The judgment of learning (JOL) task is often used to assess memory monitoring at encoding. In the JOL task, participants study cue-target word pairs (e.g., mouse-cheese) and are asked to rate the probability of correctly recalling the target in the presence of the cue at test (e.g., mouse-?). Previous research has shown that the accuracy of JOLs is sensitive to perceptual cues, and these cues can induce *metacognitive illusions* in which JOLs overpredict subsequent recall. One example is the *font-size effect* (Rhodes & Castel, 2008), which occurs when participants inflate their JOL ratings when word pairs are presented in large font relative to small font, without concomitant increases to recall. While the font-size effect has been reported under several conditions, the underlying factors driving the effect remain unclear. Two theories have been proposed to account the font-size effect—the fluency account and the beliefs account. First, the fluency account states that larger words are more perceptually fluent than smaller words. Due to their greater perceptual fluency, participants process larger words more efficiently and/or effectively, leading to greater JOLs relative to smaller words (Undorf, Zimdahl, & Bernstein). Alternatively, the beliefs account posits that participants beliefs about an item’s memorability is the primary factor influencing JOLs. Regarding the font-size effect, participants may assign higher JOLs to large items because they hold the belief that large pairs are easier to learn than small pairs (Mueller, Dunlosky, Tauber, & Rhodes, 2014).

The present study provides an additional test of the font-size effect while also testing whether other perceptual manipulations can influence the correspondence between JOLs and recall. First, Experiments 1A and 1B were designed to replicate the font-size effect and tested whether JOL overestimation extended to a highlighting manipulation. Participants studied a mix of forward (e.g., credit-card), backward (e.g., card-credit), symmetrical (e.g., king-queen) and unrelated (e.g., muffin-floor) word pairs generated using the University of South Florida Free Association Norms (USF Norms; Nelson, McEvoy, & Schreiber, 2004). These pairs were presented either as a mix of large (54-pt.) and small (12-pt.) Arial font (Experiment 1A) or as a mix of 32-pt. Arial pairs that were either presented highlighted in yellow or presented with no highlight (Experiment 1B). Importantly, all experiments included an additional pure-control group comparison in which only the standard perceptual condition is used (i.e., all pairs presented in a standard font size), rather than a mix of perceptually fluent/disfluent pairs. These control groups were included because encoding manipulations have been shown to spill over into other encoding tasks when encoding is manipulated within-subjects (Huff, Bodner, & Gretz, 2021). This additional comparison group allowed us to gauge perceptual effects on JOLs more accurately relative to a baseline condition. Overall, in Experiment 1A, large font produced an equivalent boost to JOLs and recall relative to small font. However, JOLs and recall did not differ between large or small font pairs relative to the control group (Figure 1). In Experiment 1B, JOLs and recall did not differ as a function of highlighting, and no differences were detected between the highlighted pairs, non-highlighted pairs, or the control group (Figure 2). Thus, the font-size effect was not in evidence, and the effect failed to extend to highlighting.

Because Rhodes and Castel (2008) showed that the font-size effect was moderated by pair relatedness and strongest when using unrelated word pairs, we reasoned that the effect may have failed to emerge in Experiments 1A and 1B due to our inclusion of related pairs. To account for this, Experiments 2A and 2B provided additional tests of both font-size and highlighting on JOLs using only unrelated word pairs. However, in Experiment 2A, the font-size effect again was not in evidence, as JOLs and recall were again equally boosted for large pairs relative to small pairs. Again, no differences were detected relative to the control group (Figure 3). In Experiment 2B, JOLs and recall again did not differ as a function of highlighting, and no differences were detected relative to the control group (Figure 4).

Finally, Experiment 3 tested whether presenting pairs using Sans Forgetica—a perceptually disfluent font designed to improve memory (Earp, 2018)—would result in inflated JOLs. Recent research, however, suggests that this font does not improve memory and may induce a memory cost (Geller, Davis, & Peterson, 2020; Tayler, Sanson, Burnell, Wade, & Garry, 2020). However, although this font is disfluent, participants may hold the memory belief that Sans Forgetica will improve memory, resulting in inflated JOLs for Sans Forgetica pairs. Thus, Experiment 3 used this font to test the fluency and beliefs accounts of how perceptual features affect JOLs. Based on the fluency account, the disfluent nature of Sans Forgetica should result in lower JOLs relative to a more fluent font such as Arial. However, if participants hold a memory belief that pairs presented in Sans Forgetica are more likely to be correctly recalled at test, then JOLs may be inflated for the Sans Forgetica pairs. Overall, the perceptually disfluent Sans Forgetica font in Experiment 3 yielded a memory cost relative to pairs presented in standard Arial font, though it had no effect on JOLs. Furthermore, Arial pairs presented in mixed lists with Sans Forgetica pairs showed greater JOLs and recall rates than Arial pairs in the pure group, suggesting that the mixed list context increased participants both JOLs and the encoding of Arial pairs (Figure 5).

