Memory & Cognition (2019) 47:106–116 https://doi.org/10.3758/s13421-018-0853-1

Cognitive load eliminates the effect of perceptual information on judgments of learning with sentences

Karlos Luna[1] • Pedro B. Albuquerque [1] • Beatriz Martín-Luengo [2]

1 Psychology Research Centre, School of Psychology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

2 Centre for Cognition and Decision Making, (School of Psychology), National Research University Higher School of Economics, Krivokolenniy sidewalk, 3a, Moscow, Russian Federation 101000

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13421-018-0853-1) contains supplementary material, which is available to authorized users

---

Abstract

Items presented in large font are rated with higher judgments of learning (JOLs) than those presented in small font. According to current explanations of this phenomenon in terms of processing fluency or implicit beliefs, this effect should be present no matter the type of material under study. However, we hypothesized that the linguistic cues present in sentences may prevent using font size as a cue for JOLs. Experiment 1, with short sentences, showed the standard font-size effect on JOLs, and Experiment 2, with pairs of longer sentences, showed a reduced effect. These results suggest that linguistic factors do not prevent font size from being used for JOLs. However, Experiment 3, with both short and long sentences, showed an effect of font size only for the former and not the latter condition, suggesting that the greater amount of to-be-remembered information eliminated the font-size effect. In Experiment 4, we tested a mechanism to explain this result and manipulated cognitive load using the dot-memory task. The short sentences from Experiments 1 and 3 were used, and the results replicated the font-size effect only in the low-cognitive load condition. Our results are consistent with the idea that perceptual information is used to make JOLs only with materials such as words, word pairs, or short sentences, and that the increased cognitive load required to process longer sentences prevents using font size as a cue for JOLs.

Keywords Judgments of learning . Font size . Sentences . Cognitive load

Perceptual information, such as font size, affects metamemory judgments. Items presented in large font are rated as easier to remember (i.e., rated with higher judgments of learning, JOLs) than words presented in small font (Hu, Liu, Li, & Luo, 2016; Li, Xie, Li, & Li, 2015; McDonough & Gallo, 2012; Rhodes & Castel, 2008; Susser, Mulligan, & Besken, 2013). However, font size has a small actual effect on memory performance (Halamish, 2018; Luna, Martín-Luengo, & Albuquerque, 2018; Undorf & Zimdalh, 2018), resulting in mismatched effects of font size in memory and metamemory. Here we studied the generalizability of the effect of font size on JOLs to sentences.

Two major explanations have been proposed to account for the effect of font size on metamemory. First, the phenomenon was explained in terms of experience-based factors: Words in large font are processed more fluently, and this fluency is misattributed as indicative of better future memory (Rhodes & Castel, 2008). The effect of fluency on metamemory has been supported by several studies (Besken & Mulligan, 2013; Susser et al., 2013; Undorf & Erdfelder, 2011, 2015). However, an explanation of the effect of font size on JOLs based exclusively on fluency does not account for all the results. For ex- ample, Mueller, Dunlosky, Tauber, and Rhodes (2014, Exp. 4) told participants that a word in small [or large] font was going to be presented and requested JOLs before presenting the words. Results showed higher JOLs when large font was mentioned, strongly suggesting that fluency was not playing a role. After these results, Mueller et al. (2014) proposed a second explanation: The effect of font size on JOLs is due to beliefs we hold about how our memory works and the variables and fac- tors that affect it (see also the analytic processing theory, Mueller & Dunlosky, 2017; Mueller, Dunlosky, & Tauber, 2016). It seems that people believe that words in large font are going to be better remembered than words in small font. Recent research has suggested that both experience-based fac- tors (e.g., encoding fluency) and theory-based factors (e.g., a priori beliefs) play roles in metamemory judgments (Besken, 2016; Frank & Kuhlmann, 2017; Undorf, Zimdahl, & Bernstein, 2017; Yan, Bjork, & Bjork, 2016).

Research about the effect of font size on JOLs has focused mostly on memory for words and word pairs. To date, only two studies have used other materials. Katzir, Hershko, and Halamish (2013) presented short texts in different font sizes and computed a reading comprehension index. Children aged seven reported higher reading comprehension with texts presented in large than in small font, but the opposite was found for children aged ten, suggesting that the relationship between font size and metamemory judgments can change under certain conditions.

Another study presented psychological concepts and definitions in either small or large font (Ball, Klein, & Brewer, 2014, Exp. 2). This study is more relevant for the present research because the participants were young adults, so developmental issues cannot explain the different results in studies with words or word pairs. Ball et al. found higher JOLs for definitions presented in small font, a finding opposite those from previous studies. However, there was one major methodological difference between their experiment and others dealing with the font-size effect: Study time was unlimited and self-paced. Participants spent more time studying items in large than in small font, and when the authors controlled study time the effect of font size on JOLs disappeared. This suggests that study time was more relevant as metacognitive cue than font size. The authors tentatively explained these results by arguing that under unlimited study time conditions, analytic processes based on the amount of time allocated to study a definition may be stronger than implicit processes based on heuristics (e.g., the belief that memory will be better if items are presented in large font than in small font). This explanation suggests a different effect of a cue (e.g., font size), depending on the availability of another cue more diagnostic of memory performance (e.g., study time). However, when participants are allowed to study words or word pairs for as long as they want, the higher JOLs for items in large font size is still replicated (Mueller et al., 2014, Exp. 2; Price & Harrison, 2017, Exp. 3). These studies suggest that study time may not be the key factor explaining the results in Ball et al. (2014).

An alternative explanation to Ball et al.'s (2014) results is that font size only affects metamemory judgments for simple materials such as words or word pairs. The study by Ball et al. provided insight into the effect of font size on JOLs for ecological materials. However, since the parameters of their research were designed to be as close as possible to a learning context, the possibility that the effect of font size on JOLs is limited to words and word pairs cannot be ruled out. In this research, we tested whether font size affects metamemory judgments with other types of materials: single sentences in Experiments 1 and 4, pairs of longer sentences in Experiment 2, and short and long sentences in Experiment 3.

