Journal of Memory and Language 99 (2018) 99–110

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Perceptual fluency affects judgments of learning: The font size effect

Chunliang Yang *, Tina S.-T. Huang, David R. Shanks

University College London, United Kingdom

ARTICLE INFO

Article history:
Received 18 April 2017
Revision received 15 November 2017
Available online 7 December 2017

Keywords:
Perceptual fluency
JOLs
Font size effect
Stimulus size
Continuous identification task

ABSTRACT

The font size effect on judgments of learning (JOLs) refers to the fact that people give higher JOLs to large than to small font size words, despite font size having no effect on retention. The effect is important because it spotlights a process dissociation between metacognitive judgments about memory and memory performance itself. Previous research has proposed a fluency theory to account for this effect, but this theory has been contradicted by a recent study which found no difference in response times (RTs) – and hence fluency – in a lexical decision task between large and small words (Mueller, Dunlosky, Tauber, & Rhodes, 2014). In the current research, we further tested the fluency theory by employing a continuous identification (CID) task in Experiment 1 and by explicitly comparing the CID and lexical decision tasks in Experiment 2. We show that lexical decision is an inappropriate instrument for measuring differences in perceptual fluency. The CID task, in contrast, provides direct evidence that the stimulus size effect on JOLs is substantially mediated by perceptual fluency. Experiment 3 found that fluency is at least as important as beliefs about font size in contributing to the font size effect on JOLs.

Introduction

The font size effect on judgments of learning (JOLs; i.e., estimates of the likelihood that a given item will be remembered at a future memory test) was originally reported by Rhodes and Castel (2008). They instructed participants to study words in large (48-point) or small (18-point) font sizes. After studying each word, participants made a JOL to predict the likelihood they would remember that word. Participants gave significantly higher JOLs to large than to small words, yet at a later test, recall performance was equivalent for large and small words. The font size effect on JOLs is robust and has been replicated dozens of times (e.g., Ball, Klein, & Brewer, 2014; Besken, 2016; Hu, Liu, Li, & Luo, 2016; Hu et al., 2015; Kornell, Rhodes, Castel, & Tauber, 2011; Li, Xie, Li, & Li, 2015; Miele, Finn, & Molden, 2011; Mueller, Dunlosky, Tauber, & Rhodes, 2014; Price & Harrison, 2017; Price, McElroy, & Martin, 2016; Susser, Mulligan, & Besken, 2013). The effect is important because JOLs determine individuals' study strategies (Metcalfe & Finn, 2008; Yang, Potts, & Shanks, 2017b), and hence any process dissociation between JOLs and actual memory performance can potentially induce inefficient study (e.g., Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013; Yang, Sun, & Shanks, 2017; Yang et al., 2017b). For example, an individual might study a textbook chapter for more or less time depending on whether it is written in a small or large font, even though font size is unlikely to affect retention of the chapter's content. From a theoretical perspective, understanding such process dissociations is an important step in developing interventions to improve individuals' study strategies.

Two theories have been proposed to account for the font size effect on JOLs. The first explanation is a belief theory, which postulates that people hold *a priori* beliefs that large words are easier to remember or more important than small words, and that they incorporate these beliefs into their JOLs. Research has found that perceived importance can moderate people's JOLs (Castel, 2007). Mueller et al. (2014) found that some people believe that large words are more important than small words, and Rhodes and Castel (2008) proposed that participants might believe that a large font signals the importance of a study item within the context of an experiment. Therefore, it is possible that the difference in perceived importance between large and small words may produce the font size effect on JOLs (Rhodes & Castel, 2008). Mueller et al. (2014) also found that some people believe large words are easier to remember, and therefore suggested that people apply this belief in forming their JOLs (Mueller & Dunlosky, 2017). Moreover, Hu et al. (2015) found that the font size effect on JOLs is significantly predicted by variability in people's beliefs about the difficulty of remembering large and small words. Collectively, these findings support the belief theory (based either on beliefs about importance

* Corresponding author at: Division of Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom.

E-mail address: chunliang.yang.14@ucl.ac.uk (C. Yang).

https://doi.org/10.1016/j.jml.2017.11.005
0749-596X/

or about ease of remembering) as an account for the font size effect on JOLs.

