Acta Psychologica 186 (2018) 27–38

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief☆

Adam B. Blake*, Alan D. Castel

Department of Psychology, University of California, Los Angeles, United States

A R T I C L E I N F O

Keywords:
Metamemory
Fluency
Beliefs
Judgments of learning

A B S T R A C T

A current debate in metamemory research considers the roles of fluency (Rhodes & Castel, 2008) and belief cues (Mueller, Dunlosky, & Tauber, 2015) in the construction of judgments of learning (JOLs). Under the fluency hypothesis, the processing fluency drives the construction of JOLs. The belief-based hypothesis instead asserts a strong influence of belief. The current study directly influences participants' beliefs to observe the effects on metacognitive judgments of learning. Experiment 1A attempted to variably strengthen participants' beliefs about the font-size effect: participants were informed of research implying a superiority of large-fonts in memory and between study cycles were given differing levels of feedback on the relationship between font-size and their performance. A reliable replication of the font-size effect was found, though there were no appreciable differences due to feedback. Experiment 1B introduced a counter-belief that small-fonts are superior in terms of memory recall, and participants showed minimal effects of font-size on JOLs and none on memory. Finally, in Experiment 2 participants underwent a baseline condition with no belief manipulation and then an experimental condition that manipulated beliefs as in Experiments 1A and 1B. The results provide clear confirmatory evidence for the effects of belief on JOLs, though these data neither support a pure fluency hypothesis nor a pure belief-based hypothesis. We discuss an additive effect of perceptual fluency and belief on JOLs, and present possible mechanisms that may interact to influence and bias JOLs.

For any learner, the ability to assess, accurately, the precise quality and extent of encoding in memory would be a fantastic tool. Though people often make accurate metacognitive judgments, they sometimes show interesting and important dissociations between their memories and their judgments about their memories (Rhodes, 2015). Many learners use assessments of memory to regulate their study schedules (Finn & Metcalfe, 2008; Nelson, Dunlosky, Graf, & Narens, 1994), but these assessments may be prone to error due to the inherently inferential nature of metacognitive judgments (Kelley & Rhodes, 2002; Schwartz, Benjamin, & Bjork, 1997). Research on JOLs has shown that peculiar but predictable metacognitive errors can arise due to influence from peripheral and irrelevant factors such as perceptual fluency (e.g., Rhodes & Castel, 2008) or beliefs regarding the saliency of specific features (e.g., Mueller, Dunlosky, Tauber, & Rhodes, 2014). Often, these factors do not positively correlate with memory performance yet exhibit an effect on metamemory nonetheless. A current debate in the metacognitive literature is whether belief or fluency is the primary instigator for these errors in metacognitive judgment.

There are many ways to conceptualize fluency and its effects on judgments, though generally fluency is considered the subjective ease-of-processing of information (Alter & Oppenheimer, 2009). There is a varied body of literature that has examined the ways fluency impacts metacognitive judgment. Often these studies focus on the negative influence of different types of fluency and how they can create metacognitive illusions of confidence. For example, in the case of *perceptual fluency*, learners rate items as better remembered when they are presented in a larger font (Rhodes & Castel, 2008) or louder volume (Rhodes & Castel, 2009). Similarly, learners are susceptible to *retrieval fluency* where items which come to mind easily are deemed better learned (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin, Bjork, & Schwartz, 1998).

It is a strong view in the metacognitive literature that these processing fluencies nearly exclusively influence JOLs (Begg et al., 1989; Koriat & Bjork, 2006; Koriat, Bjork, Sheffer, & Bar, 2004; Kornell, Rhodes, Castel, & Tauber, 2011). In a particular demonstration of the effects of perceptual fluency on metacognitive judgments, participants

☆ We thank Robert Bjork, Elizabeth Ligon Bjork, and Barbara Knowlton for helpful comments regarding this work; and Meenely Nazarian, Monika Holser, Angelica Sanchez, and Rachel Todd for help with data collection. Portions of this work were presented at the 55th annual meeting of the Psychonomic Society, Long Beach, CA. Please address all correspondence to Adam B. Blake, Department of Psychology, University of California, Los Angeles, 1285 Franz Hall Box 951563, Los Angeles, CA 90095.
* Corresponding author.
E-mail address: adamblake@ucla.edu (A.B. Blake).

https://doi.org/10.1016/j.actpsy.2018.04.004
Received 5 December 2017; Received in revised form 16 March 2018; Accepted 13 April 2018

have a strong bias towards words presented in larger versus smaller fonts (Rhodes & Castel, 2008). Presumably participants experience a higher degree of perceptual fluency when reading large-font words as opposed to small font words which may be perceived as harder to read. The strong argument in that study was that when the perceptual fluency of larger font words is equated with smaller font words, the bias is eliminated: words presented in alternating capital letters (e.g. "hElLo" "WoRlD") are dis-fluent in any font size and the font size bias is removed (Rhodes & Castel, 2008). Similarly, when the perception of a word has been reduced or altered via backward masking, people are likely to give lower JOLs though subsequent memory is not impaired (Besken & Mulligan, 2013). Interestingly, the perceptual characteristics only appear to have a strong effect immediately after presentation, and a delay is sufficient to remove their bias in participants (Luna, Martín-Luengo, & Albuquerque, 2017), suggesting that perceptual fluency is a factor that is weighted less when compared against more diagnostic factors like retrieval fluency. Nevertheless, perceptual characteristics *do* have an effect on participants' choices to restudy information, and thus merit further study (Luna et al., 2017).

A criticism of Rhodes and Castel (2008) suggests that the font-size manipulation is not enough to affect processing fluency, and thus the differences in JOLs across font size must be due to another factor. Specifically, when items on a lexical decision task are manipulated with the same font-sizes there are no differences in classification times across font-size (Mueller et al., 2014). However, similar effects of fluency occur for more direct manipulations of perceptual fluency as well: participants rate blurry words as less memorable than clear words, though there is little, if any, effect on memory (Yue, Castel, & Bjork, 2013). Regardless, it is a valid point that for the materials commonly used to show the font-size effect there may be no true differences in processing fluency.

The competing explanation for the variability in JOLs posits that JOLs are driven instead by participants' beliefs and expectations about the memorability of words in larger fonts (Mueller et al., 2014; Mueller, Tauber, & Dunlosky, 2013). Under this view, it is not that fluency has direct effects on JOLs, but rather that it takes the form of a heuristic: subjectively easily processed information must result in easily recalled information (Koriat, 2008; Miele, Finn, & Molden, 2011). A fairly direct method of measuring one's belief about a stimulus is to use a pre-JOL paradigm where the judgment must be made before seeing the word (see for example Castel, 2008). In the particular case of assessing belief in the face of fluency, judgments made prior to the stimulus are necessarily unaffected by the fluency of the following stimulus. With only their beliefs to guide them, participants still show a bias towards larger fonts, providing strong evidence for the influence of beliefs when making metacognitive judgments (Mueller et al., 2014).

Logically, when utilizing a belief-based system for JOLs one would expect that dispelling any erroneous belief about font-size should in turn dispel the effect. When participants are explicitly told to ignore any variances in font-size when making JOLs because they are not predictive of later recall, there is in fact a dramatic reduction, though not complete elimination, of the font-size effect (Rhodes & Castel, 2008). A very similar finding was shown in a study examining the effects of perceptual fluency on JOLs in auditory stimuli, where participants failed to discount perceptual fluency as a predictor for learning even when the experimenters gave overt warnings that it was non-diagnostic of memory (Besken & Mulligan, 2014). It would appear that the competing belief introduced by the experimenter had a direct effect on judgments. Alternatively, this belief can be introduced experientially: after studying and testing their knowledge of a word list, participants have the opportunity to assess their own metacognitive decisions. When given a second study-test cycle participants show effects of debiasing, generally exhibited through lower confidence and JOLs, for large-font words (Rhodes & Castel, 2008), though this interaction was not statistically significant.

This refinement of judgment through experience has been shown in multiple cases, providing some further evidence suggesting that there may be some effect of belief behind any differences in JOLs (Benjamin, 2003; Koriat & Bjork, 2006; Rhodes & Castel, 2008). An explanation for this refinement is that participants become more sensitive to the manipulated mnemonic cues, for example, associative direction and strength of word pairs in Koriat and Bjork (2006) or word frequency in Benjamin (2003). However, in the case of the font-size effect, participants tend to believe that font-size influences memory (Kornell et al., 2011; Mueller et al., 2014) and if they were to rely on this "mnemonic cue" one would expect an increase in the difference between large- and small-font JOLs across lists. Instead, there a numerical reduction (not statistically significant) in the effect (e.g., Rhodes & Castel, 2008, Experiment 2), which possibly indicates some type of discounting of the effects of perceptual fluency. In all of these studies, there is no change in the type of items presented across list, yet average predicted performance is more closely aligned with average memory performance in an absolute sense. Since there is no perceptual change introduced during encoding across lists, one must infer that participants developed different beliefs about the materials, beliefs that affected their judgments.

Importantly though, the above description of experiential debiasing is not necessarily affecting *all* beliefs about list-learning. That is, there are various beliefs that people have when approaching a study-test cycle, one of which is likely an overly optimistic idea of their memory capacity as a whole. In a study-test cycle using items in different font-sizes, this optimism exists alongside their beliefs regarding font-size and perceptual fluency. When a learner recalls fewer items than expected, she may refine her beliefs about her memory capacity, but not those about perceptual fluency. This refinement is reflected in an experiment showing that the basic font-size effect persists across multiple lists despite reductions in JOLs as a whole (Rhodes & Castel, 2008). It is possible that multiple biases (font-size, memory capacity, clarity, accessibility) are adjusted together through a common experience, though it is not clear in the data whether this occurs or if the basic JOL paradigm allows for it.