Our hypothesis for Experiments 1 and 2 was that the se- mantic content, grammatical rules, and syntactic structure of sentences might provide diagnostic cues for JOLs and reduce reliance on perceptual information. During encoding, participants know that there will be a memory test and that they should try to remember the information. Since the linguistic cues could be more helpful than the perceptual information for the upcoming memory task, participants might assign them a higher weight in making JOLs, effectively reducing their reliance on font size. If this were the case, the effect of font size on metamemory judgments with sentences should be reduced or eliminated. In contrast, current explanations of the effect of font size—that is, experience-based (fluency) and theory- based (beliefs)—still predict higher JOLs for items presented in larger font. Nothing seems to suggest from either explanation that a change of materials should alter the effect of font size on JOLs.

To better disentangle any effects of semantic and syntactic cues, in Experiments 1 and 2 we also manipulated semantic relatedness within the sentences. For the semantically unrelated sentences the semantic component was low, and for the semantically related sentences it was high, but both types of sentences included syntactic and grammatical cues. Therefore, if there was any difference in the font-size effects between the two groups, we could attribute it to one type of cue or another. To foreshadow, the results from Experiment 1 suggested that linguistic factors do not reduce the font-size effect. However, the effect of font size was reduced in Experiment 2 with pairs of longer sentences. These results suggested that the length of a sentence may be a more relevant factor than its linguistic properties. In Experiment 3, we tested and confirmed that the length of the sentence moderates the effect of font size. In Experiment 4, we tested whether cognitive load may explain the moderator effect. To present information in an organized fashion, we present the research relevant to cognitive load in the introduction to Experiment 4.

Experiment 1

Method

Participants and design Fifty undergraduate students (all female; age *M* = 21.26, *SD* = 4.86 years old) participated in the experiment in exchange for course credits. A 2 (font size: small [18 points], large [48 points]) × 2 (semantic relatedness: related, unrelated) mixed design was used, with font size manipulated within subjects. We did not want participants to experience both types of materials at the same time, to prevent them from making explicit comparisons that could be used to reduce the uncertainty of JOLs, so semantic relatedness was manipulated between subjects. Twenty-four participants were allocated to the unrelated-sentences condition, and 26 to the related-sentences condition. To limit potential problems with small samples and stopping rules, in addition to the usual analysis of variance (ANOVA) and *t* tests, we also conducted Bayesian analyses that are not affected by sample size.

Normative study We conducted a normative study to create the materials. First, the authors generated 52 subject–verb– object sentences with a clear semantic relationship: The subject was always a profession, and the verb and object were typical actions performed by members of that profession with the object (e.g., BThe pilot lands the plane,^ or BThe salesperson persuades the client^). We avoided sentences with any shared root words (e.g., BThe singer sings a song^). Another 52 sentences were then created using the same words but in different sentences, to form unrelated sentences that still made sense (e.g., BThe pilot persuades the client,^ or BThe salesperson lands the plane^). To check the semantic relationship of the resulting 104 sentences (52 semantically unrelated and 52 semantically related), 14 participants who did not complete the main experiment provided a relatedness score by indicating on a scale from 1 (*very low*) to 4 (*very high*) the probability of someone with that profession performing that action. To control for basic linguistic properties, we entered subjects, verbs, and objects separately into the Portuguese linguistic database P-PAL (Soares et al., 2014) and obtained the frequency and number of letters for each word. The sentences used in the experiment were selected on the basis of the relatedness scores from the normative study (see the next section) and the linguistic frequencies and numbers of letters in the words that formed the sentences.

Materials From the 52 semantically unrelated sentences tested in the normative study, 45 were selected for the experiment. Forty were used as targets, four as primacy and recency buffers (data from these sentences were not analyzed), and one as an example in the instructions. For counterbalancing purposes, the 40 target questions were separated into two groups of 20, each with similar numbers of letters and word frequencies for their subject, verb, and object (all *p*s > .20) and with matching relatedness scores ($M = 1.40$, $SD = 0.22$, range [1.00, 1.79], and $M = 1.37$, $SD = 0.23$, range [1.00, 1.71]), $t(19) = 0.62$, $p = .540$, $d_{unb} = 0.16$. The same procedure was used for the semantically related sentences. We selected 45 sentences, 40 used as targets, four buffers, and one as the example in the instructions. Also, for counterbalancing, the 40 target questions were divided into two groups of 20 with similar semantic relationships ($M = 3.87$, $SD = 0.26$, range [3.21, 4.00], and $M = 3.90$, $SD = 0.10$, range [3.71, 4.00]), $t(19) = 0.55$, $p = .586$, $d_{unb} = 0.17$. The relatedness score was higher for the semantically related sentences than for the unrelated sentences, $t(39) = 62.86$, $p <$

$.001$, $d_{unb} = 12.08$.

Procedure Participants completed the experiment on a computer in individual soundproof cabins. The experiment was programmed in LiveCode (Version 7.1.3; LiveCode, Edinburgh, UK, https://livecode.com). Participants first gave consent, provided basic demographic information, and read the instructions. The instructions stated that 44 sentences were to be presented in different font sizes, and that participants should pay attention because they would have to remember the sentences later. For each sentence the sequence was as follows: First, a fixation point (+) was presented in the center of the screen for 1 s followed by the sentence in the same place for 5 s. The sentences were displayed in a single line, in black letters over a light bluish background and in san- serif font. Half of them were displayed in 18-point font and the other half in 48-point font, counterbalanced. No more than two sentences in the same font size were presented in a row. Depending on the condition, sentences presented were semantically related or unrelated.

