



The font size effect depends on inter-item relation

Minyu Chang¹ · C. J. Brainerd²

Accepted: 22 March 2023 / Published online: 30 March 2023
© The Psychonomic Society, Inc. 2023

Abstract

The font size effect refers to the metacognitive illusion that larger fonts lead to higher judgments of learning (JOLs) but not better recall. Prior studies demonstrated robust JOL effects of font size under conditions of intra-item relation (i.e., cue–target relatedness within a word pair), even though intra-item relation is a more diagnostic cue than font size. However, it remains an open question whether the JOL effects of font size persist under conditions of inter-item relation (i.e., relations across items on a single-word list). In the current study, we examined the JOL and recall effects of font size when font size and inter-item relation were factorially manipulated in three JOL-recall experiments. Additionally, to manipulate the salience of inter-item relation, we presented related and unrelated lists in a blocked manner in Experiment 1 but in a mixed manner in Experiments 2 and 3. Our results showed that the JOL effects of font size are moderated or eliminated when inter-item relation is manipulated simultaneously with font size. Moreover, the smaller font led to better recall for related lists but not for unrelated lists across all three experiments. Therefore, our results demonstrate that individual cues may not be integrated with equal weight, and there can be a trade-off between item-specific and relational processing during the JOL process. Additionally, highlighting key information with larger fonts may not be optimal with related items.

Keywords Judgments of learning · Font size effect · Memory · Item-specific processing · Relational processing

The font size effect is a widely studied metamemory illusion, in which larger fonts elicit higher judgments of learning (JOLs; i.e., self-estimates of one's future memory) but not better memory performance than smaller fonts (Rhodes & Castel, 2008). This effect has been replicated in many experiments, mostly with unrelated word lists (e.g., Halamish, 2018; Hu et al., 2016; Kornell et al., 2011; Mueller et al., 2014; Rhodes & Castel, 2008; Susser et al., 2013; Tatz & Peynircioğlu, 2020; Undorf & Zimdahl, 2019; Yang et al., 2018), and occasionally with related or unrelated word pairs (e.g., Chen et al., 2019; Double, 2019; Luna et al., 2018; Price & Harrison, 2017; Rhodes & Castel, 2008; Wang et al., 2020). A recent meta-analysis of the font size effect (Chang & Brainerd, 2022) pointed out that related word lists were so far a missing type of study materials in this line of research. This type of material, however, can provide important theoretical and empirical

implications for the font size effect. Whereas prior studies showed that the JOL effects of font size are not mitigated by other item-specific cues like intra-item relation (e.g., Price & Harrison, 2017; Rhodes & Castel, 2008), it is unknown whether the same pattern holds for relational cues like inter-item relation. Thus, the use of related lists can provide unique insights into how people integrate item-specific cues like font size versus relational cues like inter-item relation in making JOLs. Also, because to-be-remembered information is often embedded in a web of semantic relations in real life, it is critical to investigate the font size effect under similar conditions.

Given that related lists can provide instructive findings about the font size effect while they have not yet figured in this line of research, the current study aimed to fill this gap by examining the JOL and recall effects of font size with both related and unrelated word lists. Below, we first highlight the difference between *intra-item relations* and *inter-item relations* in metamemory research. For the reader's convenience, we will refer to them as *pair relatedness* and *list relatedness* in the following text, respectively. Then, we delineate some theoretical hypotheses about the JOL and recall effects of font size under conditions of list relatedness. Last, we briefly explain the three experiments in the current study.

✉ Minyu Chang
minyu.chang@mcgill.ca

¹ Department of Psychology, McGill University, 2001 Avenue McGill College, Montréal, QC H3A 1G1, Canada

² Department of Psychology and Human Neuroscience Institute, Cornell University, Montréal, Canada

Pair relatedness versus list relatedness in metamemory research

In the metamemory literature, pair relatedness (i.e., the relation between cue and target within a word pair) is a well-established cue that affects both JOLs and recall (e.g., Koriat, 1997; Mueller et al., 2013; Undorf & Erdfelder, 2015). Related word pairs (e.g., *duke–prince*) robustly elicit higher JOLs and better recall than unrelated word pairs (e.g., *brush–cube*). The present study targeted a different type of item relatedness, which is relations among the items on lists of single words. We refer to this as list relatedness (as opposed to pair relatedness).

According to Koriat's (1997) cue-utilization account of JOLs, list relatedness and pair relatedness are distinct types of cues. In this account, people base their JOLs on three types of cues: intrinsic cues that pertain to the characteristics of study items, extrinsic cues that pertain to study conditions and encoding operations, and mnemonic cues that pertain to subjective encoding experience. As a word pair is a composite item during encoding, the relation between the cue and target *within* a word pair is a classic intrinsic cue. However, relations among items of a single-word list would be extrinsic cues, because they stimulate relational processing *across* items, which is an encoding operation applied by learners (Matvey et al., 2006).

Although the JOL effects of list relatedness have been less frequently studied than that of pair relatedness, it has been found that list relatedness affects JOLs, too. For instance, in Matvey et al.'s (2006) study, participants made item-by-item JOLs for categorized lists and unrelated lists, followed by free recall. It was found that JOLs were higher for categorized lists than for unrelated lists, although the effect was smaller than that of pair relatedness. Meanwhile, list relatedness, like pair relatedness, also affects recall accuracy (see also Kausler, 1974; Mandler, 1967), and the effect of list relatedness on actual recall is larger than its effect on JOLs. Moreover, Hourihan and Tullis (2015) showed that JOLs were sensitive to category size: JOLs were higher for categorized lists with 12 exemplars than those with four exemplars, tracking the pattern in the actual recall. In brief, there is an important theoretical difference between pair relatedness and list relatedness, and prior studies provide some evidence that list relatedness, like pair relatedness, also affects both JOLs and recall.

Do the JOL effects of font size persist under conditions of list relatedness?

A few studies show that the font size effect occurs for both related and unrelated word pairs (Price & Harrison, 2017; Rhodes & Castel, 2008; Wang et al., 2020). In particular, Price and Harrison (2017) and Rhodes and Castel (2008)

found that when font size and pair relatedness were factorially manipulated, JOLs were higher for both the larger fonts and the related word pairs. Moreover, the effects of font size and pair relatedness did not interact. Notably, pair relatedness is a more diagnostic cue than font size: Related word pairs robustly elicit both higher JOLs and better recall than unrelated pairs (e.g., Koriat, 1997; Mueller et al., 2013; Undorf & Erdfelder, 2015), whereas larger fonts only increase JOLs relative to smaller fonts.¹ Thus, these results show that the JOL effects of font size remain robust when pair relatedness is present, even though font size is a less diagnostic cue than pair relatedness.

