August 4, 2021

Dr. Bennett L. Schwartz, PhD

Associate Editor

*Metacognition & Learning*

Dear Dr. Schwartz,

Our revision of META-D-21-00035 “Perceptually Fluent Features of Study Words do not Inflate Judgments of Learning: Evidence from Font Size, Highlights, and Sans Forgetica Font Type” has been resubmitted for your consideration. We appreciate the thorough examination provided by yourself and our reviewers and are encouraged that our submission was viewed as “investigating an important issue” and that it would be “of interest to readers of *Metacognition & Learning*.” Below, we list each reviewer’s comments and provide our responses. In sections where specific revisions have been made, we cite page numbers. Finally, primary modifications to the manuscript have been made in blue-colored font to facilitate review. We look forward to hearing from you regarding our revision and hope that it is now suitable for publication in *Metacognition & Learning*.

Sincerely,

Nicholas P. Maxwell, M.S.

School of Psychology

The University of Southern Mississippi

Email: mark.huff@usm.edu

Phone: (601) 266-5411

Fax: (601) 266-5580

Cc:

Trevor Perry,

The University of Southern Mississippi

Mark Huff, PhD

The University of Southern Mississippi

**Action Editor**

**Comment #1:** I strongly resonate with Reviewer 2's point 4. You do show a font-size effect.  JOLs are higher when the font is bigger.  That replicates most other studies.  What you find that differs from the work of Rhodes and most other authors is that these JOLs are justified, that is, recall is better for larger font sizes than it is for smaller font sizes.  I really think that this point must be made more clear in the revision.  On a smaller note (no pun intended), I also agree with Reviewer 2's point 6 - a figure with the  Sans Forgetica font would be most helpful.   I also agree with Reviewer 1's point about how you monitored compliance with online data collection.  Also in keeping with Reviewer 1's last point, I think more demographic information is needed on your online participants.

***Response:*** Thank you for taking the time to provide additional feedback on our manuscript. We have edited the language throughout the manuscript to reflect the observation that while our data patterns did not exactly replicate the font-size effect as originally reported by Rhodes and Castel (2008), we still show an effect of font-size on JOLs such that large font increases the magnitude of JOLs while also providing an increase to cued-recall performance. These changes are most noticeable in the discussions for Experiment 1 (page 17) and in the General Discussion (see the Experiment 1A and 2A summaries on pages 28 and 29). Furthermore, we carefully note on pages 31 and 32 that while the standard pattern did not replicate, we did indeed find that increasing font sized increased JOLs.

Regarding the Sans Forgetica example, we now include a figure (Figure 5) which depicts three of the unrelated word pairs used in Experiment 3 in both Sans Forgetica and Arial. Finally, we have updated the participant sections of each experiment to include information on compliance checks and the recruitment criteria for Prolific participants (all participants were required to be native English speakers and have completed at least a high school diploma or equivalent.  
  
**Comment #2:** Following from Reviewer 2's point 4, one of the aspects of this paper that I found confusing was the seeming conflating of factors that change the magnitude of JOLs and the factors that change the accuracy (calibration, in this paper) of the JOLs.  Thus, I think in your Experiment 1A and 2A, you show an increase in the magnitude of JOLs for larger fonts, but not the concurrent decrease in calibration implied by the Rhodes studies.  Similarly, in Experiment 3, you find a decrease in JOLs for the special font, but it is not accompanied by any change in calibration.   When you do these changes, please also distinguish what you mean by an illusion of competence and how that indicates increase differences between mean value of the JOLs for any condition and the level of recall in that condition.   Therefore, in your revision, please make it clear when you are discussing the magnitude of JOLs, when you are discussing the calibration of JOLs, and perhaps also include some  
resolution measures.   I would be interested in whether your variables affect resolution (perhaps as measured by gamma) if you can do those analyses.

***Response*:** We have made a point of clarifying the language throughout each results section to make clear when we are discussing the illusion of competence (i.e., calibration) and when are discussing changes in the magnitude of JOLs and recall due to the perceptual manipulations. Additionally, we now include a table in the Appendix (Table A4) which contains gamma values as a function of perceptual manipulation (font-size, highlight, or control) and pair direction/relatedness (forward, backward, symmetrical or unrelated) in Experiment 1 and as a function of perceptual manipulation in Experiments 2 and 3.

**Comment #3:** I think I would also like a bit more review of the history of the development of the Sans Forgetica font  - and why the initial creators thought that it would boost memory performance.  And also please describe the data that shows it doesn't work in a bit more detail.  Can you recommend changes that would make such a dysfluent font more helpful for new learning.