Subsequently, on a different screen without the sentence, participants rated their confidence that they would later be able to recall the sentence (JOL) by selecting the appropriate button on a scale from 0% to 100% in deciles. All 11 buttons were accompanied by numerical labels ranging from 0 to 100, and participants were instructed to select 0% if they were certain that they would not remember the full sentence if they were presented later with its subject, and to select 100% if they were completely certain that would remember the full sentence. Below JOLs, they also indicated whether they wanted to see the sentence again (restudy decision), by selecting BYes^ or BNo.^ Independent of this restudy decision, participants never saw the sentences again, and there was no time limit for providing the answers. Restudy decisions were collected to test whether any effect of font size on the monitoring process (i.e., on the JOLs) would resonate to affect the control process (i.e., restudy decisions). There is a well-known relationship between JOLs and restudy decisions (Kimball, Smith, & Muntean, 2012; Luna et al., 2018; Nelson, Dunlosky, Graf, & Narens, 1994; Rhodes & Castel, 2009), and the influence of font size on restudy decisions has also been reported before (Li et al., 2015; Luna et al., 2018), but never with sentences.

After the 44 sentences had been displayed, participants completed a filler task involving writing down as many Portuguese cities as they could for 2 min, and as many coun- tries as they could for another 2 min. Finally, they completed a cued-recall test in which the subject of the sentence was presented as a cue (e.g., BThe pilot _^). Participants were instructed to write as much of the rest of the sentence as they could, or to write BI don't know^ if they recalled nothing. The subjects of the sentences were displayed one at a time in 33- point font size (the midpoint between the 18- and 48-point fonts used at study). There was no time limit for this task. The experiment lasted approximately 30 min.

Coding Responses in the cued-recall test were coded as follows: If participants provided both the verb and the object that corresponded to a given subject, the response was coded as 1. If only one of the target words was presented, the response was coded as.5. Changes in the preposition or article that linked the verb and object (e.g., from the original BThe pilot persuades the client^ to Bpersuades a client^) were ignored so that the answer was coded as 1 or.5, depending solely on the number of target words produced. Changes in number (e.g., from Bclient^ to Bclients^) were also ignored, and the responses were considered correct. If a response did not include any of the target words or was BI don't know,^ it was coded as 0.

Results

We conducted three 2 (font size: small, large) × 2 (semantic relatedness: related, unrelated) ANOVAs, one for each of our measures. When appropriate, pairwise comparisons using the Student's *t* test and Bayesian analysis were conducted. Alpha for all analyses was set to .05, and the confidence intervals reported are 95% CIs. We report partial eta-squared (η_p^2) as the effect size measure for the ANOVA, and unbiased Cohen's *d* (d_{unb}; see Cumming, 2012) for *t* tests. We also computed Goodman–Kruskal gammas between the JOLs and memory performance as a measure of resolution, for completeness. The gammas were not particularly informative for our purposes, and thus are reported in the supplemental materials.

Judgments of learning. See Table 1 for the main descriptive statistics. Participants assigned higher JOLs to sentences in large font than to those in small font, $F(1, 48) = 16.86$, $p <.001$, $\eta_p^2 = .26$. Participants also gave higher JOLs for semantically related than for unrelated sentences, $F(1, 48) = 15.49$, $p < .001$, $\eta_p^2 = .24$. The interaction was not significant, $F(1, 48) = 0.74$, $p = .395$, $\eta_p^2 = .02$.

To gain further evidence of the effect of font size on JOLs with a test independent of the number of participants, we also computed Bayesian analyses with JASP (JASP Team, 2017). A Bayes factor (BF) around 1 is not conclusive. As BF in- creases, it accumulates evidence in support of a difference between conditions, with a BF = 3 as a cutoff point to declare the evidence as moderate. As BF decreases, it accumulates evidence in support of no difference, with 0.33 as a cutoff For a more detailed explanation of Bayesian analyses applied to research on JOLs and metamemory, see Luna, Martín- Luengo, Shtyrov, and Myachykov (2016). A comparison be- tween large and small fonts with the full sample (including both the semantically related and unrelated sentences together, equivalent to the main effect from the ANOVA reported above), showed BF = 147.50, which is interpreted as extreme evidence in support of a difference. Separate analyses for both conditions also supported the effect of font size on JOLs, BF = 37.10 for related and BF = 3.72 for unrelated sentences. In conclusion, the effect of font size on JOLs was confirmed with both types of sentences.

Restudy decisions Participants more often preferred to restudy sentences in $_p$ small (*M* = 67.00, *SD* = 26.44, CI [59.67, 74.33]) than in large (*M* = 57.80, *SD* = 29.28, CI [49.68, 65.92]) font, *F*(1, 48) = 13.14, *p* = .001, η^2 = .21. Restudy decisions for semantically unrelated words (*M* = 68.23, *SD* = 27.17, CI [57.36, 79.10]) were numerically but not statistically higher than those for related sentences (*M* = 57.02, *SD* = 25.02, CI [47.40,66.64]), *F*(1, 48) = 2.31, *p* = .135, η^2 = .05. The interaction was not significant, *F*(1, 48) = 0.29, *p* = .594, η_p = .01.

Consistent with the results above and with previous studies, Pearson correlations confirmed the high and negative correlations between JOLs and restudy decisions (for *related* sentences in small font, *r* = – .71, CI [– .86, – .45], *t*(24) = 4.94, *p* = .001, and in large font, *r* = – .68, CI [– .84, – .40], *t*(24) = 4.53, *p* < .001; and for *unrelated* sentences in small font, *r* = – .66, CI [– .84, – .35], *t*(22) = 4.13, *p* = .001, and in large font, *r* = – .65, CI [– .83, – .33]), *t*(22) = – 3.97, *p* = .001).