Though Mueller et al. (2014) very importantly showed that participants *have* beliefs about font-size that may result in a font-size bias, these beliefs were assessed in the absence of the stimuli (i.e., before encoding the information). Frank and Kuhlmann (2017) show that even participants that do not report beliefs about fluency (in this case volume of a spoken word) still exhibit a pattern of responding consistent with a bias towards more perceptually fluent items. Further, there is still the unanswered question of whether the effects of belief persist in the presence of perceived differences in fluency, that is, whether beliefs are considered during a post-stimulus paradigm and affect on-line JOLs rather than fluency taking the reins, so to speak. The approach in the current study is to examine the font-size effect (Rhodes & Castel, 2008) using the basic post-stimulus JOL paradigm.

Finally, the following studies take a direct approach in manipulating participants' beliefs about font-size, a distinction that separates the current work from the current corpus of work on the effects of belief and fluency on font-size. To complement and extend the finding that participants have prior beliefs about font-size that affect their JOLs, the current study employs methods to manipulate participant beliefs about font-size to assess the direct effects of belief. These methods follow both a belief-strengthening paradigm (Experiments 1A and 1B), where participants are given information that is intended to increase the intensity of their beliefs and thus affect their subsequent JOLs, as well as a counter-belief paradigm (Experiments 1B and 2), where participants are introduced to research that runs counter to their current beliefs about their own memory.

1. Experiment 1A

If metamemory follows a belief-driven model, it should be possible to manipulate intensity of belief and observe an effect on subsequent judgments. That is, the hypothesis implies that stronger beliefs about font-size can lead to greater differences between JOLs for smaller and larger font words. In this experiment participants were informed of some suggestive findings that showed that larger-font words, relative to smaller-font words, may be easier to recall for college students (something that could be true in certain settings, but was not anticipated by the authors in the current design). Many people have pre-existing beliefs that words in larger fonts are easier to learn and remember (Kornell et al., 2011; Mueller et al., 2014; Rhodes & Castel, 2008), and the introduction of the suggestive research supported this notion. This was intended to ensure that all participants have this belief, have recently considered it, and to lay a foundation for any later confirmatory evidence.

According to the knowledge-updating literature, participants should be updating their judgments of learning following a first study-test cycle (cf. Mueller et al., 2015). Before starting a second study list, it is expected that participants will both experience a deficit in their recall performance relative to their JOLs on the first list, and if given no specific feedback, will likely not have enough information available to evaluate their performance for each font size. At this point, any information given regarding their specific item performance is expected to be utilized when participants update their understanding of their own memories.

To assess whether stronger beliefs alter JOLs, differing levels of confirmatory feedback were administered between study-test cycles. Confirmatory feedback has been shown to strengthen beliefs and can increase false memories (Zaragoza, Payment, Ackil, Drivdahl, & Beck, 2001), and it is expected that participants given confirmatory feedback suggesting they recalled more of one font than another will use this information to update their JOLs on the next study list, thereby increasing the font-size effect. If JOLs are belief-driven, more strongly held beliefs should increase the magnitude of the difference between JOLs for large- and small-font words.

1.1. Method

1.1.1. Participants and design

The participants were 88 introductory psychology students from the University of California, Los Angeles, who participated for course credit. Participants completed two study-test sets where the font-size of the studied items was manipulated within-subjects along two levels (large, 48 pt.; small, 18 pt). Instructions were manipulated between subjects prior to the second set across three levels: none, where no additional instruction was given; repeated, where the study's initial instructions were shown on-screen again; and feedback, where prescripted feedback on performance was given. Two dependent variables, JOL and recall, were measured for each participant.

1.1.2. Materials

Two study lists of 42 nouns each were taken from the Kučera and Francis (1967) norms. For each participant, each list was randomly divided into two sets of 18 items that were presented equally often in 18 pt. or 48 pt. Arial font. The remaining six items served either as primacy or recency buffers, presented equally often in 18 pt. or 48 pt. font, and were excluded from all analyses reported. The study lists were equated for frequency ($M = 45.35$), number of syllables ($M = 1.76$), and number of letters ($M = 5.74$), again using the Kučera and Francis (1967) norms. The presentation of the lists was counterbalanced such that half of the participants saw the one list first, and the other half saw the other list first.

1.1.3. Procedure

After providing informed consent, participants were told that they would study words presented in different font sizes. Additionally, they were given the information that "research has shown that, for college-age participants, words in larger fonts are easier to recall than words in smaller fonts," but that they should attempt to remember as many words as possible. After receiving these instructions participants both had to acknowledge to the experimenter that they read and understood the instructions, and then answer a question about the prompt. The answer to the question reiterated the information about font-size and memory. Participants then began the presentation of the first study list. All study stimuli were presented for 4 s each in white, lowercase letters in the center of a black background on a computer screen. Immediately following the presentation of each word participants were given 4 s to rate their confidence that they would be able to recall that item on a scale from 0% (no chance of recall) to 100% (certain of recall). Participants were encouraged to use the entire range of the scale. Immediately following the study lists participants engaged in a 4-min distractor task where they recalled either as many Presidents of the United States as they could or as many states of the United States as they could. After this distractor, they were given 4 min to recall as many of the items as they could from the study list on a blank sheet of paper.

After recall of the first list and before presentation of the second the instructions were manipulated to vary the intensity of the belief for each group. In one group, no additional instructions were given between lists (low-intensity), a second group received the exact same instructions that preceded the first list (medium-intensity), and the final group was given the instructions again along with general feedback that they had performed better on words in larger font sizes (high-intensity). Following this, participants received a new study list, distractor task, and recall phase in the same fashion as the first. After completing the study, participants were debriefed and informed of the specific instances when font-size is likely to impact memory, e.g. the Von Restorff effect (Von Restorff, 1933).

1.2. Results

To test these any differences in JOLs and performance across font size and list, as well as the effects of instruction, a 3 (instruction) × 2 (list) × 2 (font) mixed ANOVA, with instruction as the between-subjects factor, was performed. The alpha level was set to 0.05 and all effect sizes are reported in terms of η_p^2 for ANOVAs and Cohen's d for t-tests.

1.2.1. JOLs

We first examined the data for any effects of the central instruction manipulation on JOLs, and there was neither a significant three-way interaction between instruction, font, and list [$F(2, 85) = 1.08$, $\eta_p^2 = 0.025$, $MSE = 46.56$, $p = 0.34$]; nor a significant two-way interaction of instruction and font [$F(2, 85) < 1$, $\eta_p^2 = 0.008$, $MSE = 89.25$]; nor a significant two-way interaction of instruction and list [$F(2, 85) < 1$, $\eta_p^2 = 0.013$, $MSE = 127.74$]. Finally, instruction had a similar effect on JOLs regardless of whether participants received no instructions between lists ($M = 45.53$, $SD = 13.89$), repeated instructions between lists ($M = 48.25$, $SD = 18.49$), or feedback with instructions between lists ($M = 46.63$, $SD = 17.54$), $F(2, 85) < 1$, $\eta_p^2 = 0.006$, $MSE = 857.40$. The effects regarding the instruction manipulation being null, the descriptive statistics for JOLs have been collapsed across the three instructional conditions in the analyses below (see Table 1 for the full set of means and standard deviations).

Fig. 1 shows the effect of font-size in both lists collapsed across the three instruction types. On average, small-font words appeared to be given lower JOLs than words in large font. This trend held across both lists, with what appears to be a reduction in the magnitude of the effect as well as an increase in the calibration of the JOLs and performance across lists. As suggested by Fig. 1, the font-size effect for JOLs was reliably replicated in this experiment as well as the hypothesized reduction in the font-size effect across lists, yielding a significant interaction of font-size and list, $F(1, 85) = 10.49$, $\eta_p^2 = 0.110$, $MSE = 46.56$, $p = 0.002$. As shown in Fig. 1, in the first study list participants gave higher JOLs for words in 48 pt. font ($M = 57.95$, $SD = 16.17$) than

Table 1
Means and standard deviations for each cell of Experiment 1A.

Condition	Measure	List	Font-size	Mean	SD
Low-intensity (n = 30)	JOLs	1	Small	42.46	14.22
			Large	55.93	14.08
		2	Small	38.69	14.38
			Large	45.05	12.83
	Recall	1	Small	26.03	13.34
			Large	23.81	13.87
		2	Small	27.62	16.56
			Large	28.41	15.84
Medium-intensity (n = 30)	JOLs	1	Small	45.52	18.78
			Large	59.61	19.16
		2	Small	39.47	17.35
			Large	48.40	18.63
	Recall	1	Small	23.33	16.56
			Large	22.86	13.95
		2	Small	29.37	18.59
			Large	27.30	14.63
High-intensity (n = 28)	JOLs	1	Small	44.34	19.28
			Large	58.33	15.07
		2	Small	35.87	17.85
			Large	47.98	17.68
	Recall	1	Small	23.98	10.16
			Large	22.11	13.42
		2	Small	27.89	16.17
			Large	32.14	16.68

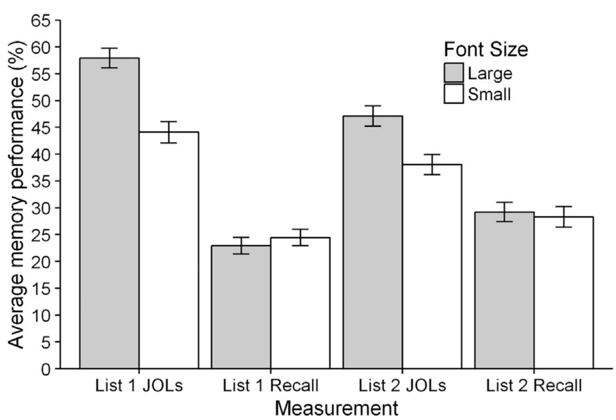

Fig. 1. Mean predicted recall (JOLs) and free recall for words in Experiment 1A, divided by list and by font-size. Error bars represent standard errors of the mean.

those in 18 pt. font ($M = 44.10$, $SD = 17.38$), $t(87) = 9.60$, $d = 1.02$, $p < 0.001$. This pattern persisted to the second study list where participants again rated words in 48 pt. ($M = 47.13$, $SD = 16.43$) with higher JOLs than 18 pt. words ($M = 38.06$, $SD = 16.45$), $t(87) = 6.66$, $d = 0.71$, $p < 0.001$. Note both the large overall reduction in JOLs across lists ($M_{list1} = 51.03$, $M_{list2} = 42.60$) and the decrease in the effect size of the font-size effect ($d_{list1} = 1.02$, $d_{list2} = 0.71$) which reflect the relative debiasing and calibration of JOLs across lists with respect to the recall scores reported below. The reduction in JOL magnitudes across lists was significant in both the judgments for large font words [$t(87) = 9.27$, $d = 0.99$, $p < 0.001$] and small font words [$t(87) = 4.64$, $d = 0.49$, $p < 0.001$].