The second explanation is a fluency theory, which postulates that large words are processed with greater perceptual fluency than small words. The experience of fluency during encoding produces a subjective *feeling-of-knowing*, and this subjective feeling acts as a basis for assessments about learning status (Koriat & Bjork, 2006; Koriat & Ma'ayan, 2005; Mueller, Tauber, & Dunlosky, 2013; Undorf, Zimdahl, & Bernstein, 2017). Previous studies have supplied convincing evidence that greater processing fluency produces higher JOLs – a fluency effect on JOLs (Ball et al., 2014; Besken & Mulligan, 2013; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Magreehan, Serra, Schwartz, & Narciss, 2016; Undorf et al., 2017; Yang et al., 2017b).

Only two studies, though, have directly examined the role of fluency in the font size effect on JOLs. The first was conducted by Rhodes and Castel (2008). In their Experiment 6, some words were presented in a standard format (e.g., *computer*) and others in a format with alternating lowercase and uppercase letters (e.g., *gArDeN*). Rhodes and Castel (2008) obtained a font size effect on JOLs in the standard format condition but not in the alternating format condition. They proposed that differences in perceptual fluency between large and small words were disrupted in the alternating format condition. However, Mueller et al. (2014) argued that Rhodes and Castel's (2008) Experiment 6 cannot provide unequivocal evidence to support the fluency theory, and that prior beliefs can equally well explain the results: Participants may simply not believe that large but alternating font words are easier to remember than small alternating font words.

Mueller et al. (2014) conducted a further study to test the fluency theory by employing a lexical decision task in their Experiment 1. Words (e.g., *chicken*) and non-words (e.g., *arage*) were sequentially presented in large or small font sizes. Participants were instructed to decide, as quickly and accurately as they could, whether the presented item was a word or a non-word. Mueller et al. (2014) found no difference in response times (RTs) between large and small words, and hence suggested that "processing fluency, as measured by the lexical decision task, is not mediating the font-size effect" (p. 4).

This finding is surprising because prior to Mueller et al.'s (2014) study, the general consensus amongst researchers was that perceptual fluency does underlie the font size effect on JOLs, and indeed many researchers had offered the font size effect on JOLs as evidence that perceptual fluency can affect JOLs (e.g., Bjork, Dunlosky, & Kornell, 2013; Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Kornell et al., 2011; Miele et al., 2011; Rhodes & Castel, 2008). It is important to note that Mueller et al. (2014) did not completely reject the fluency theory. Instead, they suggested that their results were inconsistent with the fluency theory and they encouraged future research to further explore the theory (p. 9). However, after Mueller et al.'s (2014) study was published, researchers started to acknowledge that fluency may play no role in the font size effect on JOLs (e.g., Ball et al., 2014; Finn & Tauber, 2015; Li, Jia, Li, & Li, 2016; Magreehan et al., 2016; Mueller & Dunlosky, 2017; Mueller, Dunlosky, & Tauber, 2016; Susser, Jin, & Mulligan, 2016; Susser, Panitz, Buchin, & Mulligan, 2017; Undorf et al., 2017). Taking a more neutral position, Hu et al. (2015) claimed that "Although Mueller et al. (2014) suggest that fluency does not differ... There may be other types of fluency that differ significantly between large and small words" (p. 10).

Assessing the evidence against the fluency theory

There are at least three possible reasons for the lack of a difference in RTs between large and small words in Mueller et al.'s (2014) Experiment 1. The first, as proposed by Mueller et al. (2014), is that there is truly no difference in perceptual fluency between large and small words. Secondly, their null result might be a false negative, because the number of trials (18 large and 18 small words) and sample size (31 participants) might have combined to render their experiment underpowered. It is well-known that small sample size and number of trials can lead to false negative results (Vadillo, Konstantinidis, & Shanks, 2016). The third possibility concerns the research method Mueller et al. employed, specifically, their use of RTs obtained from a lexical decision task as an index of perceptual fluency. The lexical decision task is complex (Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015): Participants need to read or identify the letter string first, judge whether it is a word or a non-word, and then select which button to press to indicate their response before the judgment RT is recorded. Participants may check the letter string letter-by-letter, and their lexical decisions may be conservative and time-consuming. Therefore, there could be considerable noise in the RTs obtained from the lexical decision task. Access to word meaning is also assumed to be involved in the lexical decision task (Chumbley & Balota, 1984). Consequently, RTs derived from Mueller et al.'s (2014) Experiment 1 might be driven by semantic processing *in addition* to perceptual processing of the words, and thus it is unclear to what extent their findings contradict accounts claiming that *perceptual* fluency underlies the font size effect on JOLs. In short, lexical decision may be a poor tool for measuring variations in perceptual fluency.