Table 1 Mean (standard deviation) [95% confidence interval] for judgments of learning in Experiments 1 and 2

	Small (18-pt)	Large (48-pt)	Total
Experiment 1			
Related	54.13 (18.28) [47.11, 61.16]	58.81 (17.17) [52.21, 65.41]	56.47 (17.45) [49.76, 63.18]
Unrelated	35.25 (14.69) [29.37, 41.13]	42.40 (15.99) [36.00, 48.79]	38.82 (13.88) [33.27, 44.38]
Total	45.07 (19.04) [39.79, 50.35]	50.93 (18.41) [45.83, 56.03]	
Experiment 2			
Related	55.24 (15.83) [49.58, 60.91]	58.31 (12.88) [53.70, 62.92]	56.78 (13.50) [51.95, 61.61]
Unrelated	42.49 (12.90) [36.44, 48.54]	43.64 (16.61) [37.70, 49.59]	43.07 (16.51) [37.16, 48.98]
Total	48.87 (17.46) [44.45, 53.29]	50.98 (16.49) [46.81, 55.15]	

Table 2 Mean (standard deviation) [95% confidence interval] for memory performance in Experiments 1 and 2

	Small (18-pt)	Large (48-pt)	Total
Experiment 1			
Related	43.94 (16.37) [37.65, 50.24]	44.71 (14.75) [39.04, 50.38]	44.33 (14.04) [38.93, 49.72]
Unrelated	10.52 (10.08) [6.49, 14.55]	11.67 (11.93) [6.90, 16.44]	11.09 (10.38) [6.94, 15.25]
Total	27.85 (21.68) [21.84, 33.86]	28.85 (21.35) [22.93, 34.77]	
Experiment 2			
Related	31.92 (13.13) [27.23, 36.62]	33.59 (12.14) [29.25, 37.94]	32.76 (11.59) [28.61, 36.91]
Unrelated	23.79 (10.82) [19.92, 27.66]	24.67 (9.40) [21.30, 28.03]	24.23 (9.24) [20.92, 27.53]
Total	27.85 (12.61) [24.66, 30.05]	29.13 (11.67) [26.17, 32.08]	

Memory performance. Table 2 shows the main descriptive statistics as percentages. Memory was similar for sentences in small and in large font, $F(1, 48) = 0.37$, $p = .544$, $\eta^2 = .01$, and better for related sentences than for unrelated sentences, $F(1, 48) = 89.35$, $p < .001$, $\eta^2 = .65$. The interaction was not significant, $F(1, 48) = 0.01$, $p = .905$, $\eta^2 < .01$.

Discussion

In general, Experiment 1 showed that perceptual information is used as a cue for metamemory judgments with sentences, even though perceptual information did not affect memory performance. In other words, the presence of grammatical, syntactical, and semantic cues did not stop participants from using font size to make JOLs, suggesting that participants do not replace font size with linguistic cues when studying sentences. Our results also extended the effect of perceptual information on JOLs and restudy decisions to materials other than words or word pairs. In Experiment 2, we further tested our hypotheses with pairs of longer sentences.

Experiment 2

In Experiment 2, we presented pairs of sentences, and each one was longer than those used in Experiment 1. In each pair, the first sentence acted as the cue and the second as target.

Method

Participants and design Sixty undergraduate students (57 females; age $M = 20.98$, $SD = 3.89$) participated in the experiment in exchange for course credit. We used the same design as in Experiment 1, with 30 participants each allocated to the related and unrelated conditions.

Materials. The related and unrelated pairs of sentences were obtained from Myers, Shinjo, and Duffy (1987), consisting of 32 sets of five sentences. Each of their sets included four different first sentences and one second sentence, with a varying degree of relatedness between them. For our experiment we selected 30 sets, and within each set, the sentence with the highest association (e.g., BRichard was invited to Karen's birthday party^) and the lowest association (e.g., BRichard left home early after eating breakfast^) with the second sentence (e.g., BHe spent a whole day shopping for the present^).

Procedure. We followed the same general procedure as in Experiment 1, with three exceptions. First, the two sentences were presented on different screens. Second, to allow enough time to read each sentence, we followed Dunlosky, Baker, Rawson, and Hertzog (2006), who had also used the materials of Myers et al. (1987) and presented each sentence for 400 ms/ word, range [2,400 ms, 5,600 ms]. Finally, restudy decisions were not collected in this study

Coding. A coding scheme like the one in Experiment 1 was used. We first identified the number of units of information in the target sentence (e.g., *He spent* [1] *a whole day* [2] *shopping* [3] *for the present* [4]). Then, the coding followed that in Experiment 1

Results

Judgments of learning The main descriptive statistics are presented in Table 1. We replicated the effect of relatedness on metamemory judgments, by finding higher JOLs for related than for unrelated sentences, $F(1, 58)= 12.39$, $p = .001$, $\eta_p^2 = .18$. The JOLs were only marginally higher for large than for small font size, $F(1, 58) = 3.93$, $p = .052$, $\eta_p^2 = .06$. The interaction was not significant, $F(1, 58) = 0.81$, $p = .373$, $\eta_p^2 =.01$.

Bayesian analyses showed anecdotal support for no difference between the JOLs for large and small font size in the full sample, BF = 0.88. This result, although inconclusive, contrasts with the extreme evidence found in Experiment 1. When both groups were analyzed separately, the results also showed anecdotal support for no differences as a function of the font size, BF = 0.65 for semantically related and BF = 0.34 for semantically unrelated sentences.

Memory performance The main statistics are presented in Table 2. Memory did not vary with font size, $F(1, 58) = 1.14$, $p = .290$, $\eta_p^2 = .02$, and was better for semantically related than for unrelated sentences, $F(1, 58) = 9.93$, $p =.003$, $\eta^2 = .15$. The interaction was not significant, $F(1, 58)= 0.11$, $p = .743$, $\eta^2 < .01$.