1.2.2. Gamma correlations

Metacognitive accuracy can also be operationalized as the degree to which the magnitude of a JOL for a word was associated with the probability that that word was recalled. This measure is generally referred to as metacognitive resolution and can be computed as the Goodman-Kruskal γ correlation between JOLs and recall for each participant (Nelson, 1984). In the following analyses, note that the degrees of freedom may differ from those in the JOL or recall analyses because the correlation cannot be computed when there is not enough variance in participants' responses.

Considering first the effects of the instructional manipulation on metacognitive resolution, participants were no more accurate when given confirmatory feedback ($\gamma = 0.30$, $SE = 0.08$), than when they were given repeated instructions ($\gamma = 0.32$, $SE = 0.08$), or no between list instructions ($\gamma = 0.26$, $SE = 0.08$), $F(2, 72) < 1$, $\eta_p^2 = 0.011$, $MSE = 0.07$. This lack of a main effect for the instructional condition was neither qualified by a significant two-way interaction with font-size [$F(2, 72) = 2.123$, $\eta_p^2 = 0.056$, $MSE = 0.12$, $p = 0.13$], nor list [$F(2, 72) < 1$, $\eta_p^2 = 0.007$, $MSE = 0.12$]; nor was there a significant three-way interaction [$F(2, 72) = 1.01$, $\eta_p^2 = 0.027$, $MSE = 0.16$, $p = 0.37$].

Next, γ was not significantly different for items in large font ($\gamma = 0.30$, $SD = 0.05$) versus small font ($\gamma = 0.29$, $SE = 0.04$), $F(2, 72) < 1$, $\eta_p^2 = 0.001$, $MSE = 0.12$. Similarly, there was no significant difference between the first list ($\gamma = 0.30$, $SE = 0.05$) and the second list ($\gamma = 0.29$, $SE = 0.04$), $F(2, 72) < 1$, $\eta_p^2 < 0.001$, $MSE = 0.12$. Further, there was no significant two-way interaction of font-size and list, $F(2, 72) < 1$, $\eta_p^2 = 0.01$, $MSE = 0.16$.

Finally, we drew from Koriat (1997) and examined the correlation between font-size and JOLs, as opposed to recall and JOLs. In Koriat (1997), the materials were manipulated to be more or less difficult, however, here we use perceptual fluency as a proxy for item difficulty. To interpret these values, values father from 0 (bounded at −1, 1) will indicate more of a reliance on font-size, where positive values indicate that participants find larger items more memorable and negative values indicate participants find smaller items more memorable. A two-way ANOVA comparing the Goodman-Kruskal gamma correlations across list and condition was computed. There were no significant differences in correlations between conditions [$F(2, 85) < 1$, $\eta_p^2 = 0.01$, $MSE = 0.15$], nor was there a significant interaction [$F(2, 85) < 1$, $\eta_p^2 = 0.01$, $MSE = 0.07$]. However, there was a decreased correlation between font-size and JOLs from the first list ($\gamma = 0.38$, $SE = 0.03$), to the second list ($\gamma = 0.29$, $SE = 0.04$), $F(1, 85) = 6.02$, $\eta_p^2 = 0.07$, $MSE = 0.07$, $p = 0.02$.

1.2.3. Recall

The results for the effects of instruction on memory performance were similar to those for the JOL measure: there was neither a neither a significant three-way interaction between instruction, font, and list [$F(2,85) = 1.16$, $\eta_p^2 = 0.026$, $MSE = 94.79$, $p = 0.32$]; nor a significant two-way interaction of instruction and font [$F(2, 85) < 1$, $\eta_p^2 = 0.016$, $MSE = 161.07$]; nor a significant two-way interaction of instruction and list [$F(2, 85) < 1$, $\eta_p^2 = 0.009$, $MSE = 129.96$]. Lastly, main effect of instructions was not significant such that participants remembered similar percentages of the word lists when no instructions were given between lists ($M = 26.47$, $SD = 14.96$), instructions were repeated between lists ($M = 25.72$, $SD = 16.03$), and feedback was given with the instructions ($M = 26.53$, $SD = 14.34$), $F(2, 85) < 1$, $\eta_p^2 = 0.001$, $MSE = 532.07$. The effects regarding the instruction manipulation being null, the descriptive statistics for recall have been collapsed across the three instructional conditions in the analyses below (see Table 1 for the full set of means and standard deviations).

Like past research, the recall scores did not differ across font-size in this experiment. A similar percentage of words in large font ($M = 26.08$, $SD = 14.68$) were recalled as words in small font ($M = 26.38$, $SD = 15.37$), $F(1, 85) < 1$, $\eta_p^2 = 0.001$, $MSE = 129.96$. This pattern was consistent across lists and there was no significant interaction of font-size and list, $F(1, 85) = 1.47$, $\eta_p^2 = 0.017$, $MSE = 94.79$, $p = 0.23$. Finally, there was a main effect of list on recall such that a higher percentage of words were recalled on the second list ($M = 28.76$, $SD = 16.34$) than the first list ($M = 23.70$, $SD = 13.59$), $F(1, 85) = 14.21$, $\eta_p^2 = 0.143$, $MSE = 161.07$, $p < 0.001$.

1.2.4. Post-task items

To ensure that the instructions and research were believed by participants, a manipulation check was included in the post-task questionnaire. When asked if they had believed the research presented in the instructions at the start of the experiment, 62.50% indicated that they definitely believed the research.

1.3. Discussion

In this experiment, we attempted to bias people's beliefs regarding font size so that they were consistent with, and possibly stronger than, the standard belief that larger font words are easier to remember than smaller font words. Particularly, we supplied participants with information on font-size and memory that coincides with the very common belief that words in larger font are easier to recall. We expected that when participants' beliefs were reinforced participants would give JOLs consistent with their strengthened beliefs: increased JOLs for large words and decreased JOLs for words in small font with larger differences at each level of the instruction manipulation.

It is unclear why the instructional manipulation did not cause any effects on the font-size effect. A possible explanation is that participants did not believe the information being presented to them via the instructions, a point which is supported by the fact that only ~63% reported accepting the "research" they were presented with. However, this check was performed after the participants had undergone the experiment, and additional analyses performed on that subset yielded the same patterns presented above. Instead we speculate that participants already had a strong belief about how font-size would affect memory. This prior belief is discussed in the introduction of this paper, and has been shown multiple times in past research (e.g., Kornell et al., 2011; Mueller et al., 2014). Additionally, the initial instructions may have already amplified the font-size bias which is suggested by the slightly larger effect size shown in list one of this experiment ($d = 1.02$) compared to Rhodes and Castel (2008) Experiment 1 ($d = 0.83$, computed from their reported statistics). It is plausible that participants were already at or near ceiling for this belief and thus unable to strengthen it.

Though the expected finding was not shown, it is important to note that participants did show a strong font-size bias towards larger font items. The expected overconfidence in JOLs on list one was found, and the overall decrease in JOLs across lists likely reflects experiential debiasing regarding memory capacity. That is, that participants experienced the limits of their memory capacity in the context of the task, and on the second list, generally felt that items were less likely remembered. That this improved accuracy was not reflected in participants' metacognitive resolution (γ) suggests that these judgments were not updated at the item level but as a global reduction in expected performance. Recent work suggests this may be because they began using number of items as a metacognitive cue, or at least that it became a more salient cue (Murayama, Blake, Kerr, & Castel, 2016).

At a glance, the effect size of the font-size effect appears to be diminished across lists, a finding which if significant would suggest that across lists participants reduced their font-size bias. However, this is a misleading line of thought because the higher-order analysis of the interaction between font-size and list was non-significant, meaning that the difference in effect size was negligible. However, there is evidence to suggest that participants were less reliant on font-size as a predictor of learning: there was a significant reduction in the correlation between font-size and JOLs from the first list to the second. If we interpret this reduction in correlation in the same light as Koriat (1997) did with difficulty and JOLs, we might conclude that participants partially shifted their focus from font-size to some other factor from the first list to the second.

On the whole, it appears that the instructional manipulation was not strong enough to alter participants' JOLs, that their bias towards font-size was already very high, and that their judgements appear to be less related font-size differences across lists.