However, whether the pattern is the same with list relatedness remains an open question. On the one hand, it is possible that the JOL effects of font size will not be mitigated by list relatedness. As mentioned above, according to Koriat's (1997) cue-utilization account, pair relatedness is an intrinsic cue whereas list relatedness is an extrinsic cue. Because extrinsic cues usually have weaker influences on JOLs (Koriat, 1997), it is reasonable to expect that the effects of font size on JOLs, which are not modified by pair relatedness (an intrinsic cue), would also not be modified by list relatedness (an extrinsic cue). Additionally, the cue-integration framework (Peynircioğlu & Tatz, 2019; Undorf & Bröder, 2020; Undorf et al., 2018) also predicts that the font size effect will persist under conditions of list relatedness. According to this account, people simultaneously process and integrate multiple cues when making JOLs, and hence, JOLs are sensitive to multiple cues if these cues affect JOLs when manipulated in isolation. In support of this prediction, prior studies have shown that the effect of font size remains robust when other, more diagnostic, cues (e.g., word frequency, word concreteness, pair relatedness) are provided along with font size (Fan et al., 2021; Price & Harrison, 2017; Rhodes & Castel, 2008; Undorf et al., 2018).

On the other hand, there are other reasons to expect that list relatedness might influence the JOL effects of font size that are rooted in the nature of encoding. Here, it is critical to pay attention to the fact that pair relatedness and font size are both item-specific cues that are processed with respect to individual list items, whereas list relatedness is processed *across* individual list items. In the memory literature, there is often a trade-off between those two types of processing, which are usually referred to as item-specific and relational encoding (e.g., Hunt & Einstein, 1981). Therefore, list relatedness shifts encoding in the direction of relational processing, which may dilute the effects of item-specific perceptual

¹ Chang and Brainerd's (2022) meta-analysis showed that there was actually a small font size effect on memory, as larger fonts slightly increased memory when pooled over 103 experiments ($g = .08$). However, this effect was usually not reliable in single experiments.

features such as font size. To sum up, there are contrasting theoretical predictions for the effect of font size on JOLs when list relatedness is an available cue: (a) It may still persist because list relatedness is an extrinsic cue that usually has weaker JOL effects than intrinsic cues, and the cue-integration framework predicts that people can simultaneously integrate multiple cues when making JOLs; or (b) it may be reduced by list relatedness, as the relational processing provoked by list relatedness may attenuate the processing of item-specific perceptual features.

Does font size affect recall with related lists?

Although many prior experiments have shown that font size has little-to-no effects on recall, it is important to point out that prior font size experiments relied overwhelmingly on unrelated word lists (e.g., Kornell et al., 2011; Mueller et al., 2014; Rhodes & Castel, 2008; Susser et al., 2013). Thus, these studies leave open the question of whether there is a memory effect (as opposed to a JOL effect) of font size with related word lists.

In the memory literature, some findings suggest that there may indeed be such an effect (Arndt & Reder, 2003; Brainerd et al., 2002). For instance, Arndt and Reder (2003) used different font types for the words on Deese–Roediger–McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) lists. DRM lists consist of a series of words (e.g., *bed, pillow, nap*) that are forward associates of a common unrepresented word (the critical distractor; e.g., *sleep*). Thus, DRM lists tend to elicit strong relational encoding, which often leads participants to falsely remember the critical distractor as being on the list. Arndt and Reder observed that presenting DRM list words in different fonts reduced false recognition of critical distractors. This seems to demonstrate that the different perceptual features associated with fonts enhance item-specific processing, which is complementary to the relational processing provoked by the high level of list relatedness.

In that connection, Hunt and his colleagues demonstrated that high levels of engagement of *both* item-specific and relational processing are most beneficial for memory performance (e.g., Hunt & Einstein, 1981; Hunt & Seta, 1984). Therefore, it is possible that font size manipulations have similar effects as font type manipulations, which can strengthen item-specific processing with related lists and thus produce better memory. Note that if font size manipulations did enhance item-specific processing, such effects were less likely to benefit recall for unrelated lists. This is because while related lists favor relational processing, unrelated lists tend to provoke more item-specific processing than relational processing. Thus, any item processing enhancement induced by font size may be redundant with unrelated lists.

The current study

As discussed above, list relatedness is a distinct type of cue from pair relatedness, and it has not yet figured in research on the font size effect. Moreover, there are contrasting theoretical predictions about whether font size affects JOLs when list relatedness is simultaneously manipulated, and it also remains an open question whether font size affects recall with related lists. The present study was designed to resolve those uncertainties by testing (a) whether the JOL effects of font size are robust under conditions of list relatedness, and (b) whether font size affects recall for related lists. To achieve these goals, we factorially manipulated list relatedness and font size in three free recall experiments. For the font size manipulation, we presented list words in either 18-pt font or 48-pt font. For the list relatedness manipulation, we used both related lists (categorized lists in Experiments 1 and 2 and DRM lists in Experiment 3) and unrelated lists. To vary the salience of list relatedness, we presented related and unrelated lists in a blocked manner in Experiment 1 and in a mixed manner in Experiments 2 and 3.

Experiment 1

Method

Participants

Fifty-one undergraduates ($M_{age} = 19.90$, $SD_{age} = 1.97$) participated in Experiment 1. All participants were compensated with extra course credits. We conducted a sensitivity analysis with G*Power 3.1 (Faul et al., 2007), which suggested that our sample size provided sufficient power ($1 - \beta > .80$) to detect a small-sized ($\eta_p^2 = .04$) main effect of font size or interactions between font size and list relatedness in mixed analyses of variance (ANOVAs).

Materials

The materials were 64 English vocabulary words, evenly divided into four 16-word lists. Two of the four lists were categorized, with each list including a category label (e.g., *animal*) and 15 categorical exemplars (e.g., *squirrel, cat, elephant*) of the given category. The categorized lists were retrieved from Van Overschelde et al.'s (2004) category norms. We used 16-item lists because the four-word categorized lists in Matvey et al. (2006) produced relatively weak effects on JOLs, and Hourihan and Tullis (2015) showed that longer lists exhibited stronger effects. Additionally, we presented the category label as the first word, so as to provoke strong relational processing. The

other two lists were semantically unrelated and consisted of words randomly selected from unused categories in Van Overschelde et al. (2004). Within each list, half of the words were presented in 18-pt Arial font whereas the other half were presented in 48-pt Arial font. We controlled word length, word frequency (Durda & Buchanan, 2006), and word concreteness (Brysbaert et al., 2014) between related and unrelated lists and between 18-pt and 48-pt words. The word lists used in all experiments and the relevant attribute values for the lists are available online (<https://osf.io/7j2pv>).