Discussion

Experiment 2 showed a marginal main effect of font size and anecdotal BFs below 1. These results suggest that with pairs of sentences, font size has a limited effect on JOLs, but the results are not conclusive and do not allow us to completely reject any effect. However, they do show that the effect of font size on JOLs was greatly reduced from Experiment 1 to 2 (from $\eta^2 = .26$ to .06, and from BF = 147.50 to 0.88). A comparison between the font-size effects (computed as the JOLs for large items minus the JOLs for small items) in Experiments 1 and 2 showed that the effect was larger in Experiment 1, $t(108) = 2.14$, $p = .035$, $d_{unb} = 0.41$.

We proposed that linguistic cues in a sentence might reduce the font-size effect. The results for semantic relatedness in Experiments 1 and 2 strongly suggest that semantic content is not responsible for the reduction in effect size. Other linguistic factors that might be stronger with pairs of longer sentences could be responsible (but see the Discussion of Experiment 4 regarding this point), but there was one other factor that varied between Experiments 1 and 2:

In Experiment 2, participants had to process and remember more information than in Experiment 1. Could the amount of information explain the reduced font-size effect? Experiment 2 allowed us to test this idea at a descriptive level. Its target sentences included three to five to-be-remembered units of information (e.g., three units: *He decided* [1] *not to run* [2] *for re-election* [3]; five units: *He spent* [1] *a sleepless* [2] *night*

[3] *thinking* [4] *about her* [5]). The font-size effect (JOL for large font minus JOL for small font) showed a tendency to decrease as the number of units of information increased: with three units, $M = 1.29$; with four units, $M = 0.98$; and with five units, $M = -0.19$.[1] After this result, in Experiment 3 we tested the idea that the longer the to-be-remembered information, the lower the font-size effect on JOLs. To this end, we manipulated the length of the sentences within a single experiment.

Experiment 3

In Experiment 3, we presented both short and long sentences, to test the hypothesis that the amount of information moderates the effect of font size on JOLs. We expected to replicate the font-size effect with short sentences but not with long sentences.

Method

Participants and design Sixty-eight participants (61 females; age $M = 20.50$, $SD = 3.93$ years old) took part in the experiment in exchange for course credit. A 2 (font size: small, large) × 2 (sentence length: short, long) within-subjects design was used.

Materials The 40 unrelated sentences from Experiment 1 were used as short sentences, and we created another set of 40 sentences by adding another four units of information. Unrelated sentences were used because they had shown a numerically higher font-size effect than the related sentences in Experiment 1, thus providing more leeway to study possible causes for the reduced font-size effect. The subject was used as the cue and the rest of the sentence as the target. In the short sentences, the target section included two units of information and three or four words (e.g., *The veterinary sells* [1] *a ring* [2]), and in the long sentences, six units of information and 10 to 13 words (e.g., *The veterinary sells* [1] *a ring* [2] *to buy* [3] *the motorbike* [4] *that he always* [5] *desired* [6]). Following the procedure in Experiment 1, we obtained linguistic frequencies for the target words in the long sentences. Finally, for counterbalancing purposes we divided both the short and long sentences into four groups of ten with matched linguistic frequencies and numbers of letters, words, and units of information in the target section.

Procedure The same general procedure was used as in Experiment 1. Participants were presented with 40 sentences, 20 short and 20 long, half presented in 18-point font and the other half in 48-point font. Both length and font size were counterbalanced. As in Experiment 2, we presented the sentences for a time that varied with the number of words (Dunlosky et al., 2006). Pretesting had shown that 400 ms/ word was too short a time to comfortably read the long sentences, so we added 1 s to the total time. The short sentences were presented for a time that ranged between 3 and 3.4 s, and the long sentences between 5.8 and 6.6 s. The experiment then proceeded as in Experiment 1.

Coding The same coding scheme was used as in Experiments 1 and 2.

Results

See Table 3 for the main descriptive statistics regarding JOLs and memory performance.

Judgments of learning JOLs were higher for short than for long sentences, $F(1, 67) = 87.67$, $p < .001$, $\eta^2 = .57$. Font size was marginally significant, $F(1, 67) = 3.14$, $p = .081$, $\eta_p^2 = .04$, but its effect was qualified by a significant interaction, $F(1, 67) = 7.30$, $p = .009$, $\eta^2 = .10$. For short sentences, JOLs were higher when presented in large than in small font, $t(67) = 3.04$, $p = .003$, $d_{unb} = 0.17$, but for long sentences, the difference was not significant, $t(67) = 0.24$, $p = .813$, $d_{unb} = 0.02$. The Bayesian analysis confirmed the pattern. For short

Table 3 Mean (standard deviation) [95% confidence interval] for judgments of learning and memory performance in Experiment 3

	Small (18-pt)	Large (48-pt)	Total
Judgments of Learning			
Short sentences	31.25 (17.85) [27.01, 35.49]	34.47 (20.39) [29.63, 39.32]	32.86 (18.66) [28.43, 37.29]
Long sentences	17.74 (16.79) [13.74, 21.73]	17.49 (15.38) [13.83, 21.14]	17.61 (15.5) [13.93, 21.29]
Total	24.49 (15.92) [20.71, 28.28]	25.98 (16.39) [22.08, 29.87]	
Memory Performance			
Short sentences	9.32 (11.99) [6.47, 12.17]	8.81 (11.80) [6.00, 11.61]	9.07 (10.32) [6.61, 11.52]
Long sentences	6.94 (8.68) [4.88, 9.00]	5.78 (6.69) [4.19, 7.37]	6.36 (6.71) [4.77, 7.96]
Total	8.13 (9.15) [5.96, 10.31]	7.29 (7.99) [5.39, 9.19]	

sentences there was moderate support for a difference, BF = 8.69, and for long sentences there was moderate support for no difference, BF = 0.14.