2. Experiment 1B

Experiment 1A was unable to show an effect of instruction in terms of strengthening belief in the font-size effect. The interaction of font-size and list suggested a possible experiential debiasing of their beliefs about font-size but this is a tenuous claim. The current experiment took a more direct approach to manipulating participants' beliefs by introducing a competing belief. Informing participants of the true, null correlation between font-size and memory has been shown to dramatically reduce bias, however this does not eliminate the font-size effect (Rhodes & Castel, 2008). The lack of elimination there could mean that participants either did not trust the information they were being given, or that the fluency of the text is truly driving the results. The current experiment avoids subtlety and provides a belief that is in direct opposition of the common belief about font size: *smaller* words are easier to remember (something that could be true in certain settings, but not in the current design). A belief-based account for JOLs should hypothesize that when participants believe that smaller words are more memorable than large words there will be a reversal in the font-size effect such that JOLs are larger for smaller words. This would be completely counter to a fluency account which would hypothesize that 18 pt. font words, which are presumably less fluent than 48 pt. words (Rhodes & Castel, 2008), should always be given lower JOLs. Additionally, as this belief runs counter to normal intuitions about font-size and recall, it is expected to have a larger, more noticeable effect than the previous attempt in Experiment 1A to strengthen what was possibly an already strong belief.

Because this manipulation was expected to possibly reduce, eliminate, or reverse the font-size effect, a power analysis was performed using the effect size from Rhodes and Castel (2008) Experiment 4 ($\eta_p^2 = 0.13$) and a modest correlation among the measures (0.3). This particular effect size was chosen because (1) in that experiment the authors warned the participants that font-size is not diagnostic of memory performance, and (2) it was the lowest effect size of the full study and thus the most conservative estimate (in comparison the largest effect size was 0.45 and the average in the study was 0.36). The results of the power analysis suggested that in any given group judging words in small and large font a sample of at least 21 should show the font-size effect in JOLs.

2.1. Method

Participants were 90 workers recruited through Amazon Mechanical Turk and were paid $4.50 to participate in the study. The experiment used the same design, procedure, and materials as in Experiment 1A with the exception of the research that was presented prior to study. Instead of informing participants of research showing that large words are easier to remember, the opposite information was given: smaller words are easier to remember. As in Experiment 1A, the high-intensity condition participants were told that their recall on the first study list was consistent with the stated research, the moderate-intensity condition participants were only reminded of the stated research after the first recall session, and the low-intensity condition participants were given the manipulation at the beginning of the study and received no further instruction regarding it.

Following the experiment, participants were told about the goals of the experiment, and thanked for their participation. After eliminating 12 participants for failure to complete the full experiment, the final group sizes for the between-subjects instruction manipulation were 30 for low-intensity, 21 for moderate-intensity, and 27 for high-intensity.

As a note, because these workers used their own computer and browser settings, the exact size of the font could not be controlled. However, the ratio of the font sizes was preserved (48:18).

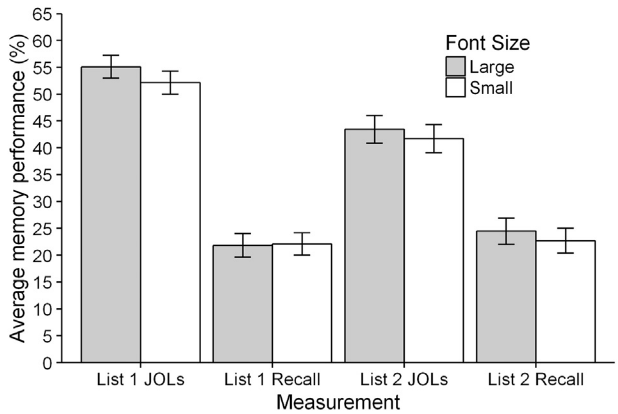

Fig. 2. Mean predicted recall (JOLs) and free recall for words in Experiment 1B, divided by list and by font-size. Error bars represent standard errors of the mean.

2.2. Results

A simplified representation of the data is presented in Fig. 2, which shows the effect of font-size in both lists collapsed across the three instruction types. There appear to be no strong differences across font-size in either the JOL or recall measures, regardless of list. A 3 (instruction) × 2 (list) × 2 (font) mixed ANOVA, with instruction as the between-subjects factor, was performed to test the effects of the manipulations on JOLs and recall scores. The alpha level was set to 0.05 for all inferential statistics, and all effect sizes are reported in terms of η_p^2 for ANOVAs.

2.2.1. JOLs

Examining the data for the effects of instruction, there was neither a significant three-way interaction between instruction, font, and list for JOLs [$F(2, 75) < 1$, $\eta_p^2 = 0.0067$, $MSE = 40.40$]; nor a significant two-way interaction of instruction and font [$F(2, 75) < 1$, $\eta_p^2 = 0.017$, $MSE = 80.51$]; nor a significant two-way interaction of instruction and list [$F(2, 75) < 1$, $\eta_p^2 = 0.017$, $MSE = 163.99$]. Lastly, participants gave similar JOLs regardless of whether they received no instructions between lists ($M = 46.68$, $SD = 20.39$), repeated instructions ($M = 43.69$, $SD = 17.94$), or feedback with repeated instructions ($M = 53.09$, $SD = 19.23$), $F(2, 75) = 1.88$, $\eta_p^2 = 0.048$, $MSE = 1215.00$, $p = 0.160$. As no significant effects regarding the instruction manipulation were found, the descriptive statistics for JOLs have again been collapsed across the three conditions in the analyses below (see Table 2 for the full set of means and standard deviations).

Interestingly, JOLs for words in large font ($M = 49.19$, $SD = 19.36$) were always higher than JOLs for words in small font ($M = 46.92$, $SD = 19.60$). The interaction between font size and list was indeed non-significant [$F(2, 75) < 1$, $\eta_p^2 = 0.01$, $MSE = 40.40$] and the main effect of font size was significant, $F(1, 75) = 5.35$, $\eta_p^2 = 0.067$, $MSE = 80.51$, $p = 0.024$. Thus, the observed interaction of font and list on JOLs found in Experiment 1A was not replicated in Experiment 1B, where the instruction manipulation was altered. A reduction of JOLs across list was evident [$F(1, 75) = 54.74$, $\eta_p^2 = 0.422$, $MSE = 163.99$, $p < 0.001$], but this reduction was consistent across font size.

2.2.2. Gamma correlations

As in Experiment 1A we considered the effects of the independent variables on metacognitive resolution. Twenty-one participants were excluded from these analyses because their correlations could not be computed due to invariance in either their JOLs or recall in one or more conditions.

First, we analyzed the effects of the instructional manipulation on metacognitive resolution and found that participants were no more accurate when given confirmatory feedback ($\gamma = 0.28$, $SE = 0.08$), than when they were given repeated instructions ($\gamma = 0.31$, $SE = 0.10$), or no between list instructions ($\gamma = 0.29$, $SE = 0.09$), $F(2, 54) < 1$, $\eta_p^2 = 0.010$, $MSE = 0.22$. There were neither a significant two-way interaction with font-size [$F(2, 54) < 1$, $\eta_p^2 < 0.001$, $MSE = 0.11$], nor with list [$F(2, 54) = 1.74$, $\eta_p^2 = 0.060$, $MSE = 0.15$, $p = 0.39$]; nor was there a significant three-way interaction [$F(2, 54) = 1.51$, $\eta_p^2 = 0.053$, $MSE = 0.11$, $p = 0.23$].

Looking at font-size, γ was not significantly different for items in large font ($\gamma = 0.30$, $SD = 0.05$) versus small font ($\gamma = 0.26$, $SE = 0.05$), $F(2, 54) < 1$, $\eta_p^2 = 0.010$, $MSE = 0.11$. Similarly, there was no significant difference between the first list ($\gamma = 0.28$, $SE = 0.08$) and the second list ($\gamma = 0.28$, $SE = 0.05$), $F(2, 54) < 1$, $\eta_p^2 < 0.001$, $MSE = 0.15$. Further, there was no significant two-way interaction of font-size and list, $F(2, 54) < 1$, $\eta_p^2 = 0.010$, $MSE = 0.11$.

Lastly, we again computed a two-way ANOVA comparing the Goodman-Kruskal gamma correlations across list and condition to assess changes in reliance on font-size as a predictor of learning. There were again no significant differences in correlations between conditions [$F(2, 74) = 1.33$, $\eta_p^2 = 0.04$, $MSE = 0.16$, $p = 0.27$], nor was there a significant interaction [$F(2, 74) < 1$, $\eta_p^2 = 0.02$, $MSE = 0.08$]. Unlike Experiment 1A, there was no change in the correlation between font-size and JOLs from the first list ($\gamma = 0.11$, $SE = 0.04$), to the second list ($\gamma = 0.04$, $SE = 0.04$), $F(1, 74) = 1.59$, $\eta_p^2 = 0.02$, $MSE = 0.08$, $p = 0.21$.

Table 2
Means and standard deviations for each cell of Experiment 1B.

Condition	Measure	List	Font-size	Mean	SD
Low-intensity (n = 30)	JOLs	1	Small	51.34	20.08
			Large	55.14	19.21
		2	Small	38.69	20.75
			Large	41.56	21.47
	Recall	1	Small	19.84	16.83
			Large	23.81	15.62
		2	Small	21.27	19.24
			Large	23.65	19.35
Medium-intensity (n = 21)	JOLs	1	Small	46.81	14.04
			Large	50.37	14.09
		2	Small	37.55	21.66
			Large	40.02	20.55
	Recall	1	Small	17.91	11.94
			Large	14.06	14.24
		2	Small	20.41	16.23
			Large	23.58	20.56
High-intensity (n = 27)	JOLs	1	Small	57.07	16.89
			Large	58.72	17.30
		2	Small	48.33	21.03
			Large	48.23	21.28
	Recall	1	Small	27.87	20.28
			Large	25.75	22.42
		2	Small	26.10	21.19
			Large	26.10	20.98

2.2.3. Recall

Similar to the findings from the JOL analyses regarding the instruction condition, recall showed neither a significant three-way interaction between instruction, font, and list [$F(2, 75) = 1.30$, $\eta_p^2 = 0.034$, $p = 0.278$]; nor a significant two-way interaction of instruction and font [$F(2, 75) = 2.17$, $\eta_p^2 = 0.055$, $MSE = 66.66$, $p = 0.121$]; nor a significant two-way interaction of instruction and list [$F(2, 75) = 1.31$, $\eta_p^2 = 0.034$, $MSE = 221.54$, $p = 0.276$]. Finally, participants recalled the same percentage of words whether they were in the no repeated instructions condition ($M = 22.14$, $SD = 17.83$), the repeated instructions condition ($M = 18.99$, $SD = 16.06$), or the feedback with instructions condition ($M = 26.46$, $SD = 21.23$), $F(2, 75) = 1.34$, $\eta_p^2 = 0.034$, $MSE = 1016.00$, $p = 0.269$. See Table 2 for the full set of means and standard deviations.