Mueller et al. (2014) tested the fluency theory more indirectly by measuring study time allocation in their Experiment 2. Participants were allowed to spend as much time as they wanted to study each word. Mueller et al. (2014) hypothesized that participants would spend less time studying large compared to small words if large words are processed more fluently than small words. However, they observed no difference between study times allocated to large and small words, and proposed that "the lack of an effect of font size on study time allocation is inconsistent with the hypothesis that encoding fluency is responsible for the font-size effect on JOLs" (p. 5).

But again, this result does not provide strong motivation to reject the fluency theory because, besides fluency, many other factors could have affected participants' study time allocation (e.g., motivation, curiosity). Participants might believe that large words are more important than small words (Mueller et al., 2014; Rhodes & Castel, 2008), and allocate more time to them accordingly (Noh, Yan, Vendetti, Castel, & Bjork, 2014). A fluency advantage for large words (leading them to be studied for less time) may have operated in opposition to a belief that large words are important (leading them to be studied for longer), thus contributing to the overall null result. Yang, Potts, and Shanks (2017a) found that participants decreased their study times across a study phase when they were allowed to spend as much time as they wanted to study each item (e.g., Euskara-English word pairs in Yang et al.'s Experiment 1 and face-name pairs in their Experiment 2), again implying that self-regulated study time allocation can be affected by other factors besides fluency.

Moreover, recent research has found that in some situations self-regulated study time allocation is not a sensitive measure of fluency. For example, Witherby and Tauber (2017) found that participants responded faster to concrete (e.g., *apple*) than to abstract (e.g., *idea*) words in a lexical decision task, but there was no difference in study times between concrete and abstract words when participants were allowed to spend as much time as they wanted to study them. Therefore, Mueller et al.'s (2014) Experiment 2 cannot be taken as providing indirect evidence against the fluency theory because self-regulated study time allocation can be affected by many other factors besides fluency, and is an insensitive measure of fluency. Overall, Mueller et al.'s (2014) Experiments 1 and 2 fall short of providing compelling evidence against the fluency theory and it remains unclear whether perceptual fluency contributes to the font size effect on JOLs.

After Mueller et al.'s (2014) study, researchers raised two other important questions. The first question is whether – moving beyond the standard font size manipulation – there exists evidence that perceptual fluency can affect JOLs (e.g., Besken, 2016; Frank & Kuhlmann, 2016; Price & Harrison, 2017; Susser et al., 2016; Undorf et al., 2017). Susser et al. (2016) addressed this question by employing an identity-priming paradigm. Participants were asked to name and make item-by-item JOLs for words (e.g., *phone*) which were preceded by either matched (*phone*) or mismatched (e.g., *doctor*) primes. Susser and colleagues found that matched priming produces greater perceptual fluency than mismatched priming, as reflected by a difference in naming latencies. They also found that higher JOLs were given to matched words than to mismatched words – a priming effect on JOLs. But a mediation analysis revealed that naming latencies did not mediate the priming effect on JOLs. Thus Susser and colleagues concluded (p. 660) that "effects of perceptual fluency on JOLs do not exist."

On the other hand, Undorf et al.'s (2017) results contradicted Susser et al.'s (2016) conclusion. Undorf et al. (2017) instructed participants to identify stimuli (objects, faces, or words in different experiments) and make item-by-item JOLs. For each stimulus, 30 images were created in which the object became progressively larger and larger: Image size increased with image number. In a slow clarification condition, images were presented for 1 s each, in the following number sequence: 1, 2, 3 ... 30; in a fast condition the images were presented in the sequence: 1, 3, 5 ... 29. Thus the maximum image size occurred after 15 image presentations in the fast condition and after 30 images in the slow condition. The results showed that stimuli were identified faster in the fast condition than in the slow condition, and the size level at which a stimulus was identified was larger in the fast condition than in the slow condition. The results also showed that higher JOLs were given to stimuli in the fast condition than in the slow condition – a clarification speed effect on JOLs. Most importantly, Undorf et al. (2017) found that identification RTs significantly mediated the clarification speed effect on JOLs (for similar findings, see Besken, 2016). Evidently, Undorf et al.'s (2017) and Susser et al.'s (2016) results support mutually conflicting conclusions. Therefore, it is still controversial whether perceptual fluency can affect JOLs and more research is needed to explore this question.