Procedure

The experiment was constructed in Qualtrics. Each participant studied four 16-word lists, evenly distributed across two blocks, with two related lists in one block and two unrelated lists in the other. The order of the blocks and the order of lists within each block were counterbalanced between participants.

The experiment consists of two blocks, with the second block immediately following the first. The two blocks followed the same study-buffer-test sequence, and they only differed in the word lists presented during the study phase. At the beginning of each block, participants were informed that they would study two word lists, and they would make a JOL for each word after it was presented. They were also informed that they would take a memory test after studying the two lists. During the study phase, all the words were presented at a 2-s rate, and there was a 5-s interval between consecutive lists. After each word was presented for 2 seconds, the word disappeared and participants were asked to make a JOL—namely, to judge the likelihood that they would be able to recall it later on a 0–100 scale (0 = *not likely at all*, 100 = *very likely*). They were required to type a whole number between 0 and 100 and were encouraged to use the full scale. There was no time limit for making JOLs. In each block, after the two word lists were studied, participants worked on a buffer task (simple mathematical problems) for 2 minutes. Then, they were given 3 minutes to complete a free recall test. No feedback was provided for the recall test.

Results

In this and the following experiments, we first report the results for JOLs and then for recall. Although JOL resolution is not of primary interest in the current study, we still reported those results for archival purposes. Here, JOL resolution indicates whether items that are recalled with higher probabilities are rated with higher JOLs, which is indexed by Goodman–Kruskal gamma correlations

between JOLs and recall. All the ANOVAs were conducted with the afex package in R (Singmann et al., 2015), and the Goodman–Kruskal gamma correlations were calculated using the Hmisc package in R (Harrell, 2019). Outliers were identified and removed if they were 1.5 interquartile ranges (IQRs) above or below the median (Höhne & Schlosser, 2018). Based on this criterion, no participant was excluded in Experiment 1.

JOL results

The descriptive results for JOLs in Experiment 1 are presented in Table 1. The JOLs were submitted to a 2 (list relatedness: related, unrelated) \times 2 (font size: 18 pt, 48 pt) \times 2 (block order: related/unrelated, unrelated/related) mixed ANOVA. The ANOVA revealed a main effect of list relatedness, $F(1, 49) = 28.64$, $MSE = 207.00$, $\eta_p^2 = .37$, $p < .001$, and a main effect of font size, $F(1, 49) = 8.47$, $MSE = 22.59$, $\eta_p^2 = .13$, $p = .005$, as JOLs were overall higher for the related lists and the larger font. There was no interaction between list relatedness and font size, $F(1, 49) = .00$, $MSE = 17.75$, $\eta_p^2 = .00$, $p = .978$.

In addition, there was a List Relatedness \times Block Order interaction, $F(1, 49) = 9.49$, $MSE = 207.00$, $\eta_p^2 = .16$, $p = .003$. Post hoc t tests² revealed that related lists only elicited higher JOLs when they are presented before unrelated lists, $t(49) = 6.57$, $p < .001$, but not when they are presented after unrelated lists, $t(49) = 1.48$, $p = .145$. Additionally, there was a List Relatedness \times Font \times Block Order interaction, $F(1, 49) = 5.91$, $MSE = 17.75$, $\eta_p^2 = .11$, $p = .019$. When related lists were presented in the first block and the unrelated lists were presented in the second, the JOL effect of font size was non-significant in either unrelated lists, $t(49) = .32$, $p = .748$, or related lists, $t(49) = 2.15$, $p = .072$. However, when the unrelated lists were presented first, there was a significant effect of font size on JOLs in unrelated lists, $t(49) = 3.20$, $p = .010$, but not in related lists, $t(49) = .96$, $p = .455$. Thus, when related lists or unrelated lists were presented in the first block, there was a significant JOL effect of font size with unrelated lists but not with related lists. When related lists or unrelated lists were presented in the second block, there was no JOL effect of font size with either list type.

Recall results

Table 1 also summarizes the descriptive results for the proportion of correct recall in Experiment 1. A 2 (list relatedness: related, unrelated) \times 2 (font size: 18pt, 48pt) \times 2 (block

² The p values for all post hoc t tests are corrected for multiple comparisons using the Benjamini–Hochberg method (Benjamini & Hochberg, 1995).

Table 1 Mean and standard deviations of judgments of learning and recall in Experiment 1

List type	Block order			
	Unrelated/Related		Unrelated/Related	
	18 pt	48 pt	18 pt	48 pt
JOLs				
Related	63.44 (18.00)	66.02 (18.62)	55.49 (13.16)	56.87 (13.24)
Unrelated	47.65 (16.68)	47.29 (16.56)	49.40 (10.87)	53.66 (14.51)
Recall				
Related	.69 (.19)	.55 (.19)	.67 (.23)	.57 (.21)
Unrelated	.37 (.23)	.40 (.24)	.35 (.20)	.40 (.18)

Standard deviations are in the parentheses. The measure of recall is the proportion of correct recall

order: related/unrelated, unrelated/related) mixed ANOVA for recall revealed a list relatedness main effect, $F(1, 49) = 124.99$, $MSE = .02$, $\eta_p^2 = .72$, $p < .001$, and a font size main effect, $F(1, 49) = 10.35$, $MSE = .01$, $\eta_p^2 = .17$, $p = .002$. There was also a List Relatedness \times Font Size interaction, $F(1, 49) = 28.15$, $MSE = .01$, $\eta_p^2 = .36$, $p < .001$, as smaller words produced higher recall for related lists, $t(49) = 6.60$, $p < .001$, but not for unrelated lists, $t(49) = -1.75$, $p = .087$. Thus, although font size did not affect recall for unrelated lists, as reported in many prior studies (e.g., Hu et al., 2016; Kornell et al., 2011; Rhodes & Castel, 2008; Yang et al., 2018), the smaller font actually led to better recall in related lists.