In regards to the font-size effects, the percentage of words recalled was consistent across large ($M = 23.03$, $SD = 19.24$) and small fonts ($M = 22.27$, $SD = 18.29$) [$F(1, 75) < 1$, $\eta_p^2 = 0.005$, $MSE = 66.66$] and did not interact with list, $F(1, 75) = 1.37$, $\eta_p^2 = 0.034$, $MSE = 88.18$, $p = 0.278$.

2.2.4. Post-task items

To ensure that the instructions and research were believed by participants, a manipulation check was included in the post-task questionnaire. When asked if they had believed the research presented in the instructions at the start of the experiment, 84.62% indicated that they definitely believed the research.

2.3. Discussion

Akin to Experiment 1A, the procedure did not result in a graded, strengthening effect across instruction condition. The lack of any effects across this manipulation is not particularly surprising given the similar results in Experiment 1A. However, the belief presented in this experiment is not a commonly pre-held belief about font-size and memory: for example, in Mueller et al. (2014) only 10 participants of 48 reported small-font words as more memorable. We expected that the novelty of the belief would result in varying levels of acceptance, as well as gradation across the intensity of the manipulation. Instead, participants showed a uniform font-size bias in their JOLs, indicating a similar change in behavior towards font-size across all instructional conditions. The lack of effect here may reflect some strong tendency of participants to accept counterintuitive research about their own memory, or alternatively may show that any effects are too variable to show fine-grained differences in the studied sample.

Performance across lists again showed a reduction in JOLs which brought them closer to actual recall performance. However, as in Experiment 1A the improved global calibration of JOLs was not qualified by a significant interaction effect between font-size and list, nor were there any improvements in resolution across lists in this experiment. As such, the difference between Experiments 1A and 1B appears to be solely due to the change in the initial instructions and any effects they may have had on participants' beliefs. Here, participants were under the impression that smaller words would be easier to remember than larger words, and though there was not a complete reversal of the font size effect, there is a definite effect of the opposing belief crippling the effect.

The fluency argument and the belief argument for JOLs make competing predictions for this experiment. For fluency, smaller words are less perceptually fluent than large words and thus no matter what information the experimenter gives the participant, JOLs should be larger for the more perceptually fluent words and should be correlated with font-size to some degree. Alternatively, belief that smaller words are more easily remembered should produce larger JOLs for words in small font and should be negatively correlated with font-size. The results show that presenting a belief that small font words are easily remembered nearly eliminates the font-size effect and a near-null correlation between font-size and JOLs ($\gamma = 0.04$), which suggest a very low reliance on font-size as a predictor of learning. Rhodes and Castel (2008) showed a similar reduction in the effect after instructing participants to actively discount perceptual fluency as a diagnostic cue for memory. Critically, here we show a dramatic reduction in the font-size effect even when participants are aware of and even primed to attend to font-size specifically. The fact that the effect is not eliminated possibly suggests a lingering effect of perceptual fluency on JOLs that is not easily overcome by changing one's beliefs.

3. Experiment 2

In Experiments 1A and 1B we attempted to manipulate participants' beliefs about font-size and memory and the strength of those beliefs and it was shown that participants' responses are not particularly affected by the attempts made to strengthen those beliefs. Unfortunately, these experiments lacked an appropriate control group that was not given any instruction regarding the effects of font-size on memory. It is thus unclear if the instructions themselves, regardless of the intensity manipulation, had any effect at all. Further, participants were all given reason to believe in font-size effects at the beginning of the study. If participants had given substantial effort to applying those beliefs over the course of a study list, it is not surprising that the instructional manipulations did not have a strong (or any) effect.

Importantly though, the patterns of data presented in 1A and 1B suggest that the content of the beliefs and their effects on behavior may be relatively malleable via instructional manipulations at the start of the study. That is, though people tend to show strong beliefs about font-size prior to starting studies like these (Kornell et al., 2011; Mueller et al., 2014), participants in Experiment 1B appeared to easily accept that words in small font might be easier to remember that words in large font. After being presented with this information, there was no effect of font-size on their JOLs though a sample size that large should have shown the standard effect under the fluency hypothesis.

In the previous experiments the instruction manipulation was performed prior to any baseline measurements regarding the font-size effect in the participant pool. That is, participants likely had a priori beliefs about the effects of font-size that could potentially confound the data (e.g. if participants in Experiment 1B had a strong belief that font-size does not affect learning at all), and further, participants may not have truly believed the information given to them by the researchers. To confirm that this nullification of the font-size effect was reliable, this experiment analyzes the belief-instruction manipulation against participants' non-biased (i.e. not influenced by the researchers) beliefs about font-size.

We expected that participants would all show the standard font-size effect on a first list, such that JOLs for large font words are higher than those given for small font words. We further expected that participants' JOLs on a second list would be similar across font-size if given the belief instructions from Experiment 1B or show only a slightly reduced effect of font-size if given the instructions from Experiment 1A. Even though processing fluency should remain constant across lists, it is hypothesized that the introduction of a counter-belief will have an additive effect reducing the differences across font-size caused by fluency and prior beliefs.

3.1. Method

The participants were 62 workers recruited through Amazon Mechanical Turk, who were paid $4.50 to participate. Participants completed two study-test sets where the font-size of the studied items was manipulated within-subjects along two levels (large, 48 pt.; small, 18 pt). These were the same word lists that were constructed for Experiment 1. Again, it should be noted that because these workers used their own computer and browser settings, the exact size of the font could not be controlled, but the ratio of the font sizes was preserved (48:18). Information about the effects of font-size was manipulated between subjects prior to the second list across two levels: participants were either informed that large or small words are easier to learn.

3.1.1. Procedure

After providing informed consent, participants were told that they would study words presented in different font sizes and that they should attempt to remember as many words as possible. Participants indicated their understanding of the instructions by answering a multiple choice reading check question before moving on. All study stimuli were presented in black, lowercase letters in the center of a white background on a computer screen. Immediately following the presentation of each word participants were instructed to rate their confidence that they would be able to recall that item on a scale from 0% (no chance of recall) to 100% (certain of recall). Participants were encouraged to use the entire range of the scale. Immediately following the study lists participants engaged in a 2-min distractor task where they played a game of Tetris. After this distractor, they were given 4 min to recall as many of the items as they could from the study list by typing them into a textbox.

After recall of the first list and before presentation of the second, participants were given the information regarding the effects of font-size on memory. The wording here was identical to Experiments 1A and 1B: "research has shown that, for college-age participants, words in [larger/smaller] fonts are easier to recall than words in [smaller/larger] fonts." Again, participants answered a multiple choice reading check question before moving on, and in this case the correct answer reiterated the font-size information briefly. The participants then studied the second list, played another Tetris distractor for 2 min, and finally recalled the second list for 4 min.

After completing the study, participants were debriefed and informed of the specific instances when font-size is likely to impact memory, e.g. the Von Restorff effect (Von Restorff, 1933), and also completed other questionnaire regarding how they study information and their beliefs about font-size and memory.

3.2. Results

3.2.1. JOLs

As shown in Figs. 3 and 4, which are the graphs for the instructed-large condition and the instructed-small conditions, it appears that the instruction manipulation produces different results depending on the instruction. The baselines in each condition (list one) have a similar pattern, as expected, but after the belief is introduced the JOLs move in different directions, such that in Fig. 3 we see a clear font size effect in the JOLs for list two and in Fig. 4 there are no differences between the JOLs in list two. This direction of change is made more apparent in Fig. 5, which shows the average difference in JOLs across list for each font-size. In the condition where participants were instructed that large words are easier to learn the font-size effect appears to grow: JOLs for large words do not change across lists but JOLs for small words decrease across lists. In the condition where participants were instructed that small words are sometimes easier to learn the opposite is found: JOLs for words in large font become smaller and JOLs for words in small font do not change. Thus, for the instruct-large condition, it appears that the JOLs for small words become even smaller, whereas for the instruct-small condition the JOLs for the large words are reduced due to the manipulation. To test these apparent effects, a 2 (instruction) × 2 (list) × 2 (font size) mixed ANOVA, where the belief

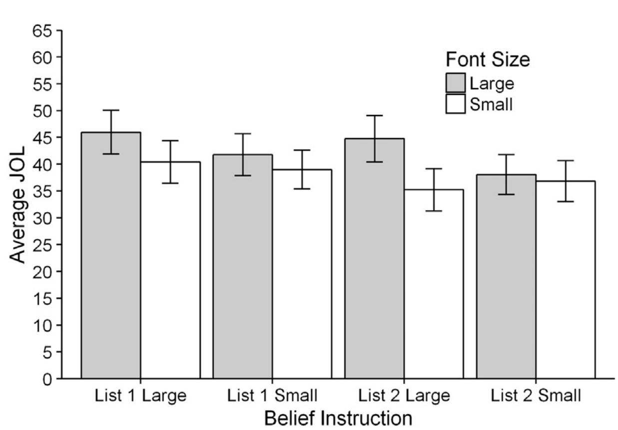

Fig. 3. Mean JOLs for large and small font words in Experiment 2, divided by list and the belief instruction condition ("Large" indicates that participants were instructed that larger words are easier to recall). Error bars represent standard errors of the mean.