The second question is whether perceptual fluency underlies the stimulus size effect on JOLs. For example, after Mueller et al.'s study, Undorf et al. (2017) noted that "there is no evidence that perceptual fluency contributes to the stimulus size effect on JOLs" (p. 294), and they further investigated this question by manipulating stimulus clarification speed. Nonetheless, Undorf et al.'s (2017) study cannot provide direct evidence that perceptual fluency underlies the stimulus size effect on JOLs because it manipulated the rate of change in the sizes of stimuli, rather than directly manipulating stimulus size. All stimuli in their study had the same (dynamically-changing) size, except that the identified size was determined by the participants' response. For example, on a slowly-identified trial, the stimulus size displayed on screen would be larger *at the moment* of identification relative to the stimulus size displayed on screen if the participant could identify the stimulus more rapidly. This means that the relationship between identification RTs and JOLs is confounded by the different levels of stimulus size at which the words were identified across the two clarification conditions.

Undorf et al. suggested that the greater JOLs in the fast clarification condition relative to the slow condition could be mediated by greater perceptual fluency (i.e., shorter RTs). However, since stimulus identifications tended to be made at a larger size in the fast condition than in the slow condition, an alternative explanation for the aforementioned finding is that the higher JOLs observed in the fast condition occurred as a direct consequence of their larger stimulus size at identification. Similarly in the slow condition, for a given trial with a fast identification RT, stimulus size would have been smaller at the moment of identification compared to the size corresponding to the same RT if the trial had been in the fast condition. Direct evidence should demonstrate that a large (versus small) stimulus size, which is processed with greater perceptual fluency, produces higher JOLs, and that perceptual fluency mediates that stimulus size effect on JOLs. This demands an explicit experimental manipulation of stimulus size – something which was not part of Undorf et al.'s method. Therefore, despite Undorf et al.'s (2007) demonstration of perceptual fluency contributing to the effect of stimulus enlargement speed on JOLs, there is still no direct evidence that perceptual fluency underlies the stimulus size effect on JOLs when stimulus sizes are pre-determined and stationary.

To summarise, lexical decision and self-regulated study time allocation are the two most widely-used methods to measure fluency in metamemory research (e.g., Ball et al., 2014; Jia et al., 2015; Mueller et al., 2013, 2014, 2016; Undorf & Erdfelder, 2015; Witherby & Tauber, 2017). By employing these two methods, Mueller et al. (2014) found no difference in fluency between large and small words. However, as discussed, the null outcome could have been produced by alternative factors. Following Muller et al.'s study, researchers examined whether perceptual fluency can affect JOLs. By employing different experimental methods and types of stimuli, Undorf et al. (2017) and Susser et al. (2016) observed different results supporting mutually conflicting conclusions. Undorf et al. (2017) investigated whether perceptual fluency underlies the stimulus size effect on JOLs by manipulating stimulus classification speed, but their study cannot provide conclusive evidence because they did not experimentally manipulate processing fluency independently of stimulus size at the point of classification.

Motivation of the current research

The main aim of the current research is to further test whether perceptual fluency underlies the font size effect on JOLs by employing a new experimental paradigm – a continuous identification (CID) task. The CID task, a variety of perceptual identification task (Sanborn, Malmberg, & Shiffrin, 2004), is a method frequently used to measure fluency in memory (e.g., repetition priming) research (e.g., Berry, Shanks, Speekenbrink, & Henson, 2012; Stark & McClelland, 2000; Ward, Berry, & Shanks, 2013). However, to our knowledge, no previous metamemory research has employed the CID task to measure fluency.