Memory performance Participants recalled more information for short than for long sentences, $F(1, 67) = 9.00$, $p = .004$, $\eta_p^2 = .12$. Font size did not affect memory, $F(1, 67) = 1.01$, $p = .318$, $\eta^2 = .01$, and the interaction was not significant, $F(1, 67) = 0.13$, $p = .715$, $\eta^2 < .01$.

Discussion

These results suggest that the kind of cues used to make metamemory judgments vary with sentence length. In this case this variation was positive, because participants stopped using, or relied less on, an undiagnostic cue such as font size. Why did font size affect JOLs for short but not for long sentences? Our initial hypothesis regarding the presence of linguistic factors, again, does not seem to explain this result, since both types of sentences included syntactic, grammatical, and at least a certain amount of semantic information. A possible explanation is that with increasingly longer materials, the cognitive load required for processing them also increases. Cognitive load might carry over and affect the immediate JOLs. Carryover effects, in which the subproducts of processing a task are carried over and affect other tasks, have been used to explain several psychological and other phenomena, for example, the cost of task switching (e.g., Hsieh & Liu, 2005), the effect of cognitive involvement in advertisement processing in radio programs (Martín-Luengo, Luna, & Migueles, 2013), or the impact of competition on prejudice toward outgroups (Sassenberg, Moskowitz, Jacoby, & Hansen, 2007). Alternatively, cognitive load may arise when rating whether one will remember a longer sentence simply because there is more information to deal with. In either case, increased cognitive load might have limited the available re- sources while participants made the JOLs, which in turn might have limited the number of cues used to make them. With this explanation, encoding fluency or beliefs might still have been present, but they might not have been used to make JOLs because participants did not use all the cues available. We tested this explanation in Experiment 4, in which we presented the sentences from Experiment 1 in conditions of either high or low cognitive load.

Experiment 4

Cognitive load is known to affect metamemory judgments. For example, JOLs made under time pressure (one way of reducing cognitive resources) are lower than those without time pressure (Benjamin, 2005). In addition, without time pressure, JOLs are based on retrievability, and consequently are better adjusted to memory performance, whereas under pressure, JOLs are based on familiarity (Benjamin, 2005). Similarly, participants with lower working memory capacity, and thus lower cognitive resources, showed poorer monitoring accuracy than participants with higher working memory capacity, although there was no difference in their JOL magnitudes (Griffin, Wiley, & Thiede, 2008). Finally, resolution is lower in the presence of auditory distraction than under full attention (Beaman, Hanczakowski, & Jones, 2014). Other metamemory measures are also sensi- tive to the manipulations of cognitive load and divided attention (e.g., global JOLs in Barnes & Dougherty, 2007, or feelings of knowing in Sacher, Taconnat, Souchay, & Isingrini, 2009). In general, past research has supported the hypothesis that higher cognitive load impairs monitoring, maybe because it reduces the amount of resources for checking and using the available cues. If this were the case, we expected that if we increased the cognitive load of the task, participants would have fewer resources available to make their metamemory judgments, and thus the chances that they would rely on font size would be lower. In sum, we expected that under high cognitive load, participants would not use font size as a cue to JOLs.

Method

Participants and design Eighty-five participants completed the experiment (80 females; age *M* = 21.25, *SD* = 4.38 years old) in exchange for course credit. One participant was re- moved due to unusually long reaction times during the JOL phase (*M* = 22.6 s, as compared to an average for the rest of the sample below 3 s; see the supplemental materials for a full report of the reaction times for JOLs). Thus, the data from 84 participants were entered into the analyses. In this experiment the unrelated sentences from Experiment 1 were used, and the sample size was determined as what would be enough to find the effect size found earlier in that condition. To find a d_{unb} = 0.45, a power analysis with alpha = .05 and power = .80 showed that 41 participants were needed for each of the between-subjects condition. One extra participant was collect- ed in each group to match the sample in the counterbalanced conditions (sentences presented in small or large font).

A 2 (font size: small, large) × 2 (cognitive load: high, low) mixed design was used, with the first variable manipulated within subjects and the last between subjects. Forty-two participants were allocated to each of the cognitive load conditions. The main measures were JOLs and memory performance, but to further test the effect of the cognitive load manipulation, we also recorded the number of dots correctly remembered in the dot-memory task (see below) and the reaction times for JOLs and for the dot-memory task.

Materials The 40 unrelated sentences from Experiment 1 were used. We selected the unrelated sentences because the effect of font size was larger in that condition, and thus would provide more room to detect a decrease in the effect.

To manipulate cognitive load, we used the dot-memory task (Bethell-Fox & Shepard, 1988) that had been used in previous research for this purpose (e.g., De Neys, 2006; Trémolière, Gagnon, & Branchette, 2016). In this task, participants are shown a 3×3 matrix with dots in a certain number of cells and are asked later to replicate the pattern of dots in an empty matrix. This visual task allowed us to manipulate cognitive load without interfering with the verbal nature of the sentences or the numerical nature of the JOLs. We developed 46 patterns with four dots, based on the most complex patterns of Bethell-Fox and Shepard. Some examples are presented in the supplemental mate- rials (see also Fig. 1). Two patterns were used for training, and four as primacy and recency buffers. These patterns were the same for participants in the high- and low- cognitive load conditions

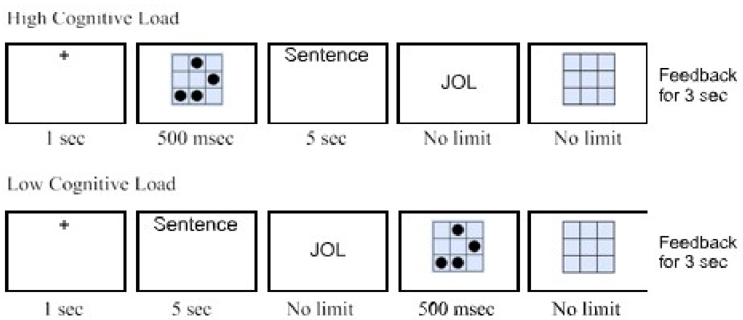

Fig. 1 Sequences of tasks for each trial in the high- and low-cognitive load conditions in Experiment 4