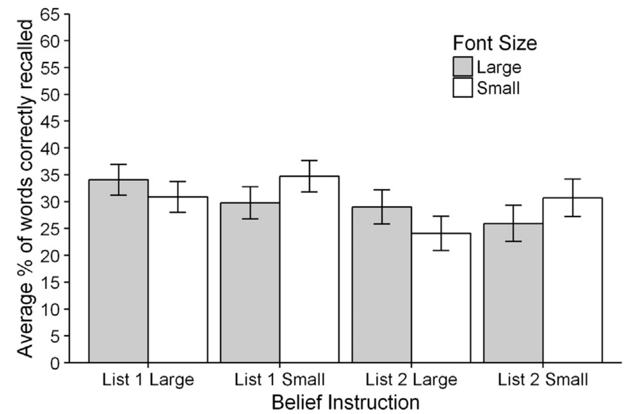

Fig. 4. Mean percentage of words correctly recalled for large and small font words in Experiment 2, divided by list and the belief instruction condition ("Large" indicates that participants were instructed that larger words are easier to recall). Error bars represent standard errors of the mean.

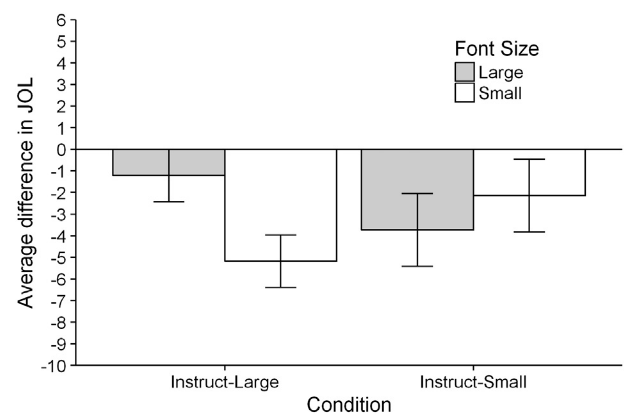

Fig. 5. The average difference of JOLs from list one to list two for words in large and small font in Experiment 2, divided by the instruction manipulation condition. Error bars represent standard errors of the mean.

instruction was between subjects, was performed. The three-way interaction between font size, list, and belief instruction was significant, $F(1, 60) = 4.73$, $\eta_p^2 = 0.073$, $MSE = 92.54$, $p = 0.034$.

The three-way interaction seems to be driven by the simple two-way interaction between font size and the critical instruction manipulation, $F(1, 60) = 4.76$, $\eta_p^2 = 0.073$, $MSE = 74.15$, $p = 0.033$. This interaction was such that in the condition where participants were given reason to believe words in large font are easier to learn, words in large font ($M = 46.75$, $SD = 21.55$) were given higher JOLs than words in small font ($M = 39.16$, $SD = 21.59$) [$t(61) = 4.43$, $d = 0.35$, $p < 0.001$], but this effect was reduced in the opposite belief condition where words in large font ($M = 41.30$, $SD = 20.99$) were given only slightly higher JOLs than those in small font ($M = 38.47$, $SD = 18.92$), $t(61) = 2.08$, $d = 0.14$, $p = 0.046$. The magnitude of JOLs was significantly different across the instruction manipulation such that words in large font were given larger JOLs in the condition where participants had reason to believe words in larger fonts are easier to learn [$t(61) = 2.99$, $d = 0.26$, $p = 0.004$] but there were no differences in the JOLs for small font across the instruction conditions [$t(61) = 0.37$, $d = 0.03$, $p = 0.72$].

The simple two-way interaction between the list and instruction manipulation was not significant [$F(1, 60) = 0.25$, $\eta_p^2 = 0.004$, $MSE = 96.26$, $p = 0.621$]. On average, participants gave similar magnitudes of JOLs regardless of whether they were in the believe-large condition ($M = 42.95$, $SD = 22.57$) or the believe-small condition ($M = 39.88$, $SD = 21.19$), $F(1, 60) = 0.35$, $\eta_p^2 = 0.006$, $MSE = 1654.5$, $p = 0.555$. On average participants gave lower JOLs to items on the second list (M = 40.06, SD = 21.86) than the first list (M = 42.78, SD = 21.84), $F(1, 60) = 4.80$, $\eta_p^2 = 0.074$, MSE = 96.26, $p = 0.032$. Finally, the general font size effect was shown such that words in large font (M = 44.02, SD = 22.50) were given larger JOLs than words in small font (M = 38.81 SD = 21.27), $F(1, 60) = 22.70$, $\eta_p^2 = 0.275$, MSE = 74.15, $p < 0.001$.

3.2.2. Gamma correlations

We computed Goodman-Kruskal γ correlations to analyze the effects of the independent variables on metacognitive resolution. Eight participants were excluded from these analyses because their correlations could not be computed due to invariance in either their JOLs or recall in one or more conditions.

Considering the effects of the instructional manipulation on metacognitive resolution, we found that participants were no more accurate with their judgments when given information that suggested words in large font were more memorable (γ = 0.37, SE = 0.09), than when given the same information about small-font words (γ = 0.32, SE = 0.08), $F(1, 51) < 1$, $\eta_p^2 < 0.001$, MSE = 0.28. Similarly, there were no significant effects of font-size on metacognitive resolution: large-font words (γ = 0.36, SE = 0.06) were judged just as accurately as small-font words (γ = 0.34, SE = 0.06), $F(1, 51) < 1$, $\eta_p^2 < 0.001$, MSE = 0.17. However, a marginally significant main effect of list was found such that judgments in the first list (γ = 0.40, SE = 0.05) were more accurate than judgments in the second list (γ = 0.29, SE = 0.07), $F(1, 51) = 3.54$, $\eta_p^2 = 0.065$, MSE = 0.14, $p = 0.07$.

Looking at the more complex effects of the list variable, there was neither a significant interaction with font-size [$F(1, 51) < 1$, $\eta_p^2 = 0.008$, MSE = 0.15], nor with condition [$F(1, 51) = 2.33$, $\eta_p^2 = 0.044$, MSE = 0.15, $p = 0.39$]; however, there was a marginally significant three-way interaction [$F(1, 51) = 3.24$, $\eta_p^2 = 0.060$, MSE = 0.15, $p = 0.08$].

Though the three-way interaction was only marginally significant, a brief exploratory suggests the interaction may be driven by differences in judgments about small-font words. Judgments for words in large font do not appear to differ across condition regardless of whether they are made in the first list [$t(26) = 1.02$, $d = 0.20$, $p = 0.32$] or the second list [$t(25) = 0.54$, $d = 0.11$, $p = 0.60$]. However, though accuracy for words in small-font in list one is better in the bigger-is-better condition (γ = 0.46, SE = 0.07) than the opposite (γ = 0.31, SE = 0.07) [$t(28) = 1.56$, $d = 0.29$, $p = 0.13$], this appears to be reversed on the second list, such that the bigger-is-better condition is less accurate (γ = 0.20, SE = 0.11) than the other (γ = 0.36, SE = 0.08), [$t(23) = -1.08$, $d = 0.22$, $p = 0.29$]. Further, this may be due to an increase in accuracy for small font words in the smaller-is-better condition from the first list (γ = 0.31, SE = 0.07) to the second (γ = 0.36, SE = 0.08) [$t(27) = -0.90$, $d = 0.17$, $p = 0.38$] coupled with a decrease in the same words in the bigger-is-better condition from the first list (γ = 0.46, SE = 0.07) to the second (γ = 0.31, SE = 0.07) [$t(25) = 2.059$, $d = 0.40$, $p = 0.05$].

Lastly, we again computed a two-way ANOVA comparing the Goodman-Kruskal gamma correlations across list and condition to assess changes in reliance on font-size as a predictor of learning. In this experiment, there were no changes in correlations across lists, $F(1, 56) < 1$, $\eta_p^2 = 0.003$, MSE = 0.11. However, there was a significant effect of condition, such that participants in the bigger-is-better condition (γ = 0.36, SE = 0.04) showed higher correlations than those in the smaller-is-better condition (γ = 0.14, SE = 0.04), $F(1, 56) = 10.97$, $\eta_p^2 = 0.16$, MSE = 0.12, $p = 0.002$. This main effect was driven by the two-way interaction, $F(1, 56) = 7.80$, $\eta_p^2 = 0.12$, MSE = 0.11, $p = 0.007$. Two dependent samples t-tests were run to determine the nature of the interaction. In the bigger-is-better condition, there was no difference from the first list (γ = 0.26, SE = 0.06) to the second list (γ = 0.21, SE = 0.06), $t(28) = 1.60$, $d = 0.15$, $p = 12$. However, in the smaller-is-better condition, there was a decrease from the first list (γ = 0.36, SE = 0.04) to the second list (γ = 0.36, SE = 0.04), $t(28) = 2.56$, $d = 1.05$, $p = 0.02$.

3.2.3. Recall

Again, referring to Figs. 3 and 4, the recall scores appear to have a slight decline across lists. Additionally, there appears to be a difference in the pattern of which font size was better recalled depending on the condition that the participant was in. Specifically, in the condition where participants were instructed that large words would be easier to recall, the small words were recalled less often on both lists. This pattern was reversed in the opposite belief condition. To test these apparent effects, a 2 (instruction) × 2 (list) × 2 (font size) mixed ANOVA, where the belief instruction was between subjects, was performed. The three-way interaction between font size, list, and belief instruction was not significant, $F(1, 60) = 0.11$, $\eta_p^2 = 0.002$, MSE = 84.14, $p = 0.740$.