JOL resolution results

The JOL resolution data can be found in Table 2. A 2 (list relatedness: related, unrelated) \times 2 (font size: 18pt, 48pt) \times 2 (block order: related/unrelated, unrelated/related) mixed ANOVA was conducted on JOL resolution, with five participants' data removed because their gamma correlations were not calculable in at least one condition. The ANOVA revealed a main effect of font size, $F(1, 44) = 11.90$, $MSE = .10$, $\eta_p^2 = .21$, $p = .001$, as resolution was better for the smaller font than for the larger font. The main effect of list relatedness, $F(1, 44) = 2.24$, $MSE = .10$, $\eta_p^2 = .05$, $p = .142$,

Table 2 Mean and standard deviations of JOL resolution for all experiments by list relatedness and font size

Experiment	Related		Unrelated	
	18 pt	48 pt	18 pt	48 pt
Experiment 1	.45 (.35)	.37 (.34)	.39 (.41)	.19 (.42)
Experiment 2	.38 (.39)	.22 (.37)	.35 (.36)	.24 (.42)
Experiment 3	.33 (.41)	.24 (.41)	.42 (.39)	.47 (.31)

Standard deviations are in the parentheses. The measure of JOL resolution is Goodman–Kruskal gamma correlations

and the List Relatedness \times Font Size interaction, $F(1, 44) = .73$, $MSE = .09$, $\eta_p^2 = .02$, $p = .396$, were not significant.

Experiment 2

Experiment 1 showed that font size had a significant JOL effect with unrelated word lists but not with related lists in the first block, and font size had no significant JOL effect with either related or unrelated lists in the second block. Thus, the effects of font size on JOLs seem to be mitigated by list relatedness. Additionally, although font size did not affect recall for unrelated lists, the smaller font produced better recall for related lists.

In Experiment 1, related lists and unrelated lists were presented in a blocked manner, so participants saw a set of lists that were homogeneously related or homogeneously unrelated. In Experiment 2, we increased the salience of list relatedness by presenting related and unrelated lists within each block. Given that the JOL-making process is comparative and inferential in nature, we expected that the mixed-list design would enhance participants' awareness of variations in list relatedness and prompt them to weigh relatedness more in JOLs. In short, our goal was to further examine the JOL and recall effects of font size when list relatedness is made more salient.

Method

Participants, materials, and procedure

Fifty-one undergraduates ($M_{age} = 19.82$, $SD_{age} = 1.14$) participated in Experiment 2. All participants were compensated with extra course credits. The sample size, materials and procedure were identical to Experiment 1 with only one change: The lists were presented in a mixed rather than blocked manner in that there was one related list and one unrelated list presented in each block.

Results

JOL results

The descriptive results for JOLs in Experiment 2 can be found in Table 3. The JOLs were submitted to a 2 (list relatedness: related, unrelated) × 2 (font size: 18 pt, 48 pt) × 2 (block: first, second) × 2 (list order: related/unrelated, unrelated/related) mixed ANOVA. Because Experiment 1 demonstrated considerable differences between blocks, we included block as a factor in the analyses. Additionally, because we presented a related list and an unrelated list in each block, with the order of lists counterbalanced across participants, we also included list order as a factor.

The ANOVA revealed a main effect of list relatedness, $F(1, 49) = 65.17, MSE = 209.89, \eta_p^2 = .57, p < .001$, as related lists produced higher JOLs than unrelated lists, and a main effect of block, $F(1, 49) = 5.35, MSE = 238.07, \eta_p^2 = .10, p = .025$, as JOLs were higher in the first block than the second. The main effect of font size was not significant, $F(1, 49) = 2.96, MSE = 107.10, \eta_p^2 = .06, p = .092$.

The ANOVA also revealed a List Relatedness × Font Size interaction, $F(1, 49) = 4.46, MSE = 37.78, \eta_p^2 = .08, p = .040$. Post hoc tests revealed that the larger font only produced significantly higher JOLs for related lists, $t(49) = 3.12, p = .010$, but not for unrelated lists, $t(49) = .36, p = .721$. Moreover, there was a List Relatedness × Font Size × Block interaction, $F(1, 49) = 4.22, MSE = 21.59, \eta_p^2 = .08, p = .045$. Font size affected JOLs for related lists only in the second block, $t(49) = 2.64, p = .044$, but not in the first block, $t(49) = 1.98, p = .106$, and font size had no JOL effect for unrelated lists in either block, $ts = 1.10$ and $-.58, ps = .367$ and $.564$.

Additionally, the ANOVA revealed a List Relatedness × Block interaction, $F(1, 49) = 4.92, MSE = 161.23, \eta_p^2 = .09, p = .031$, as the list relatedness effect was stronger in the second block, $t(49) = 14.70, p < .001$, than in the first block, $t(49) = 9.00, p < .001$. There was also a List Relatedness × List Order

interaction, $F(1, 49) = 10.89, MSE = 209.89, \eta_p^2 = .19, p = .002$, as the list relatedness effect was stronger when a related list preceded an unrelated list, $t(49) = 16.81, p < .001$, than when an unrelated list preceded a related list, $t(49) = 7.01, p = .003$. Last, there was a List Relatedness × Block × List Order interaction, $F(1, 49) = 21.29, MSE = 161.23, \eta_p^2 = .30, p < .001$. Specifically, in the first block, the effect of list relatedness was significant when a related list preceded an unrelated list, $t(49) = 8.85, p < .001$, but not significant when an unrelated list preceded a related list, $t(49) = -.64, p = .524$. However, in the second block, the effect of list relatedness was significant regardless of list order, $ts = 5.16$ and $4.81, ps < .001$.

Recall results

The descriptive results for the proportion of correct recall in Experiment 2 are displayed in Table 3. Based on the outlier exclusion criterion described in Experiment 1, one participant with 100% recall accuracy was identified as an outlier and was removed from the analyses. The qualitative pattern of the results remained the same before and after the removal of the outlier. A 2 (list relatedness: related, unrelated) × 2 (font size: 18 pt, 48 pt) × 2 (block: first, second) × 2 (list order: related/unrelated, unrelated/related) mixed ANOVA revealed a main effect of list relatedness, $F(1, 48) = 242.61, MSE = .03, \eta_p^2 = .84, p < .001$, and a main effect of font size, $F(1, 48) = 4.24, MSE = .02, \eta_p^2 = .08, p = .045$, as recall was higher for related lists and for the smaller font.