In the CID task, a word and a mask are alternately presented, with the presentation time of the word increasing and the presentation time of the mask decreasing in each fixed-duration cycle (see Fig. 1). Across cycles, the word gradually becomes clearer and easier to perceive as the stimulus-to-mask ratio increases via progressive demasking. Participants' only task is to identify the presented word as quickly and accurately as possible, and their identification RT is used as an index of fluency. On the basis of prior research (Ferrand et al., 2011; Grainger & Segui, 1990), we anticipated that the CID task would be more sensitive than lexical decision to variations in perceptual fluency. By employing the CID task, we explored whether there is a difference in perceptual fluency between large and small words, and whether perceptual fluency mediates the font size effect on JOLs. If both answers are affirmative, the current research will support the fluency theory as an account for the font size effect on JOLs, which will also imply that perceptual fluency can affect JOLs. At the same time, through directly manipulating font size, the current research will provide

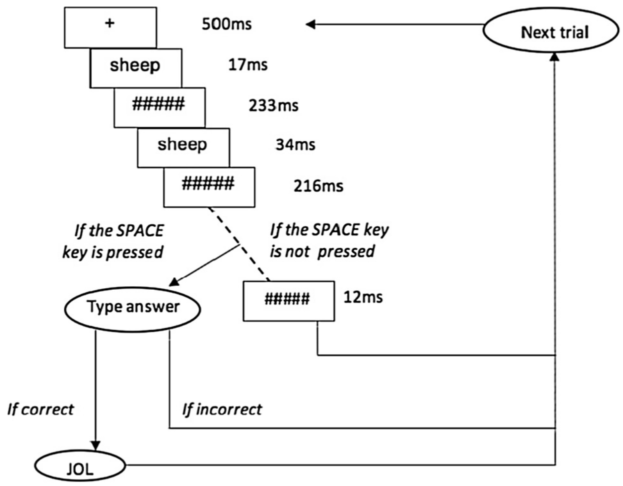

Fig. 1. Experimental design schema of Experiment 1's study phase. For each identification trial, a word and a mask were alternatively presented in the same font size, which was randomly decided by the computer. Participants' task was to identify each word as soon and accurately as they could.

firm evidence about whether or not perceptual fluency underlies the stimulus size effect on JOLs.

Experiment 1

In Experiment 1, we employed the CID task to investigate whether perceptual fluency underlies the font size effect on JOLs. As discussed above, the small number of trials in Mueller et al.'s (2014) Experiment 1 might have contributed to their null result. Therefore we increased the number of trials to 100.

Method

Participants

We conducted a power analysis using G∗power to determine the required sample size (Faul, Erdfelder, Lang, & Buchner, 2007). By using the effect sizes from previous studies in which Cohen's *d*s ranged from 0.58 to 0.74 (Hu et al., 2016; Rhodes & Castel, 2008), we found that about 22–34 participants are required to observe a significant (α = 0.05) font size effect on JOLs at 0.9 power. Therefore, we recruited 28 participants, with a mean age of 22.21 (*SD* = 7.10) years, 21 females, from the University College London (UCL) participant pool.[1] Participants reported normal or corrected-to-normal vision, received £3 or course credit as compensation, and were tested individually in a single sound-proofed cubicle. In all the experiments reported here, participants' first language was English, and ethical approval was provided by the UCL Department of Experimental Psychology.

Materials

The principal stimuli were 110 monosyllabic English nouns selected from the MRC Psycholinguistic Database (Coltheart, 2007). Each word had 5 letters, a Kučera-Francis Frequency score of 3–50, a Concreteness score of 300–670, and an Imageability score of 300–600. We strictly controlled the letter length to 5 to ensure that the mask (#####) would completely cover each word. Ten words were used for practice and the other 100 were used in the main experiment. To prevent any potential item effects, the program randomly selected half the words to be presented in large and the other half in small font sizes for each participant, and the presentation sequence of words was also randomly determined. Stimuli were displayed on an LCD monitor (resolution: 1920 × 1080 at 60 Hz) via the MATLAB *Psychtoolbox* package (Kleiner, Brainard, & Pelli, 2007).

Design and procedure

The experiment involved a within-subjects design (font size: large/small). Participants were asked to identify 100 English words as quickly and accurately as they could, and to remember them for a later memory recall test. They were informed that at a later memory test they would be asked to recall as many words as possible.

