communication History**

**MS Title**: Angry, Sad, or Scared? Within-valence Mapping of Emotion Words to Facial and Body Cues in 2- to 4-year-old Children

**Author Names**: Yang Wu, Hannah M. Matteson, Claire M. Baker, & Michael C. Frank

**Submitted:** Sep 29, 2022

**Editor First Decision**: Revise & Resubmit

Jan 15, 2023

Dear Dr. Wu,

I have now received a review of your manuscript, “Angry, sad, or scared? Within-valence mapping of emotion words to facial and body cues in 2- to 4-year-old children”, from a researcher with special expertise in emotion regulation and identification and development. I also independently read the manuscript before consulting these reviews. The reviewer and I had mostly positive reactions to your manuscript. It provides very useful data and the manuscript itself was very well written. I agree that your manuscript has important strengths and also that there are some issues that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

I apologize for how long the review process has taken for this paper. In short, after struggling a bit to secure reviewers, one reviewer asked for extensions but I ultimately lost confidence in their ability to provide a review. However, because the review and my own reading were largely positive, I felt comfortable inviting a revision for your manuscript.

The reviewer did an outstanding job in their review. In your resubmission, please include a document with a point-by-point response to the reviewer’s comments, outlining each change made in your manuscript or providing a suitable rebuttal. In particular, I think you should pay particular attention to clarifying a few methodological and analytic decisions (e.g., trails that could be skipped). Although I do not consider the “strength of contribution” to be a crucial criteria when making decisions on manuscripts, I would like you to temper your claims in the introduction a little bit more given that previous research has tested at least some of these processes. The reviewer recommends other ways to highlight the unique contribution, which you may or may not find useful (it’s up to you if you’d like to integrate those particular points). I also think it’s reasonable to discuss power a bit more (as the reviewer notes). I saw that there wasn’t much (or any?) justification of the sample sizes in the pre-registrations. Finally, providing some descriptive information on things mentioned in the pre-registrations of both studies would be worth doing (e.g., both studies’ pre-registrations say that you would examine mask use).

**When submitting a revision, please mark any changed text in the manuscript in red font**. I will likely send the manuscript back your way to highlight these changes if they are not submitted.

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision. Please see the instructions below for submitting your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

William Chopik

# Reviewer 1

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Thank you for the opportunity to read this manuscript. The present study examined how well 2-, 3-, and 4-year-old children could identify the image that corresponded with an emotion label in a 2-alternative forced choice design. Study 1 tasked children with identifying only faces, and Study 2 presented children with faces on matching body postures. Results revealed that children at all ages were significantly above chance in selecting the matching image. Further, older children performed better than younger children, children performed better on cross-valence than within-valence trials, and the addition of body postures did not significantly change children’s accuracy.

Overall, there is a lot to like about the present manuscript. It addresses an interesting and important research question, as the role of emotion words in early development has received increased attention in recent years and may impact theories of emotional development. The manuscript is also very clear and well-written. The fact that the study was pre-registered is an additional strength. That said, I have several concerns about the work, which I have described in detail below:

Major Concerns

1. In the introduction and discussion, the authors emphasize the fact that previous work has not looked at whether young children (2-year-olds, specifically) can identify within-valence emotions in a comprehension task, and that the present study is unique in this respect. Although the study designs are not identical, both Russell & Widen (2002) and Price, Ogren, and Sandhofer (2022) have previously addressed these broader points. In Study 1 and Study 3 of Russell & Widen (2002), 2-year-olds were tasked with putting faces into appropriate boxes. In Study 1, when children were sorting into an “angry” box, “sad”, “fearful”, and “disgusted” faces were all included in the study design. Thus, in this comprehension task (i.e., which faces are ”angry”), children did have to make decisions about other within-valence emotions without requiring verbal responses. Similarly, 2-year-olds in Study 3 were tasked with deciding whether angry, sad, and fearful faces belong in an “angry” box or not. In recent work by Price et al. (2022), 2-year-olds identified whether faces belonged in “happy”, “sad”, or “angry” categories in Study 1, versus “surprised”, “disgusted”, and “afraid” categories in Study 2. Thus, both studies involved two within-valence negative emotions, and both studies found that 2-year-olds identified the emotions at above-chance levels. Given what has already been demonstrated by these two manuscripts, I believe the scope of the unique contribution of the present manuscript is a bit overstated both in the introduction and in the discussion (e.g., page 15 “Our study goes beyond prior work, however, in finding a much earlier ability to differentiate same-valence emotions”). That said, I believe that 1) additional research examining this topic is still extremely important but 2) the unique contribution and value of the present study could be more clearly specified. I would encourage the authors to reframe the introduction and discussion slightly in terms of the specific, unique contributions of the present study.
2. Related to the above point, emotion cue integration (knowing the word associated with body posture + face vs just face) seems to be a key unique research question in the present design. Perhaps emphasizing this unique contribution a bit more would be helpful. Additionally, I perceive the use of the CAFÉ dataset as a strength. Here, children are tasked with identifying the faces of other children, rather than adults (which have been most commonly used in developmental work previously). It may make sense to emphasize this contribution a bit more within the manuscript.
3. When describing the specific stimuli selected from the CAFÉ dataset, the authors state that to be included in the present study, “the face was rated among the highest for displaying a target emotion”. Can this be elaborated upon? What was the specific inclusion criterion here? E.g., Did the face need to be in the top 10 highest rated faces for that category? Above a certain % recognition threshold? Additional information about how the specific stimuli were selected would be helpful.
4. On page 9 the authors state that “If the child did not know the answer, the parent could skip the trial by directly clicking the next button”. How were these trials then coded? How many trials were skipped in this manner? If these trials were omitted from analyses, this seems particularly problematic, as children could skip trials they did not know and only respond on trials in which they were more confident, which would of course bias their own final score toward higher overall accuracy. If these trials were coded as incorrect, this may also introduce some bias because then children who are more willing to take a guess have at least a 50/50 chance of responding correctly, which means that children who said they “didn’t know”/skipped more often would have lower overall scores as a function of guessing less. Thus, additional information about how skipped trials were coded/handled for analyses would be important to include, as well as the overall number of skipped trials.
5. Regarding parental interference- The online design with parents clicking to respond on behalf of their child does appear to introduce opportunity for bias, although I appreciate the authors including information about efforts made to reduce bias. The authors report that “we excluded trials that had parental or sibling interference”, however it was not clear how this was determined. Can the authors describe the criteria for excluding trials based on interference, as well as the number of trials excluded for interference? Further, given that blind coders did not always corroborate parent clicks, were trials excluded when the coder disagreed with the parent click? Why didn’t the authors choose to use the blind coder’s decisions as opposed to parent clicks for analyses? The blind coder data may be more objective.
6. When looking at the adult validation of the face stimuli, why were adults so much worse at identifying the “scared” faces relative to the other emotion categories? Elaborating on this may be helpful.
7. Related to the above point, it appears as though the adult rating of the sad stimuli got worse when body postures were added as opposed to when the faces were presented alone (rating dropped from .92 to .87). It seems very surprising to me that providing additional, congruent emotional information would cause the accuracy rating to drop for adults. Can the authors include an interpretation of why this may be the case? Is this indicative of the sad body postures not being reliably identified as such? Speculation as to why this pattern was observed may be informative.

Minor Comments

1. I can see that the specific sample size was preregistered, and I greatly appreciate this practice. That said, 16 participants per age group seems like a relatively small sample. Can the authors include additional justification for the selected sample size?
2. On a very minor note, on page 11 near the bottom of the page, I believe this should be “child-friendly” rather than “child-friend”.

**Author Response**
Apr 7, 2023

See file “1992429-cover-letter.pdf”

**Editor Final Decision:** Accept

Apr 10, 2023

Dear Dr. Wu,

I have now had a chance to read over your manuscript “Angry, sad, or scared? Within-valence mapping of emotion words to facial and body cues in 2- to 4-year-old children”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
William Chopik