Collectively, this set of experiments provides a greater understanding of how perceptual features influence JOLs and recall, particularly within the context of cued-recall testing. In doing so, this research supports the Pychonomic Society’s mission of fostering the science of cognition via the advancement and communication of basic research in experimental psychology by (1) attempting to replicate the font-size effect, which is an established finding in metamemory research, (2) testing whether the effect can be produced via two additional manipulations (highlighting and Sans Forgetica), and (3) our novel inclusion of control groups, which allowed us to assess the effects of context on fluency effects, providing a more complete assessment of how these processes affect both JOLs and recall. In sum, this set of experiments suggests that perceptually fluent and disfluent study pairs do not appear to inflate JOLs at study.

References

Earp, J. (2018). Q&A: Designing a font to help students remember key information

Geller, J., Davis, S. D., & Peterson, D. J. (2020). Sans forgetica is not desirable for learning.

*Memory*, *28*(8), 957-967.

Huff, M. J., Bodner, G. E., & Gretz, M. R. (2021). Distinctive encoding of a subset of DRM lists yields not only benefits, but also costs and spillovers. *Psychological Research, 85*, 280-290.

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on

judgments of learning: Does it exemplify fluency effects or reflect people’s beliefs about

memory? *Journal of Memory and Language*, *70*, 1-12.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods,* *Instruments, & Computers*, *36* (3), 402–407.

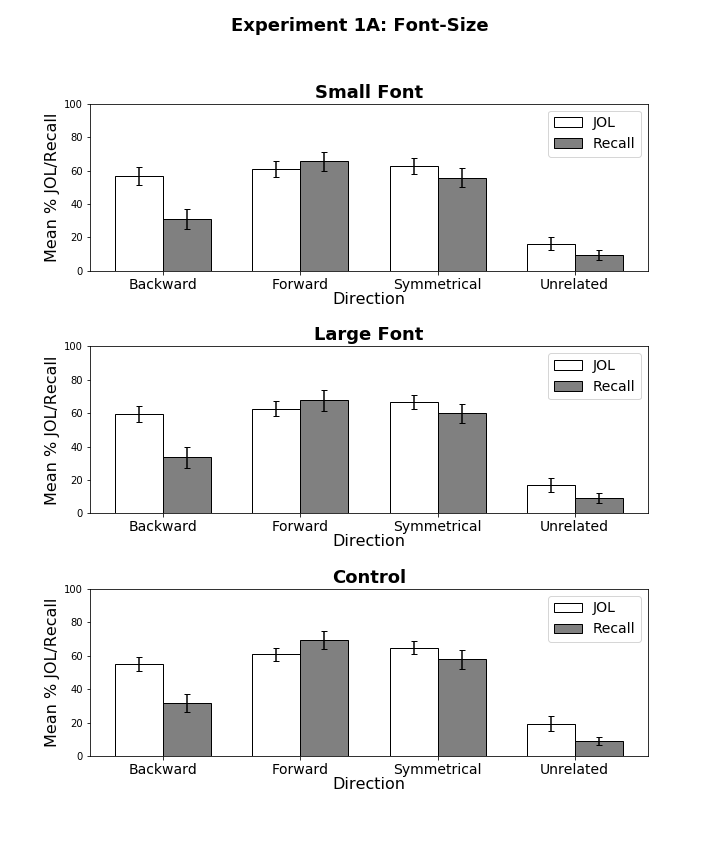
Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions: *Journal of Experimental Psychology: General, 137*(4), 615-625.