There were three tasks: study, distraction, and test. In the study task, a cross was presented at the center of the screen in a medium font size (30-point) for 500 ms. Then a word and a mask were alternately presented in Arial font as in Mueller et al. (2014), and using the same font sizes (48 or 18-point). For each identification trial, there were 14 cycles in total. At the first cycle, the word was presented for 17 ms followed by the mask for 233 ms. At the second cycle, the word was presented for 34 ms, followed by the mask for 216 ms. Thus across cycles, the presentation duration of the word increased in 17 ms steps with the duration of the mask decreasing in 17 ms steps. The word-mask cycle was repeated until participants responded or until the end of the 14th cycle. Participants were instructed to press the SPACE key as soon as they could identify the word. If they did not respond before the end of the 14th cycle, the next identification trial began. If they responded, the word and mask disappeared, and participants typed in their answer (the word) via the keyboard. Then the computer automatically checked whether or not their answer was correct. If correct, a slider ranging from 0 (I'm sure I'll not remember it) to 100 (I'll definitely remember it) was presented at the center of the screen for participants to predict the likelihood that they would remember that word at a later test. If incorrect, the next trial began (see experiment design schema of the study phase in Fig. 1). After participants identified all 100 words, they were asked to solve as many math problems (e.g., *24 + 32 = ___?*) as they could in 2 min. Then they were instructed to recall as many words as possible in any order and to type their answers. Their answers were shown on screen in a medium font size (30-point).

All experimental instructions were presented in a medium font size. Participants were told to place their left hand above the SPACE key while they used the mouse to make JOLs, which enabled them to press the SPACE key as soon as they could identify the word. They were allowed to freely adjust their distance from the monitor.

Results

Table 1 reports participants' identification accuracy which was similar for large and small words, difference = −1.1%, 95% confidence interval (CI; Cumming, 2012) = [−3.3%, 1.0%], $t(27)$ = 1.08, p = .29, Cohen's d = 0.20. All data from incorrectly identified trials were removed from the following analyses.

Participants' recall accuracy for large and small words was calculated using the formula:

$$Recall\ accuracy = \frac{\text{Number of words correctly recalled}}{\text{Number of words correctly identified}} \times 100\%$$

[1] This sample-size specification is conservative. Morey (2016) showed that effect sizes change with varying numbers of experimental trials, because a larger number of trials yields a smaller mean squared error (MSE) and hence a greater effect size. Because we increased the number of trials compared to previous studies, we expect to observe a greater effect size. Thus the power to detect a significant font size effect on JOLs is expected to be greater than specified.

Table 1
M (*SD*) of participants' identification and judgment accuracy in Experiments 1–3.

	Large	Small
Experiment 1	93.4% (6.6%)	94.6% (3.5%)
Experiment 2		
CID	94.9% (4.4%)	94.4% (5.3%)
Lexical Word	97.2% (3.8%)	96.8% (5.0%)
Lexical Non-word	89.4% (14.3%)	92.6% (7.2%)
Experiment 3	94.3% (5.8%)	93.2% (6.3%)

Consistent with previous studies, we found no difference in recall accuracy between large and small words, difference = 0.9%, 95% CI = [−3.3%, 5.1%], $t(27) = 0.44$, $p = .66$, Cohen's $d = 0.08$ (see the right pair of bars in Fig. 2A). In contrast participants gave significantly higher JOLs to large ($M = 51.56$, $SD = 14.90$) than to small words ($M = 47.50$, $SD = 14.63$), difference = 4.05, 95% CI = [1.98, 6.13], $t(27) = 4.01$, $p < .001$, Cohen's $d = 0.76$ (see the left pair of bars in Fig. 2A), reflecting a font size effect on JOLs.

The key data concern the measure of perceptual fluency. As can be seen in Fig. 2B, participants' median identification RTs were significantly shorter for large ($M = 1.19$ s, $SD = 0.34$) than for small words ($M = 1.44$ s, $SD = 0.30$), difference = −0.25 s, 95% CI = [−0.33, −0.17], $t(27) = -6.60$, $p < .001$, Cohen's $d = -1.25$. Twenty-seven participants responded faster to large than to small words while only one showed the reverse pattern, $\chi^2(1) = 24.14$, $p < .001$. This is a very substantial effect of font size on perceptual fluency, as measured via the CID task.