There was also a List Relatedness × Font size interaction, $F(1, 48) = 11.57, MSE = .02, \eta_p^2 = .19, p = .001$, as the smaller font only produced better recall for related lists, $t(48) = 3.69, p = .001$, but not for unrelated lists, $t(48) = -1.02, p = .312$. Further, there was a List Relatedness × Block interaction, $F(1, 48) = 4.55, MSE = .03, \eta_p^2 = .09, p = .038$, and a List Relatedness × List Order interaction, $F(1, 48) = 7.00, MSE = .03, \eta_p^2 = .13, p = .011$. The list

Table 3 Mean and standard deviations of judgments of learning and in Experiment 2

List type	Block			
	Block 1		Block 2	
	18 pt	48 pt	18 pt	48 pt
JOLs				
Related	60.52(19.05)	63.15 (18.15)	57.56 (19.44)	61.38 (20.20)
Unrelated	49.44 (18.99)	51.57 (17.22)	45.50 (19.45)	44.49 (17.55)
Recall				
Related	.66 (.18)	.52 (.22)	.65 (.25)	.61 (.24)
Unrelated	.35 (.21)	.36 (.17)	.29 (.21)	.32 (.20)

Standard deviations are in the parentheses. The measure of recall is the proportion of correct recall

relatedness effect was stronger in the second block, $t(48) = 12.30$, $p < .001$, than in the first block, $t(48) = 9.61$, $p < .001$, and it was stronger when the related list preceded the unrelated list, $t(48) = 14.78$, $p < .001$, relative to the other way around, $t(48) = 8.21$, $p < .001$. In brief, the recall results in Experiment 2 are highly consistent with Experiment 1.

JOL resolution results

The JOL resolution data are presented in Table 2. We no longer included list order and block as factors in the ANOVA because splitting data by block and list order would render nearly half of the participants' gamma correlations incalculable in at least one condition. Thus, a 2 (relatedness: related, unrelated) \times 2 (font: 18 pt, 48 pt) repeated-measures ANOVA was conducted, with one participant's data removed as an outlier and three additional participants' data removed because gamma correlations were not calculable for them in at least one condition. The ANOVA revealed the same pattern as in Experiment 1: There was a main effect of font size $F(1, 46) = 6.28$, $MSE = .17$, $\eta_p^2 = .12$, $p = .016$, as resolution was better for the smaller font than for the larger font. The main effect of list relatedness, $F(1, 46) = .17$, $MSE = .15$, $\eta_p^2 = .00$, $p = .682$, and the List Relatedness \times Font Size interaction, $F(1, 46) = .29$, $MSE = .16$, $\eta_p^2 = .01$, $p = .592$, were not significant.

Experiment 3

Experiment 2 showed that the JOL and recall effects of font size were both modified by list relatedness when list relatedness was manipulated in a mixed-list design. However, we observed a puzzling result: Font size produced a significant JOL effect for related lists in the second block. We suspect that this result is a statistical aberration, considering that the JOL effects of font size were not significant for related lists in Experiment 1 and in the first block of Experiment 2. To resolve the uncertainty, we conducted Experiment 3 to (a) replicate the results of Experiments 1 and 2 and (b) examine whether provoking stronger relational processing would completely eliminate the JOL effects of font size.

To stimulate stronger relational processing in related lists, we replaced categorized lists with DRM lists, which are traditionally administered in false memory experiments. DRM lists produce higher levels of false memory than categorized lists, indicating that they trigger even stronger relational processing that focuses on connecting the gist across individual items (e.g., Brainerd & Reyna, 2005). The reason is that there are multiple overlapping semantic relations among list

words (e.g., category membership, situational membership, synonymy) rather than only one type of relations (Brainerd et al., 2008, 2020; Cann et al., 2011). As our principal interest does not lie with false memory, we presented the critical distractor of each DRM list as the first word on the list to enhance relational processing, just as we did for the categorized lists in Experiments 1 and 2.

Method

Participants, materials, and procedure

Forty-nine undergraduates ($M_{\text{age}} = 19.38$, $SD_{\text{age}} = 1.27$) participated in Experiment 3. All participants were compensated with extra course credits. The sensitivity analysis indicated that our sample size provided sufficient power ($1 - \beta > .80$) to detect a small effect size ($\eta_p^2 = .04$) for the main effect of font size or the interaction between font size and list relatedness in mixed ANOVAs. The materials and procedure were the same as in Experiment 2, except for one change: We replaced categorized lists with DRM lists that were chosen from the Appendix of Roediger et al. (2001). Thus, among the four 16-word to-be-remembered lists, half were DRM lists, with each list including a critical distractor and the 15 forward associates of the critical distractor. The other two lists were semantically unrelated lists that consisted of words randomly selected from other unused DRM lists. No participant was excluded in Experiment 3 based on the outlier exclusion criterion described in Experiment 1.

Results

JOL results

The descriptive results for JOLs in Experiment 3 appear in Table 4. We conducted a 2 (list relatedness: related, unrelated) \times 2 (font size: 18 pt, 48 pt) \times 2 (block: first, second) \times 2 (list order: related/unrelated, unrelated/related) mixed ANOVA on JOLs. The ANOVA revealed a main effect of list relatedness, $F(1, 47) = 13.12$, $MSE = 242.36$, $\eta_p^2 = .22$, $p < .001$, whereas there was again no significant main effect of font size, $F(1, 47) = .001$, $MSE = 151.59$, $\eta_p^2 = .00$, $p = .975$. Additionally, there was a List Relatedness \times List Order interaction, $F(1, 47) = 33.47$, $MSE = 242.36$, $\eta_p^2 = .42$, $p < .001$, and a List Relatedness \times List Order \times Block interaction, $F(1, 47) = 7.34$, $MSE = 295.89$, $\eta_p^2 = .14$, $p = .009$. In the first block, JOLs were higher for related lists than for unrelated lists regardless of list order, $t_s = 5.51$ and 3.39, $p_s < .002$. However, in the second block, related lists only elicited higher JOLs when presented before the unrelated lists, $t(47) = 3.82$, $p < .001$, but not when presented after the unrelated lists, $t(47) = 1.15$, $p = .258$.

Recall results

Table 4 lists the descriptive results for the proportion of correct recall in Experiment 3. Those data were submitted to a 2 (list relatedness: related, unrelated) \times 2 (font size: 18 pt, 48 pt) \times 2 (block: first, second) \times 2 (list order: related/unrelated, unrelated/related) mixed ANOVA. The ANOVA revealed a main effect of list relatedness, $F(1, 47) = 116.84$, $MSE = .04$, $\eta_p^2 = .71$, $p < .001$, and a main effect of font size, $F(1, 47) = 7.96$, $MSE = .02$, $\eta_p^2 = .14$, $p = .007$. Additionally, there was a List Relatedness \times Font Size interaction, $F(1, 47) = 5.19$, $MSE = .03$, $\eta_p^2 = .10$, $p = .027$, as the smaller font improved recall for related lists, $t(47) = 4.67$, $p < .001$, but not for unrelated lists, $t(47) = -.52$, $p = .609$. Again, the recall results in Experiment 3 were highly consistent with Experiments 1 and 2.