***Response:*** We agree that it would be beneficial to the reader for us to provide a brief history on the development of Sans Forgetica font. Page 23 now contains a passage discussing the creation of this font as well as providing some overview of a study conducted by the Sans Forgetica research team when developing this font and their rationale for selecting the font that would be branded Sans Forgetica. Additionally, the following paragraph on page 23 provides more detail on the studies conducted by Geller et al. (2020) and Taylor et al. 2020 which showed Sans Forgetica did not provide a memory boost as claimed.  
  
**Comment #4:** You might consider including some discussion of the work of Jemstedt et al (2018) on ease or learning judgments.  We were able to show both fluency and belief effects on ease of learning judgment – while keeping memory effects to a minimum.  Take a look at our paper and then decide if it is relevant to the development of your ideas. It might also make sense to return to the work of Undorf et al (2017) in your discussion and perhaps discuss her ideas about when different cues are used in making judgments, particularly in the context of the interesting results you have not only for the font-size manipulation but also the forward-backward association manipulation.

***Response:*** We now include a paragraph in the Introduction (page 7) briefly discussing findings by Jemstedt et al. (2020) and use this section to discuss the relationship between fluency and beliefs.

**Reviewer 1**  
  
**Comment #1:** On page 3 you refer to "the judgment of learning (JOL) task" but you should instead call the JOL a measure rather than a task.

***Response:*** We have corrected the sentence on page 3 and now refer to JOLs as a “measure.” Additionally, we have updated the language throughout the manuscript to be consistent with this.

**Comment #2:** On page 5 you note that "items classified as more perceptually fluent are processed more quickly." This is circular reasoning.  The faster processing is how these stimulus types are determined to be more perceptually fluent in most studies, so the sentence feels a bit nonsensical as written.

***Response*:** Thank you for bringing up this point. We have opted to remove the sentence in question.

**Comment #3:** For each experiment, you need to include some rationale for the sample size tested. I understand that you included Bayesian tests to help if there was no a priori power analysis conducted, but it still helps the Methods of the paper to justify the sample size and stopping rule for data collection. I presume that there was no pre-registration of the study since it was not mentioned, but if there was that should be included.

***Response:*** Our sample size in all experiments was based on Rhodes and Castel (2008), who initially found the font-size effect using a sample of 20 participants. Given that our experiments were conducted online, we reasoned that we should double their original to account for any discrepancies in participant performance that might occur due to the lack of an experimenter being physically present. Therefore, we attempted to collect data for 80 participants (40 in the font-size group which doubled Rhodes and Castel and an additional 40 in the control group so that the groups could be matched on sample size). We have added a sentence on this to the Experiment 1 participants section (page 9), and again for Experiments 2 (page 19) and 3 (page 25). Finally, this study was not preregistered.

**Comment #4:** Given that your study used online data collection, you should also address the issue of compliance checks used in the study and how your participants fared with these. If the performance measures of interest were the only checks on compliance, state that explicitly.

***Response:*** In all experiments, recall performance was used as a compliance check. Participants were excluded if their recall performance fell below 5%, as this suggested that participants were not properly following task instructions. Language describing this compliance check have been added on page 9 (Experiment 1), page 19 (Experiment 2), and page 25 (Experiment 3).

**Reviewer 2**  
  
**Comment #1:** P.11, lines 28: Were the cued recall lists timed or did participants have unlimited retrieval time?

***Response:*** While participants were given unlimited time at retrieval, they were informed that they could press Enter and advance to the next pair if they could not successfully recall a pair. We have updated the language in the Experiment 1 procedure (page 12 ) to reflect the fact that recall time was unlimited.

**Comment #2:**  P. 13, lines 41-46:  Last sentence before Exp 1B is very difficult to parse.

***Response:*** We have edited this sentence for clarity. It now reads “Collectively, increasing font size increased both JOLs and recall percentages equally relative to small-font sizes. However, JOLs and recall for large-font sizes did not differ relative to a standard font size control group.”

**Comment #3:** P. 14, line 17: The authors mention fluency and distinctiveness being enhanced by bolding word pairs.  See though Price, McElroy, & Martin (2016) in Aging, Neuropsychology, and Cognition, 23, who found that participants gave lower JOLs to bolded items than to italic and regular font items and recalled bolded and italicized items at lower rates than the regular font items.