The simple two-way interactions were non-significant for both the font size and list [$F(1, 60) < 1$, $\eta_p^2 = 0.002$, MSE = 84.14, $p = 0.691$] and the list and instruction pairings [$F(1, 60) = 0.43$, $\eta_p^2 = 0.007$, MSE = 146.66, $p = 0.517$], though the effects of font size do vary across the instruction variable, [$F(1, 60) = 12.62$, $\eta_p^2 = 0.174$, MSE = 97.70, $p < 0.001$]. In the condition where participants were instructed that large fonts are sometimes easier to learn, words in large font (M = 31.57, SD = 15.75) were recalled more often than words in small font (M = 27.49, SD = 15.82), $t(30) = 2.57$, $d = 0.26$, $p = 0.015$. In the opposite instruction condition, the recall scores were reversed and participants recalled a higher percentage of words in small font (M = 32.72, SD = 16.84) than words in large font (M = 27.88, SD = 16.84), $t(30) = 2.49$, $d = 0.29$, $p = 0.019$.

Participants recalled similar amounts of items whether they were in the condition where they believed larger words were easier to learn (M = 29.53, SD = 15.78) or the condition where they believed smaller words were easier to learn (M = 30.30, SD = 16.69), $F(1, 60) = 0.05$, $\eta_p^2 = 0.001$, MSE = 726.53, $p = 0.822$. Lastly, participants recalled a higher percentage of items on average on the first list (M = 32.38, SD = 15.25) than the second list (M = 27.46, SD = 17.24), $F(1, 60) = 10.24$, $\eta_p^2 = 0.146$, MSE = 146.66, $p = 0.002$.

3.2.4. Post-task items

Again, to ensure that the instructions and research were believed by participants, a manipulation check was included in the post-task questionnaire. When asked if they had believed the research presented in the instructions at the start of the experiment, 70.97% indicated that they definitely believed the research. Considering each instructional manipulation separately, 83.87% of participants that received instructions about small-fonts providing memory benefits reported believing the research compared to only 58.06% of participants in the opposite condition. Also, when asked about their a priori beliefs about font-size 62.90% indicated an a priori belief that large font is more memorable, 8.06% small font, and 29.03% no bias.

3.3. Discussion

In this experiment, participants viewed words in large and small font and made JOLs about these words before attempting to recall them. As expected, when participants were given no instruction other than to judge the likelihood that they would recall the information later, JOLs were larger for words in large font than words in small font, either because the fonts caused a difference in processing fluency (Rhodes & Castel, 2008) or prior beliefs about font-size (Mueller et al., 2014).

We do note that there were slight differences in recall scores in this experiment that differ from prior work with the font-size effect. These differences are puzzling and may be due to a slight demand characteristic of the instruction manipulation. That is, it is possible that participants gave differential levels of effort to memorizing words printed in the "easier" and "harder" font-sizes. This explanation falls short when considering that these differences were not found in

Experiment 1A or 1B, which had a similar manipulation. Further, the difference is present in the recall for the first list, at which point both groups had gone through the same exact procedure. We conclude that these differences must be due to some random noise in the data, and as they are small differences on the order of 1–2 words, we believe they do not present any complications when considering the more central JOL effects.

Though the processing fluency did not differ between lists one and two of this experiment, there was a significant effect of the instructions given before the second list. For participants that were given information suggesting that words in large font would be easier to recall than those in small font, the font-size bias in JOLs was shows for words in large font across both lists. In contrast, for participants who were given information suggesting that words in small font would be easier to recall, there was only an effect of font-size on JOLs on the first list but no difference in JOLs on the second list. It is hypothesized that this counter-belief, counter in that it is the opposite of what participants have indicated they believe prior to these types of experiments (Kornell et al., 2011; Mueller et al., 2014), is competing with the effects of processing fluency. We suggest that when participants see a large font word and give a JOL, they are giving the sum of the effects of fluency (positive) and belief (negative). This is supported by the results in the condition where the information given was in concert with prior beliefs that large font words are easier to learn and participants showed a strong font-size effect on their JOLs.

To reiterate a point from the discussion of Experiment 1B, a pure-belief driven account of JOLs would predict that if participants believe the instructional manipulation and update their JOLs accordingly, JOLs and font-size would have a strong positive correlation in the bigger-is-better condition and a strong negative correlation in the smaller-is-better condition. Here, participants all showed a small-to-moderate correlation between JOLs and font-size on list one which is concordant with the font-size bias. Participants in the bigger-is-better retained this bias at list two, but participants in the smaller-is-better condition showed a negligible correlation. This shift in correlations suggests that participants have changed their beliefs about the effects of font-size on learning. There is no reversal to a negative correlation which leaves three possible interpretations: participants discarded their beliefs about font-size entirely, participants changed their beliefs and failed to fully updated their judgments (similar to Mueller et al., 2015), or they changed their beliefs and those beliefs are now in competition with perceptual fluency.

One important consideration regarding the critical manipulation of the instructions is that there was no independent measure of the impact of instructions. Conclusions can only be made regarding the differences in the instruction sets, and readers should be cautious in making assertions about the impact of instructions alone. Future studies concerning paradigms of this sort should consider a control group that does not receive any instructions. This control group would allow for a cleaner analysis of metacognitive changes as a result of both task experience and instruction set content.[1]

In Experiments 1A and 1B participants did not show any changes in metacognitive resolution scores across any variable. An increase in resolution would reflect either changes in JOL production leading to more accurate judgments, similar JOL production along with changes in recall that better reflect judgments, or some combination of both. Overall in this experiment there were no significant changes in JOLs across any of the variables, though it may be of interest to consider the marginal trends in the data.

There was a trend towards decreased metacognitive resolution across lists, which is probably best explained by the apparent reduction in resolution from list one to list two for participants in the smaller-is-better condition. These participants gave less accurate JOLs for words in small font on list two than they did on list one. This is in contrast with their performance on large-font words, and with participants in the opposing condition, both of which showed no trends. The reduced resolution was accompanied by decreased JOLs for words in small font from list one to list two, along with decreased recall performance for words in small font on list two. Together, these trends may suggest a tendency for participants to give less effort to memorizing words in small-font because they believe they will be easier to recall. If so, this would not only support the notion that beliefs about font-size influence JOLs but that they also influence the memory task performance.

Concerning the relevant manipulation checks from the post-task questionnaire, it is somewhat surprising to see that only ~58% of participants reported believing the instructional manipulation when the research presented indicated that large-fonts are more memorable. These participants produced a font-size effect in their data, nonetheless, suggesting that they were unaware of how the font-size was affecting their judgments (consistent with findings from Frank & Kuhlmann, 2017). This contrasts with the high acceptance of the instructional manipulation in the opposing condition (~84%). It is possible that this large difference in acceptance rates is due to the difference in the relationship between the presented information and prior beliefs about font-size and memory. That is, we speculate that participants are more likely to accept research that directly contrasts with their own beliefs.

4. General discussion

The literature strongly supports an effect of font-size on JOLs which is not reflected in recall, and this effect is replicated here in several experiments. The mechanism driving this effect is in debate currently and there are two competing hypotheses. The fluency hypothesis argues that increased processing fluency, which is inherent in more perceptually fluent items, leads to a more familiar, more accessible representation in working memory. A more accessible representation can lead participants to judge more perceptually fluent items as better remembered, though their judgments are often wrong because perceptual fluency is not necessarily a useful cue. The belief hypothesis argues that participants make their judgments based almost solely on belief, and that it is belief about large font words, and larger font in general, which is driving the font-size effect. To test these hypotheses against one another, the experiments presented here attempted to manipulate belief while holding fluency constant. In this way, any differences in JOLs would be driven by belief.

In Experiment 1A, participants were presented with evidence confirming the font-size effect before study began and between the first and second study list. It was expected the instruction manipulation would strengthen participants' pre-existing beliefs in the font-size effect and that their JOLs would reflect this strengthened belief. While the basic finding was replicated and large-font words were rated as more memorable than small font words, there were no differences in the effect across instruction. These findings suggest that either manipulation did not strengthen the belief, the belief was not strengthened appreciably, or the font-size effect is not mediated by belief. Taken alone, the results of Experiment 1A provide no direct confirmation or refutation of either the fluency or belief hypotheses but do show that the effect persists across conditions.

When presented with the information that small-font words are more easily remembered than large-font words (1B), participants showed a markedly reduced difference between their JOLs for words in large-font and words in small-font. Importantly, it was only the introduction of an incompatible belief that led participants to give similar JOLs to items in different font-size, a behavior which has not been

[1] We attempted to run a version of Experiment 2 with this no-instructions control condition twice during the development of this study. There were issues replicating the font-size effect in all groups despite the effect being relatively robust in the literature (see Luna et al., 2017 for a meta-analysis). The samples obtained had comparable recall scores and average JOLs with respect to the other studies reported here, however there was no systematic bias towards large-fonts in JOLs for all groups. Interested readers should contact the authors for more information.

observed in past studies. However, the effects found in Experiment 1B do not necessarily refute the fluency hypothesis.