JOL resolution results

The descriptive data for JOL resolution are presented in Table 2. Four participants' gamma correlations were not calculable in at least one condition, and thus, their data were removed from the analyses. Owing to the same reason described in Experiment 2, list order and block were not included as factors in the ANOVA for JOL resolution. A 2 (relatedness: related, unrelated) \times 2 (font size: 18 pt, 48 pt) repeated-measures ANOVA revealed a main effect of relatedness, $F(1, 44) = 8.74$, $MSE = .13$, $\eta_p^2 = .17$, $p = .005$, as JOL resolution was worse for related lists than for unrelated lists. The main effect of font size, $F(1, 44) = .05$, $MSE = .13$, $\eta_p^2 = .00$, $p = .819$, and the List Relatedness \times Font Size interaction, $F(1, 44) = 2.00$, $MSE = .12$, $\eta_p^2 = .04$, $p = .165$, were not significant.

General discussion

In the present study, we factorially manipulated list relatedness and font size and examined (a) whether the JOL effects of font size persist when list relatedness is varied, and (b)

whether font size affects recall accuracy with related word lists. The effects of font size on JOLs and recall in Experiments 1–3 are summarized in Table 5. Across these experiments, we found that both the JOL and recall effects of font size were modified by list relatedness. For JOLs, the effects of font size were moderated or eliminated when list relatedness was varied simultaneously. For recall, the effects of font size were only found for related lists. Below, we first discuss the JOL results and their implications. Then, we turn to the recall results. Last, we make some recommendations for future research on the font size effect and metamemory in general.

The JOL effects of font size depends on list relatedness

In Experiment 1, list relatedness was manipulated in a blocked-list design such that there were either two categorized lists or two unrelated lists in each block. As can be seen in Table 5, in the first block of Experiment 1, we found that the larger font led to significantly higher JOLs for unrelated lists but not for categorized lists, and the JOL effects of font size disappeared in the second block. Experiments 2 and 3 provide additional support that the JOL effects of font size are mitigated by list relatedness. In those two experiments, lists were presented in a mixed manner as a related list (categorized or DRM list) and an unrelated list was intermixed in each block. Such a mixed-list design should highlight the variability in list relatedness, and thus list relatedness should be a more salient cue than in the blocked-list design. As shown in Table 5, there was no effect of font size on JOLs for either related or unrelated lists in the first block of both Experiments 2 and 3. However, in the second block of Experiment 2, we observed an interaction between font size and list relatedness, as the larger font led to higher JOLs for related lists but not for unrelated lists. As we did not replicate the interaction in Experiment 3, we think it is most likely that this interaction was merely a statistical aberration.

Table 4 Mean and standard deviations of judgments of learning and recall in Experiment 3

List Type	Block			
	Block 1		Block 2	
	18 pt	48 pt	18 pt	48 pt
JOLs				
Related	57.91(20.57)	54.57 (20.30)	56.86 (20.55)	57.71 (19.27)
Unrelated	52.26 (20.04)	52.75 (20.10)	47.02 (21.20)	49.37 (20.94)
Recall				
Related	.59 (.25)	.50 (.25)	.64 (.25)	.57 (.22)
Unrelated	.37 (.22)	.38 (.23)	.37 (.26)	.34 (.28)

Standard deviations are in the parentheses. The measure of recall is the proportion of correct recall

Table 5 Summary of the effects of font size on JOLs and recall in Experiments 1–3

Experiment	List type	JOLs		Recall	
		Block 1	Block 2	Block 1	Block 2
Experiment 1	Related	18 = 48	18 = 48	18 > 48	18 > 48
	Unrelated	18 < 48	18 = 48	18 = 48	18 = 48
Experiment 2	Related	18 = 48	18 < 48	18 > 48	18 > 48
	Unrelated	18 = 48	18 = 48	18 = 48	18 = 48
Experiment 3	Related	18 = 48	18 = 48	18 > 48	18 > 48
	Unrelated	18 = 48	18 = 48	18 = 48	18 = 48

18 = 18-pt font (the smaller-font condition), 48 = 48-pt font (the larger-font condition), “=” means not significantly different from each other ($p > .050$), “>” means significantly higher ($p < .050$), “<” means significantly lower ($p < .050$)

To sum up, the three experiments provided converging evidence that the JOL effects of font size depend on list relatedness: The effect is either moderated or eliminated when list relatedness is manipulated in parallel with font size; that is, differences in relatedness remove some or all of the variance that is normally due to differences in font size. To the best of our knowledge, related word lists have not figured in prior research on the font size effect. Thus, our study provides novel evidence that list relatedness may be a boundary condition for this effect. Moreover, our study supports the notion that there is a trade-off between item-specific and relational processing when making JOLs. Specifically, the availability of a relational cue (list relatedness) reduced the effects of an item-specific cue (font size). This has important implications for the cue-integration framework of JOLs (Peynircioğlu & Tatz, 2019; Undorf & Bröder, 2020; Undorf et al., 2018). To recap, this account predicts that people can integrate multiple cues when making JOLs, and thus JOLs should reflect the effects of multiple cues simultaneously if those cues affect JOLs when manipulated in isolation. In that connection, prior findings that the JOL effects of font size persist when font size and pair relatedness are manipulated simultaneously are in harmony with this account (e.g., Price & Harrison, 2017; Rhodes & Castel, 2008). However, our results showed that list relatedness could eliminate the JOL effects of font size. This suggests that even if some cues have robust JOL effects when manipulated individually, their effects can be overshadowed by the effects of other preferentially processed cues that are simultaneously present. Specifically, there is likely a trade-off between item-specific cues like font size and relational cues like list relatedness, in that the effects of some item-specific cues can be mitigated by the presence of relational cues.

The recall effects of font size depends on list relatedness

Across the three experiments, we found a robust recall effect of font size. Specifically, the smaller font always led to better recall than the larger font for related lists but not for unrelated lists. Here, although the classic finding is that font size affects JOLs but not recall, we stress that most prior studies used unrelated word lists as study materials (for a review, see Chang & Brainerd, 2022). On the one hand, our findings were consistent with the prior findings that font size has minimal effects on recall for unrelated lists. On the other hand, our results demonstrated that font size can affect recall for related lists.