Procedure The procedure was the same as in Experiment 1, with several modifications. Figure 1 outlines the procedures for both the high- and low-cognitive load conditions. Under high cognitive load, in each trial participants saw a fixation cross for 1 s, the study matrix with four dots for 500 ms, the sentence for 5 s, the JOL with no time limit, and an empty test matrix for participants to indicate the positions of the dots, also with no time limit. To encourage participants to remember the dots, feedback on their performance was shown, indicating the number of dots correctly recalled, for 3 s. In the low-cognitive load condition the tasks were the same, but the study matrix was presented after the JOL. Thus, the only difference between the high- and low-cognitive load conditions was that in the former condition participants had to keep the positions of the dots in working memory while reading the sentence and rating the JOLs, whereas in the latter condition this was not necessary. The experiment then followed as in Experiment 1. Participants completed two training trials in order to familiarize themselves with the task.

Results

See Table 4 for the main descriptive statistics. The number of dots correctly remembered in the dot-memory tasks and the reaction times for JOLs and the dot-memory task were not directly related to our objectives, and thus a full report is presented in the supplemental materials. In general, the results from these measures supported the idea that participants com- plied with the dot-memory task.

Judgments of leaning JOLs were higher for items presented in large font than for items in small font, *F*(1, 82) = 7.10, *p* =.009, η_p^2 = .08, and cognitive load did not affect the JOLs, *F*(1, 82) = 1.14, *p* = .288, η_p^2 = .01. The interaction was marginally significant, *F*(1, 82) = 3.74, *p* = .057, η^2

= .04. Analysis of the interaction showed that, as expected, with low cognitive load the JOLs were higher for items in large than in small font, $t(41) = 2.62$, $p = .012$, $d_{unb} = 0.17$, but that with high cognitive load there was no difference, $t(41) = 0.76$, $p = .450$, $d_{unb} = 0.02$. Bayesian analyses confirmed this pattern, showing moderate support for a difference with low cognitive load, BF = 3.36, and moderate support for no difference with high cognitive load, BF = 0.22.

Memory performance Font size did not affect memory performance, $F(1, 82) = 0.55$, $p = .460$, $\eta^2 = .01$. Memory performance was numerically higher for the low-cognitive load than for the high-cognitive load group, but the difference was not significant, $F(1, 82) = 2.65$, $p = .108$, $\eta^2 = .03$. The interaction was also not significant, $F(1, 82) = 0.10$, $p = .751$, $\eta^2 < .01$.

Discussion

Experiment 4 showed that increased cognitive load prevents participants from using font size to make JOLs. This result may explain why longer sentences, such as those used in Experiment 3, or definitions (Ball et al., 2014) do not show the typical font-size effect found with words, word pairs, or short sentences. With longer materials the cognitive load in- creases and the amount of resources available to process cues for metamemory judgments decreases, so that perceptual in- formation is no longer used. What this research cannot answer is whether our participants under cognitive load relied on few- er cues or replaced perceptual information with another cue. Considering the latter alternative, we might suspect that participants did not replace font size with linguistic cues to make their JOLs. Had they done so, we would have expected to find higher gammas under cognitive load, because linguistic information is likely more diagnostic of later retrieval. Instead, the gammas reported in the supplemental materials showed marginally lower resolution under cognitive load, consistent with previous studies Experiment 4 also indirectly suggested that increased grammatical and syntactical cues in long sentences may not have been the reason why the font-size effect was reduced in Experiment 2, and eliminated in Experiment 3. Long sentences may include more linguistic information, but in Experiment 4 the same sentences with the same linguistic characteristics showed different font-size effect, depending on cognitive load. In general, it seems that the cognitive load associated with a higher amount of to-be-remembered information offers a simpler explanation for the elimination of the font-size effect with longer sentences.

We found that JOLs and memory were not affected by cognitive load. One explanation is that our manipulation may have increased the cognitive load in the participants in the high-cognitive load group, as expected, but that it also affected the low-cognitive load group to a lesser extent. JOLs and memory were numerically in the expected direction—for example, lower JOLs and memory for the high- cognitive load group—but the difference in cognitive load between the conditions may have not been enough to affect these measures.

General discussion

In this research we studied the font-size effect, the phenome- non in which items presented in larger fonts are judged as being more memorable. We tested whether this effect extends to other materials, such as sentences with different amounts of information. Our results showed that for simple subject–verb– object sentences, font size is used as a cue for JOLs, but as sentence length increases, the effect seems to progressively decrease (Exp. 2) until it disappears (Exp. 3). A direct com- parison of short and long sentences confirmed that font size affects JOLs for short but not for long sentences. Experiment 4 showed that the difference might be explained by the in- creased cognitive load required for processing long sentences. This may be due to a carryover effect from the processing of the sentences to the immediate

JOLs, or because the task of rating a JOL for a long sentence inherently consumes more cognitive resources. In either case, our results suggest that the font-size effect reported so far is limited to certain types of materials, such as words and word pairs or simple subject– verb–object sentences, but it is highly unlikely that it will affect learning in more ecological scenarios.

Cognitive load may have eliminated the effect of perceptual information on JOLs by means of at least two mechanisms. First, it may have limited the resources available to process cues that could be used to reduce the uncertainty of the JOLs. Recent research has shown that most people can integrate at least four cues when making JOLs (Undorf, Söllner, & Bröder, 2018). However, with fewer resources available, participants may be limited in the number of cues they can integrate. If this is the case, font size may be among the first cues to be discarded. This idea is consistent with the explanation of Ball et al. (2014) that different conditions may trigger different weights of metamemory cues. How do participants decide which cues to keep using and which ones to disregard? A tentative explanation is that they continue to use cues assigned higher weights, perhaps because these cues are subjectively considered more diagnostic of future performance, and that they more rapidly disregard cues assigned lower weights. Future research should test this hypothesis.