If JOLs were driven entirely by belief, participants who believed that words in smaller font should have given those items larger JOLs. In the raw data for Experiment 1B, a negative difference between large and small font words was only found in 3 of the 79 participants; it was much more common to see only slight one-to-two-point differences in JOLs between the font-sizes. As these results further showed, participants did not systematically alter JOLs according to font-sizes, suggesting that perceptual fluency may still be significantly influencing JOLs. JOL construction might be represented in summation sense where base memory, (e.g. if an item has repeated presentations or longer presentation times), item effects (e.g. familiarity, word frequency, or personal relevance), fluency, and beliefs are related:

$$[\text{JOL}] = [\text{base memory} + \text{item effects}] + [\text{fluency}] + [\text{beliefs}],$$

where beliefs would encompass beliefs about memory capacity, the effects of font-size, etc. In the present study, base memory and item effects were not manipulated, but prior studies have shown direct effects of these factors (cf. Rhodes, 2015). In this model, any negative belief would counter the positive effects of fluency, and any positive effects of fluency would mitigate negative beliefs. This equation is, of course, likely an over-simplification, but represents an attempt to better articulate the multiple bases for JOLs, how they interact, and can be altered. Additionally, it could be argued that the current set of studies does not unambiguously show a direct effect of belief on JOLs; though this study shows some evidence that belief has a tangible effect on metamemory and previous studies have shown that there must be some effects of belief, they may only be mediated effects or otherwise indirect. What is clear though is that belief is in some way affecting JOLs, and when beliefs are changed or challenged JOLs similarly change. This is consistent with recent research showing that participants' JOLs are affected by novel beliefs that were not held prior to the experiment (Mueller & Dunlosky, 2017).

A counterargument to this hypothesis is that participants are simply exhibiting demand characteristics. That is, because the instructions include information about the relative sizes of the words in the subsequent lists, participants are catering to the experimenters' expectations when making their judgments. It is difficult to disentangle any demand characteristics in these data, however there is some evidence to show that participants are responding in earnest. Namely, there was no reversal of the font-size effect—that is, if participants were responding entirely based on demand characteristics we should expect a pattern in Experiments 1B and 2 such that smaller words are being rated as more memorable. Instead the effect is null. Further, if participants *are* guiding their beliefs using the information given in the instruction sets, this is simply further evidence that belief is a main component in the formation of JOLs. This new information *should* be used to inform JOLs as it is presumably more diagnostic (as a piece of researcher coming from a scientist) than other naïve heuristics like perceptual fluency. Again, this pattern of responding is not fully endorsed as we do not see a reversal of the font-size bias.

It appears as though the font-size effect is not being driven solely by belief or solely by fluency, but rather that there are additive effects of both factors. Past research has strongly suggested that fluency must be playing a role in metacognitive monitoring (Baddeley & Longman, 1978; Begg et al., 1989; Benjamin, 2003; Koriat & Bjork, 2005; Koriat & Ma'ayan, 2005; Kelley & Rhodes, 2002; Rhodes & Castel, 2008, 2009). Mueller et al. (2014) challenges whether it is in fact fluency that is mediating those effects or merely beliefs about fluency. The current study seems to suggest that while beliefs play a large role, there is still a residual effect of fluency. However, it remains unclear whether fluency is a direct factor or if it is instead beliefs about fluency driving these effects. That is, the influence of perceptual fluency may be due to a direct impact of fluency, or alternatively, it may be mediated by participants' beliefs about fluency with little to no direct effect of fluency. It is plausible that that instead of fluency, there are competing beliefs about how font-size may affect memory: prior beliefs about positive effects of large fonts on memory (Kornell et al., 2011) and the information about font-size and memory given by the instruction sets in these experiments.

Additionally, participants may simply not be fully implementing any new beliefs. They may fully believe that the smaller font words are easier to remember but not fully adapt their heuristics in such a short time. This line of reasoning resonates with recent research showing that participants do not fully-update their strategies for the formation of JOLs after learning from their experiences (Mueller et al., 2015; Tullis, Finley, & Benjamin, 2013). The aforementioned research (Mueller et al., 2015) also highlights the effects of scaling artifacts over time, which are likely present to some extent in this experiment as well—as JOLs drop on list 2 the effects of belief may be hidden. A lack of complete knowledge updating coupled with the possible presence of these kinds of scaling artifacts certainly casts doubt on any direct effects of perceptual fluency reported here.

If it were entirely the effects of belief, the effects of the currently study might simply be explained by altered, and not replaced beliefs: participants may not have switched over to the new counter-belief presented to them, rather they partially adjusted current beliefs. The current set of experiments cannot speak to this alternate explanation directly, though we speculate that the strong acceptance of the belief-instruction in Experiment 1B (~84%) coupled with the lack of reversal of the font-size effect is compelling evidence. Despite our speculations, to fully understand whether this is a direct effect of perceptual fluency, it would be useful to have clear measures of perceptual and processing fluencies which would allow for the evaluation of their contributions. Without an independent measure of fluency, it is difficult to make strong claims as to whether these effects regarding fluency are distinctly different from other research suggesting they are instead effects of beliefs about fluency (Mueller et al., 2014).

Another approach to clarify this ambiguity is to focus on some key individual differences which are known to affect beliefs. For example, Miele et al. (2011) showed that differences in theory of intelligence (Dweck, 2000) predict differences in whether or not the "easily learned; easily remembered" (ELER) heuristic (Koriat, 2008) is used. Dweck (2000) has shown that there are different views on how intelligence can be increased or decreased. Incremental theorists, or growth theorists, have the view that a person can work hard and improve one's own intelligence. Fixed theorists on the other hand feel that intelligence is something you are born with and it cannot be changed. Fixed theorists tend to be more susceptible to using heuristics, and Miele et al. (2011) has specifically shown that fixed theorists are more likely to use the ELER heuristic.

Heuristics, though more or less automatic processes, reflect participants' beliefs about certain situations. In theory, individual differences on theory of intelligence should yield differential effects of belief such that fixed theorists show much greater effects of belief than incremental theorists. If the effects of fluency are in truth effects due to beliefs about fluency, people who are closest to the incremental end of the theory of intelligence scale would likely display minimal differences between *any* factors, be they belief or fluency.

In conclusion, the perceptual characteristics of words certainly have an effect on how people judge their learning which is evident across many studies (see Luna et al., 2017 for a meta-analysis of these effects). In these experiments, we believe that the results show both belief and fluency to be strong cues utilized by people when monitoring their learning. Whether it is perceptual fluency or beliefs about fluency that is a prime mover in the development of JOLs is uncertain and may be hard to entirely decouple. Yet this study does show that belief and competition of belief about something as simple as font-size can have a direct and observable effect on an individual's perceived level of learning.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. http://dx.doi.org/10.1177/1088868309341564.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*(5), 610–632. http://dx.doi.org/10.1016/0749-596X(89)90016-8.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. http://dx.doi.org/10.3758/BF03194388.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55–68. http://dx.doi.org/10.1037/0096-3445.127.1.55.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition, 41*(6), 897–903. http://dx.doi.org/10.3758/s13421-013-0307-8.

Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 429–440. http://dx.doi.org/10.1037/a0034407.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36*(2), 429–437. http://dx.doi.org/10.3758/MC.36.2.429.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development.* Philadelphia, PA: Taylor & Francis.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*(1), 19–34. http://dx.doi.org/10.1016/j.jml.2007.03.006.

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693. http://dx.doi.org/10.1037/xlm0000332.

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. *Psychology of Learning and Motivation, 41*, 293–320. http://dx.doi.org/10.1016/S0079-7421(02)80010-X.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. http://dx.doi.org/10.1037//0096-3445.126.4.349.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36*(2), 416–428. http://dx.doi.org/10.3758/MC.36.2.416.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187–194. http://dx.doi.org/10.1037/0278-7393.31.2.187.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1133–1145. http://dx.doi.org/10.1037/0278-7393.32.5.1133.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*(4), 478–492. http://dx.doi.org/10.1016/j.jml.2005.01.001.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*(4), 643–656. http://dx.doi.org/10.1037/0096-3445.133.4.643.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability Bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*(6), 787–794. http://dx.doi.org/10.1177/0956797611407929.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Dartmouth Publishing Group.

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2017). Do delayed judgments of learning reduce metamemory illusions? A meta-analysis. *The Quarterly Journal of Experimental Psychology,* 1–11. http://dx.doi.org/10.1080/17470218.2017.1343362.

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered?: It depends on your beliefs about intelligence. *Psychological Science, 22*(3), 320–324. http://dx.doi.org/10.1177/0956797610397954.

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245–258. http://dx.doi.org/10.1016/j.jml.2016.10.008.

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2015). Why is knowledge updating after task experience incomplete? Contributions of encoding experience, scaling artifact, and inferential deficit. *Memory & Cognition, 43*(2), 180–192. http://dx.doi.org/10.3758/s13421-014-0474-2.

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. http://dx.doi.org/10.1016/j.jml.2013.09.007.

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review, 20*(2), 378–384. http://dx.doi.org/10.3758/s13423-012-0343-6.

Murayama, K., Blake, A. B., Kerr, T., & Castel, A. D. (2016). When enough is not enough: Information overload and metacognitive decisions to stop studying information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 914–924. http://dx.doi.org/10.1037/xlm0000213.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133. http://dx.doi.org/10.1037/0033-2909.95.1.109.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*(4), 207–213. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00502.x.

Rhodes, M. G. (2015). Judgments of learning: Methods, data, and theory. In J. Dunlosky, & S. K. Tauber (Eds.). *Oxford handbook of metamemory* (pp. 1–46). New York: Oxford University Press.

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. http://dx.doi.org/10.1037/a0013684.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*(3), 550–554. http://dx.doi.org/10.3758/PBR.16.3.550.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6*(5), 132–137. http://dx.doi.org/10.1111/1467-8721.ep10772899.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory and Cognition, 41*(3), 429–442. http://dx.doi.org/10.3758/s13421-012-0274-5.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung, 18*(1), 299–342. http://dx.doi.org/10.1007/bf02409636.

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—And is not—A desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition, 41*(2), 229–241. http://dx.doi.org/10.3758/s13421-012-0255-8.

Zaragoza, M. S., Payment, K. E., Ackil, J. K., Drivdahl, S. B., & Beck, M. (2001). Interviewing witnesses: Forced confabulation and confirmatory feedback increase false memories. *Psychological Science, 12*(6), 473–477. http://dx.doi.org/10.1111/1467-9280.00388.