The recall advantage of the smaller font has also been reported in some prior studies (Halamish, 2018; Undorf & Zimdahl, 2019). According to Halamish (2018), smaller fonts may improve memory because they are processed less fluently than larger fonts, which triggers more effortful processing. In that connection, many studies provide evidence that smaller fonts are processed less fluently than larger fonts (Susser et al., 2013; Undorf et al., 2017; Yang et al., 2018). Therefore, it is likely that the effects of smaller fonts are a “desirable difficulty” effect (Bjork, 1994), which reduces processing fluency and provoke more thorough processing that redounds to memory performance.

Whereas Halamish (2018) reported that the recall benefit of smaller fonts was not robust and was eliminated by the solicitation of JOLs, we found that the smaller font consistently led to better recall, when JOLs were administered. Again, the different patterns may be due to the difference in study materials, as Halamish used pure unrelated lists whereas we used both related and unrelated lists. A possible reason for our more robust recall effect for the smaller font with related lists is that the item-specific processing triggered by the smaller font complements the relational processing triggered by list relatedness. As suggested by Hunt and his colleagues, memory is optimal when encoding is characterized by high levels of both item-specific and relational processing (e.g., Hunt & Einstein, 1981; Hunt & Seta, 1984). While related lists clearly favor relational processing, unrelated lists primarily provoke item-specific processing. Thus, the item processing enhancement induced by the smaller font should be complementary for related lists but not for unrelated lists.

The finding that the smaller font produced better recall for related lists has important practical implications. In daily life, to-be-remembered information is often presented as part of related materials. For instance, students need to remember multiple concepts that are relevant to a topic in textbooks, and advertisers list multiple advantages of a product in advertisements. In such circumstances,

particularly important information is often highlighted in a larger font. However, this may not optimize memory, as our results suggest that the smaller font enhanced memory for related items relative to the larger font. Further, Experiments 1 and 2 also showed that the smaller font produced higher JOL resolution, suggesting that it also helps people make more accurate estimates of their future memory. This is also potentially beneficial for memory outcomes, as metacognitive monitoring can guide people's study strategies (Nelson & Leonesio, 1988; Thiede & Dunlosky, 1999). Last but not least, it is worth mentioning that although some prior studies suggest that perceptual manipulations that reduce processing fluency (e.g., disfluent font type, blurring) produce desirable difficulty affects (Diemand-Yauman et al., 2011; Geller et al., 2018; Rosner et al., 2015; Sungkhasettee et al., 2011), there is also contrary evidence (Huff et al., 2022; Taylor et al., 2020; Yue et al., 2013). Thus, whether smaller fonts produce robust memory benefits requires further examination.

Future directions

The current study examined the font size effect under the novel condition of list relatedness (as opposed to pair relatedness). An obvious next step would be to investigate the same questions with more complex materials. For example, future research could focus on textual materials. In that regard, Luna et al. (2019) examined the font size effect with related and unrelated sentences and found the standard pattern in which JOLs were higher for larger-font sentences, but cued recall was not affected by font size. However, Luna et al. manipulated relatedness within sentences rather than across sentences, so their relatedness manipulation was still focused on item-specific rather than relational processing. Thus, it is an open question whether the results remain the same when relatedness is manipulated across sentences. Additionally, Luna et al. manipulated font size at the sentence level rather than at the word level. For the sake of comparability with the extant literature and ecological validity, it is important to discover whether results change when font size is manipulated for specific words within the sentences.

It would also be instructive to manipulate a wider range of font sizes. As Chang and Brainerd's (2022) meta-analysis showed, the memory effects of font size vary with the specific font sizes being compared, such that results differ when very small fonts are compared to intermediate-size fonts versus when the smaller and larger fonts are both in the intermediate range. Therefore, it is particularly important to identify which font sizes produce the optimal memory outcomes when compared to the most conventionally used font size in typographical practices.

Last and perhaps most important, future studies should aim to develop a more coherent theoretical framework for

JOLs that weights different types of cues differently in cue integration. In our studies, font size was deemphasized in the JOL process when list relatedness was simultaneously varied, suggesting that relational cues were preferentially processed relative to certain item-specific cues. Why? An obvious possibility is that list relatedness is more diagnostic of eventual memory accuracy than font size. However, prior studies demonstrated that the JOL effects of font size are still robust when other more diagnostic cues are manipulated. For instance, Price and Harrison (2017) and Rhodes and Castel (2008) showed that font size and pair relatedness had independent effects on JOLs, and Undorf et al. (2018) showed that font size, number of study presentations, concreteness, and emotionality all affected JOLs, and their effects did not interact. Thus, it seems that diagnosticity is not sufficient for a cue to be preferentially processed. It would be important to (a) recognize that cues are not necessarily weighted equally during cue integration, and (b) investigate which cues are naturally weighted more heavily than the others.

Author contribution Conceptualization: Minyu Chang; Methodology: Minyu Chang; Formal analysis and investigation: Minyu Chang; Writing—original draft preparation: Minyu Chang; Writing—review and editing: Minyu Chang, C. J. Brainerd; Supervision: C. J. Brainerd.

Data availability All data and materials are available at <https://osf.io/tj2pv/>.

Code availability Available upon request to the first author.

Declarations

Conflicts of interest/Competing interests Not applicable

References

- Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, 48(1), 1–15. [https://doi.org/10.1016/S0749-596X\(02\)00518-1](https://doi.org/10.1016/S0749-596X(02)00518-1)
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.
- Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford University Press.
- Brainerd, C. J., Chang, M., & Bialer, D. M. (2020). From association to gist. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000938>
- Brainerd, C. J., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review*, 15(6), 1035–1053. <https://doi.org/10.3758/PBR.15.6.1035>
- Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002). Dual-retrieval processes in free and associative recall. *Journal of*