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects

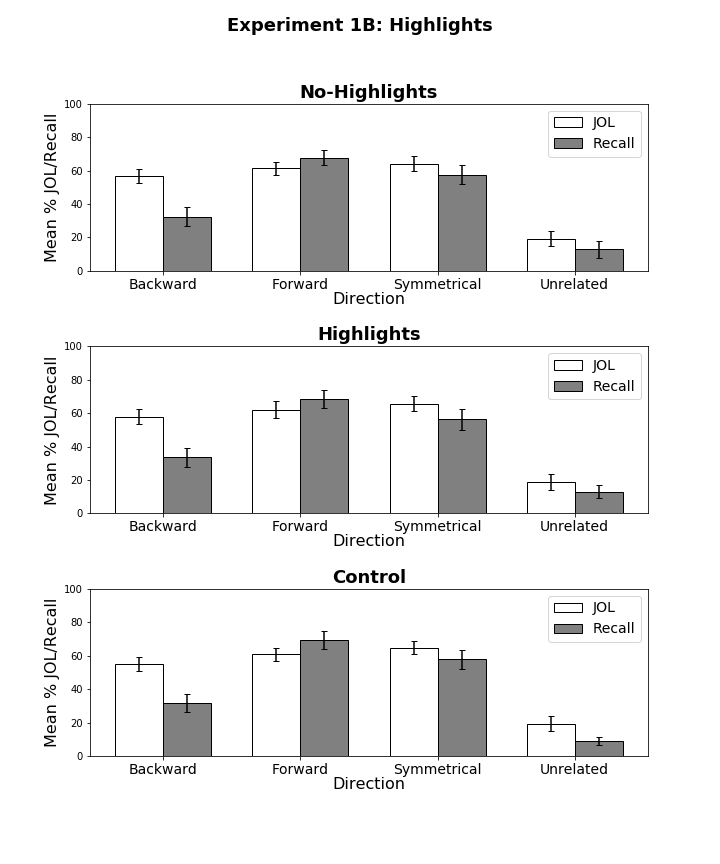
of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293-304.

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are

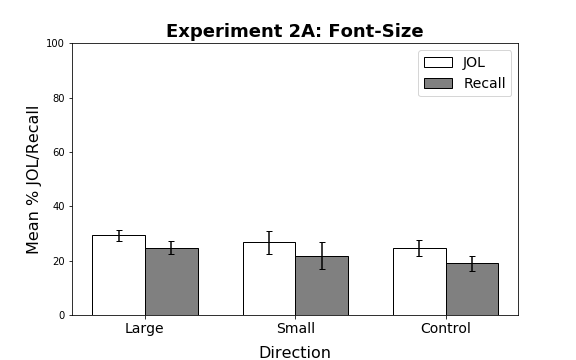
not desirable difficulties: the (lack of) effect of Sans Forgetica on memory. *Memory*, *28*(7), 850-857.



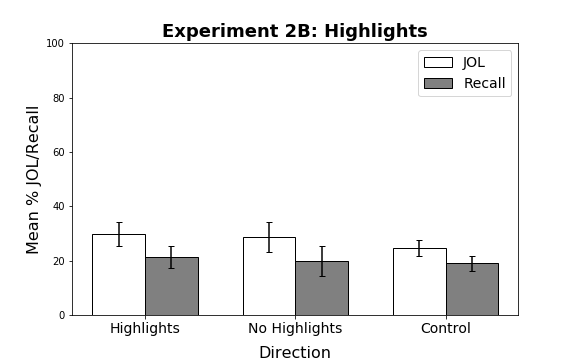
*Figure 1*. Mean JOL and recall rates as a function of pair type for pairs presented in small font (top panel), large font (middle panel), and the control group (bottom panel) as a function of pair direction in Experiment 1A. Bars represent 95% confidence intervals.



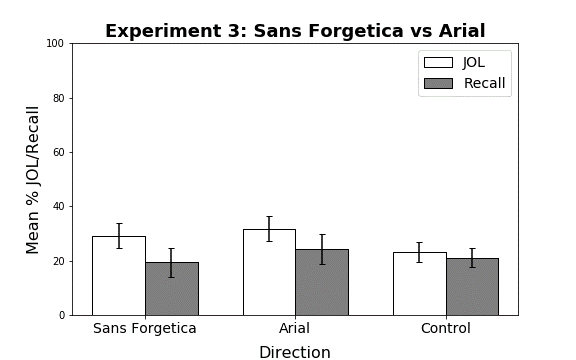
*Figure 2*. Mean JOL and recall rates as a function of pair type for highlighted pairs presented in mixed lists (top panel), non-highlighted pairs presented in mixed lists (middle panel), and non-highlighted pairs presented the control group (bottom panel) as a function of pair direction in Experiment 1B. Bars represent 95% confidence intervals.



*Figure 3*. Mean JOL and recall rates as function of pair type in Experiment 2A. Bars represent 95% confidence intervals. All study pairs were unrelated.



*Figure 4*. Mean JOL and recall rates as function of pair type in Experiment 2B. Bars represent 95% confidence intervals. All study pairs were unrelated.



*Figure 5*. Mean JOL and recall rates as function of pair type in Experiment 3. Bars represent 95% confidence intervals. All study pairs were unrelated.