Second, the increased cognitive load may have complicated the development of a plausible belief, or the access to a preexisting one, about the effect of font size on memory. The analytic processing theory proposes that making a JOL triggers an analytic problem-solving mode in which participants search for cues to reduce the uncertainty associated with the task (Mueller & Dunlosky, 2017; Mueller et al., 2016). Since participants have limited knowledge of the factors that affect future memory performance, they search for any cue available and develop online beliefs and hypotheses about the effects of these cues on performance. If participants notice a variation in the items, but they do not have enough resources to develop a belief about how that variation is related to memory, it then will not affect JOLs. In support of the relevance of developing a plausible belief, Mueller and Dunlosky found that the color in which words were presented only affected JOLs when there was a plausible explanation for its effect on memory. When the explanation was not presented, the beliefs were not developed, and color did not affect JOLs.

At the applied level, this research suggests that the font size in which materials are presented may not be particularly relevant in most educational contexts. This is not to suggest that the phenomenon is not of interest. This research has been the first to show that short sentences are rated as more memorable when they are presented in larger rather than in smaller font, but also that when long sentences are presented the effect of perceptual information on JOLs is eliminated, and that cognitive load is a good explanation for this difference. These results suggest a complex picture of the kind of cues used for metamemory judgments with different materials, and that we should be cautious when generalizing from one type of mate- rial to another.

Author note This study was conducted at the Psychology Research Centre (PSI/01662), University of Minho, and sup- ported by the Portuguese Foundation for Science and Technology and the Portuguese Ministry of Science, Technology and Higher Education through national funds, and co-financed by FEDER through COMPETE2020 under the PT2020 Partnership Agreement (POCI-01-0145-FEDER- 007653). This article was also funded by the

Basic Research Program at the National Research University Higher School of Economics and supported within the framework of a subsidy by the Russian Academic Excellence Project B5-100.^ Thanks to Sara Cadavid and Filipa Gonçalves for their help with data collection.

References

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied*, *20*, 336–348.

Barnes, K. A., & Dougherty, M. R. (2007). The effect of divided attention on global judgment of learning accuracy. *American Journal of Psychology*, *120*, 347–359.

Beaman, C. P., Hanczakowski, M., & Jones, D. M. (2014). The effects of distraction on metacognition and metacognition on distraction: Evidence from recognition memory. *Frontiers in Psychology*, *5*, 439.

Benjamin, A. (2005). Response speeding mediates the contributions of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review*, *12*, 874–879.

Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Condition*, *42*, 1417–1433.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation be- tween metamemory and memory performance. *Memory & Cognition*, *41*, 897–903.

Bethell-Fox, C., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 12–23.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. New York: Routledge.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433.

Dunlosky, J., Baker, J. M. C., Rawson, K. A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*, 390– 400.

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*, 680–693.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*, 93–103.

Halamish, V. (2018). Can very small size enhance memory? *Memory & Cognition*, *46*, 979–993.

Hsieh, S., & Liu, L. C. (2005). The nature of switch cost: Task set con- figuration or carry-over effect? *Cognitive Brain Research*, *22*, 165– 175.

Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgment of learning. *Memory*, *24*, 383–398.

JASP Team. (2017). JASP (Version 0.8.1.1) [Computer software].

Retrieved from https://jasp-stats.org

Katzir, T., Hershko, S., & Halamish, V. (2013). The effect of font size on reading comprehension on second and fifth grade children: Bigger is not always better. *PLoS ONE*, *8*, e74061.

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 923–954.

Li, F., Xie, R., Li, X., & Li, W. (2015). The influence of perceptual information on control processes involved in self-regulating learning: Evidence from item selection. *Psychonomic Bulletin & Review*, *22*, 1007–1013.

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*. https:// doi.org/10.1080/17470218.2017.1343362

Luna, K., Martín-Luengo, B., Shtyrov, Y., & Myachykov, A. (2016). Judgments of learning for words in vertical space. *Frontiers in Psychology*, *7*, 1894.

Martín-Luengo, B., Luna, K., & Migueles, M. (2013). Memory for radio advertisements: Effects of programme and typicality. *Spanish Journal of Psychology*, *16*, E80.

McDonough, I. M., & Gallo, D. A. (2012). Illusory expectations can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 391–404.

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.

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about mem- ory? *Quarterly Journal of Experimental Psychology*, *69*, 781–799.

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 of reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453– 465.

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*, 207–213.

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.

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

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*, 550–554.

Sacher, M., Taconnat, L., Souchay, C., & Isingrini, M. (2009). Divided attention at encoding: Effect on feeling-of-knowing. *Consciousness and Cognition*, *18*, 754–761.

Sassenberg, K., Moskowitz, G. B., Jacoby, J., & Hansen, N. (2007). The carry-over effect of competition: The impact of competition on prejudice towards uninvolved outgroups. *Journal of Experimental Social Psychology*, *43*, 529–538.

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo [Procura-PALavras (P-PAL): A new measure of word frequency for contemporary European Portuguese]. *Psicologia: Reflexão e Crítica*, *27*, 1–14. Retrieved from http://p-pal.di.uminho.pt 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*, 1000–1011.

Trémolière, B., Gagnon, M. E., & Blanchette, I. (2016). Cognitive load mediates the effect of emotion on analytical thinking. *Experimental Psychology*, *63*, 343–350.

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1264–1269.

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*, 647–658.

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, *46*, 507–519.

Undorf, M., & Zimdahl, M. F. (2018). 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*. 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.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). Difficulty of mending metacognitive illusions: A priory theories, fluency effects, and mis- attributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*, 918–933