- Memory and Language*, 46(1), 120–152. <https://doi.org/10.1006/jmla.2001.2796>
- Brysaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Cann, D. R., McRae, K., & Katz, A. N. (2011). False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength. *The Quarterly Journal of Experimental Psychology*, 64(8), 1515–1542. <https://doi.org/10.1080/17470218.2011.560272>
- Chang, M., & Brainerd, C. J. (2022). Association and dissociation between judgments of learning and memory: A meta-analysis of the font size effect. *Metacognition and Learning*, 17(2), 443–476. <https://doi.org/10.1007/s11409-021-09287-3>
- Chen, Y., Li, F., & Li, W. (2019). The influence of learner’s beliefs about processing fluency on font-size effect. *Acta Psychologica Sinica*, 15(2), 154–162. <https://doi.org/10.3724/SP.J.1041.2019.00154>
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17–22. <https://doi.org/10.1037/h0046671>
- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, 118(1), 111–115. <https://doi.org/10.1016/j.cognition.2010.09.012>
- Double, K. S. (2019). Do judgments of learning impair recall when uninformative cues are salient? *PsyArXiv*. <https://doi.org/10.31234/osf.io/a5bxw>
- Durda, K., & Buchanan, L. (2006). WordMine2 <http://www.web2.uwindsor.ca/wordmine/>. Accessed 19 May 2021.
- Fan, T., Zheng, J., Hu, X., Su, N., Yin, Y., Yang, C., & Luo, L. (2021). The contribution of metamemory beliefs to the font size effect on judgments of learning: Is word frequency a moderating factor? *PLoS One*, 16, e0257547. <https://doi.org/10.1371/journal.pone.0257547>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Geller, J., Still, M. L., Dark, V. J., & Carpenter, S. K. (2018). Would disfluency by any other name still be disfluent? Examining the disfluency effect with cursive handwriting. *Memory & Cognition*, 46(7), 1109–1126. <https://doi.org/10.3758/s13421-018-0824-6>
- Halamish, V. (2018). Can very small font size enhance memory? *Memory & Cognition*, 46(6), 979–993. <https://doi.org/10.3758/s13421-018-0816-6>
- Harrell, F. E., Jr. (2019). Package ‘Hmisc.’ <https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf>. Accessed 4 Oct 2022.
- Höhne, J. K., & Schlosser, S. (2018). Investigating the adequacy of response time outlier definitions in computer-based web surveys using paradata SurveyFocus. *Social Science Computer Review*, 36(3), 369–378. <https://doi.org/10.1177/0894439317710450>
- Hourihan, K. L., & Tullis, J. G. (2015). When will bigger be (recalled) better? The influence of category size on JOLs depends on test format. *Memory & Cognition*, 43(6), 910–921. <https://doi.org/10.3758/s13421-015-0516-4>
- Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgement of learning. *Memory*, 24(3), 383–398. <https://doi.org/10.1080/09658211.2015.1009470>
- Huff, M. J., Maxwell, N. P., & Mitchell, A. (2022). Distinctive sans Forgetica font does not benefit memory accuracy in the DRM paradigm. *Cognitive Research: Principles and Implications*, 7(1), 102. <https://doi.org/10.1186/s41235-022-00448-9>
- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 497–514. [https://doi.org/10.1016/S0022-5371\(81\)90138-9](https://doi.org/10.1016/S0022-5371(81)90138-9)
- Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(3), 454–464. <https://doi.org/10.1037/0278-7393.10.3.454>
- Kausler, D. H. (1974). *Psychology of verbal learning and memory*. Academic Press.
- 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. <https://doi.org/10.1037/0096-3445.126.4.349>
- 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 judgment. *Psychological Science*, 22(6), 787–794. <https://doi.org/10.1177/0956797611407929>
- Luna, K., Albuquerque, P. B., & Martín-Luengo, B. (2019). Cognitive load eliminates the effect of perceptual information on judgments of learning with sentences. *Memory & Cognition*, 47(1), 106–116. <https://doi.org/10.3758/s13421-018-0853-1>
- Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology*, 71(7), 1626–1636. <https://doi.org/10.1080/17470218.2017.1343362>
- Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 327–372). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60516-2](https://doi.org/10.1016/S0079-7421(08)60516-2)
- Matvey, G., Dunlosky, J., & Schwartz, B. (2006). The effects of categorical relatedness on judgements of learning (JOLs). *Memory*, 14(2), 253–261. <https://doi.org/10.1080/09658210500216844>
- 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. <https://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. <https://doi.org/10.3758/s13423-012-0343-6>
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the “labor-in-vain effect.” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 676–686. <https://doi.org/10.1037/0278-7393.14.4.676>
- Peynircioglu, Z. F., & Tatz, J. R. (2019). Intensifying the intensity illusion in judgments of learning: Modality and cue combinations. *Memory & Cognition*, 47(3), 412–419. <https://doi.org/10.3758/s13421-018-0875-8>
- Price, J., & Harrison, A. (2017). Examining what prestudy and immediate judgments of learning reveal about the bases of metamemory judgments. *Journal of Memory and Language*, 94, 177–194. <https://doi.org/10.1016/j.jml.2016.12.003>
- 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. <https://doi.org/10.1037/a0013684>
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814. <https://doi.org/10.1037/0278-7393.21.4.803>
- Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression

- analysis. *Psychonomic Bulletin & Review*, 8(3), 385–407. <https://doi.org/10.3758/BF03196177>
- Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, 160, 11–22. <https://doi.org/10.1016/j.actpsy.2015.06.006>
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). Afex: Analysis of factorial experiments (R package version 0.13–145) [computer software]. <http://CRAN.R-project.org/package=afex>. Accessed 4 Oct 2022.
- Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, 18(5), 973–978. <https://doi.org/10.3758/s13423-011-0114-9>
- Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, 41(7), 1000–1011. <https://doi.org/10.3758/s13421-013-0323-8>
- Tatz, J. R., & Peynircioğlu, Z. F. (2020). Judgments of learning in context: Backgrounds can both reduce and produce metamemory illusions. *Memory & Cognition*, 48(4), 581–595. <https://doi.org/10.3758/s13421-019-00991-9>
- Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: The (lack of) effect of sans Forgetica on memory. *Memory*, 28(7), 850–857. <https://doi.org/10.1080/09658211.2020.1758726>
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 1024–1037. <https://doi.org/10.1037/0278-7393.25.4.1024>
- Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, 73(4), 629–642. <https://doi.org/10.1177/1747021819882308>
- Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, 43(4), 647–658. <https://doi.org/10.3758/s13421-014-0479-x>
- Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, 46(4), 507–519. <https://doi.org/10.3758/s13421-017-0780-6>
- Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 97–109. <https://doi.org/10.1037/xlm0000571>
- Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, 92, 293–304. <https://doi.org/10.1016/j.jml.2016.07.003>
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battag and Montague (1969) norms. *Journal of Memory and Language*, 50(3), 289–335. <https://doi.org/10.1016/j.jml.2003.10.003>
- Wang, J., Qu, K., & Zhang, H. (2020). The effect of processing fluency on the font size effect of judgments of learning. *Journal of Psychological Science*, 43(1), 17–22.
- Yang, C., Huang, T. S.-T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, 99, 99–110. <https://doi.org/10.1016/j.jml.2017.11.005>
- 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. <https://doi.org/10.3758/s13421-012-0255-8>
- Open practice statement** All data and materials are available (<https://osf.io/7j2pv/>). None of the experiments was preregistered.
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.