***Response:*** We appreciate you bringing this to our attention. We have updated this section (now on page 15) to reflect findings from Price et al. (2016)

**Comment #4:** P. 16, line 43: the authors say that there was no evidence for the font-size effect given that large fonts increased both JOLs and recall performance relative to small fonts. Again, on p. 18, line 46, the authors stress that font size did not affect JOLs and recall rates differently. The font size effect (FSE) is manifested if participants provide higher JOLs to large than small font items, with no differences in recall as a function of font size. Technically, the authors are correct that there was no FSE since recall was affected too.  However, this glosses over the fact that participants did tend to give higher JOLs to large font items than to small font items.  Given that these experiments are framed in terms of beliefs versus fluency, it would be ideal to not gloss over these differences in JOLs since they do provide insight as to the roles of each of these.

***Response:***This is an interesting point. While we initially interpreted our findings within the framework of Rhodes and Castel’s (2008) original font-size effect (where large font produced an increase to JOLs without a concomitant increase in recall), we agree that the observation that large font increased both JOLs and recall merits some discussion. We have edited the language throughout the manuscript to reflect this. Specifically, on pages 17 and 18, we now state that “Experiment 1A did not show evidence consistent with font-size effect as originally reported by Rhodes and Castel (2008), as the expected interaction was not observed. Instead, relative to small font, large font increased both JOLs and correct recall.” Furthermore, we discuss these differences in the General Discussion (pages 32-33) and note that while the standard font-size effect pattern did not occur, large font indeed produced an increase to JOLs.

**Comment #5:** P. 18, lines 38-43: The authors provide two means for comparison in multiple places, but it is unclear whether these are the means for JOLs vs. recall or large vs. small given the wording.

***Response:*** Means presented in parentheses are listed in the same order in which they are discussed in the sentence. We clarify this on page 13.

**Comment #6:** P. 21, line 19:  if this is the example of Sans Forgetica (SF), it doesn't convey what the font actually looks like.

***Response:***We originally typed a word pair using the Sans Forgetica font and were under the impression that it would display in the final .pdf file of the manuscript. We have removed the example on page 21 and now reference a figure (Figure 5), which provides three examples of the unrelated word pairs used in Experiment 3 typed in Sans Forgetica and positioned next to their Arial counterparts for comparison.

**Comment #7:** P. 21, line 36: the authors speculate that the disfluent font may cause participants to believe they will be more likely to recall items in SF than in Arial.  However, one reason that you typically see disconnect between participants' JOLs and memory performance with desirable difficulties is because they are more challenging to encode, which in turn yields lower JOLs, but higher memory performance. Thus, it seems unlikely that SF would yield a belief of better memory than Arial, if SF is in fact classified as a desirable difficulty.  All this to say that the logic of this manipulation providing insight about beliefs vs. fluency seems to be a stretch as framed.  It seems likely that the fluency and beliefs accounts would propose the exact same thing for JOLs: Arial > SF.

***Response:*** We have updated the predictions for Experiment 3 in consideration of this comment. On page 24, we now discuss the disconnect between JOLs and recall and predict that Sans Forgetica will result in lower JOLs relative to Arial pairs. To support this prediction, we discuss research by Sungkhassettee, Friedman, and Castel (2011), who showed that while disfluent inverted pairs were recalled at a greater rate relative to pairs presented in a standard upright position, this memory benefit was not reflected in JOLs. We have updated our predictions for the fluency and beliefs accounts to reflect this and state that both accounts predict lower JOLs for Sans Forgetica relative to Arial font.

**Comment #8:**  P. 24, line 36: in keeping with the last comment, I'm concerned about the authors arguing that the beliefs account was not supported.  There is no reason to expect people to form a belief that a desirable difficulty will yield higher memory performance.

***Response:*** We have updated the Experiment 3 discussion on pages 27-28 to be consistent with our revised predictions regarding the beliefs account. This section now states that our finding that JOLs are reduced for Sans Forgetica pairs compared to Arial pairs is in line with predictions from both the fluency and beliefs accounts.

**Additional Minor Edits:**  
1.      The word "judgments" is spelled differently (judgements, judgments) throughout the manuscript.  
2.      P. 14, line 12: we expected highlighting to operate similarly (add word to)  
3.      P. 16, line 41: consistent with the font size effect  
4.      P. 16, line 43: relative to small fonts and the expected interaction . . .  
5.      P. 23, line 26: exceeded

***Response*:** We appreciate your attention to detail. Each of these minor spelling and grammar mistakes have been corrected.