To explore the statistical relationship between RTs and JOLs, we conducted a multilevel regression analysis using the R *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), with RTs as the independent variable and JOLs as the dependent variable. The results showed that the fixed effect of RTs on JOLs was −4.35, 95% CI = [−6.64, −2.11], indicating that every decrease of 1 s in RTs increases JOLs by 4.35 points on the 100-point scale. These results revealed a fluency effect on JOLs: The faster a word is identified, the higher the JOL it is given.

To directly test the fluency theory, we explored whether RTs mediate the font size effect on JOLs using a multilevel mediation analysis method with the R *bmlm* package (Vuorre, 2017). The package provides a Bayesian estimation of multilevel mediation models (Vuorre, 2017) and the mediation effect was estimated with 4 Markov Chain Monte Carlo (MCMC) chains and 10,000 iterations for each chain. In this multilevel mediation analysis, we took font size (small = 0; large = 1) as the independent variable, RTs as the mediator, and JOLs as the dependent variable. Table 2 reports the mediation results. The total effect of font size on JOLs was 4.11, 95% CI = [2.10, 6.12]. The indirect effect of font size on JOLs through RTs was 0.84, 95% CI = [0.31, 1.50], indicating that large fonts increase JOLs indirectly by increasing perceptual fluency. Fluency (RTs) explained 21%, 95% CI = [8%, 42%], of the font size effect on JOLs. The direct effect of font size on JOLs was 3.27, 95% CI = [1.34, 5.22], indicating that fluency did not explain all of the font size effect on JOLs: The direct effect of font size on JOLs was still significant when RTs were controlled.

Discussion

Perceptual fluency differs between large and small words as reflected by the significant difference in identification RTs. The faster a word is identified, the higher the JOL given to that word, as revealed by the inverse relationship between RTs and JOLs. Most importantly, perceptual fluency contributes to the font size effect on JOLs, as shown by the significant mediation results. In sum, these results demonstrate that perceptual fluency can affect JOLs and provide direct evidence that perceptual fluency underlies (at least in part) the stimulus size effect on JOLs.

Experiment 2

As previously discussed, the null result observed in Mueller et al.'s (2014) Experiment 1 might be due to a range of factors. In Experiment 2, we directly compared the lexical decision and CID tasks in the same participants, with the same number of trials and the same materials, to explore whether the CID task is more sensitive to variations in perceptual fluency than the lexical decision task.

Method

Participants

Twelve participants, with a mean age of 21.67 ($SD = 3.17$) years, 8 females, were recruited from the UCL participant pool (see Appendix A for the sample size calculation). They reported normal or corrected-to-normal vision, and received £2 or course credit as compensation.

Materials, design, and procedure

Eighty words were selected from Experiment 1 and 40 non-words (e.g., *dralp*) from the English Lexicon Project (Balota et al., 2007), following Mueller et al. (2014). The length of the non-words was 5 and all were monosyllabic. The words were randomly divided into two sets, one assigned to the CID task and the other to the lexical decision task. Set assignment to tasks was counterbalanced across participants. In the CID task, four words were used for practice and 36 for the main experiment. For each participant, the program randomly selected half the words to be presented in large and the remainder in small font sizes. In the lexical decision task, four words and four non-words were used for practice and 36 words and 36 non-words for the main experiment. For each participant, half the words and half the non-words were randomly chosen to be presented in large and the remainder in small font sizes. In both the CID and lexical decision tasks, the presentation sequence of items was randomly determined.

Experiment 2 involved a 2 (font size: large/small) × 2 (task: CID/lexical decision) within-subjects design. Half of the participants performed the CID task first followed by the lexical decision task, and the task order was reversed for the remainder of the participants. The procedure in the CID task was identical to that in Experiment 1 except that participants did not make item-by-item JOLs and did not take a free recall test. In the lexical decision task, words and non-words were randomly presented, one at a time, half in large and half in small font sizes. Participants were asked to judge whether the presented item was a word or a non-word as rapidly and accurately as they could by pressing the 'f' (word) or 'j' (non-word) key.

One reason for omitting item-by-item JOLs was that participants experienced non-words in the lexical decision task but not in the CID task. In the lexical decision task, the word type (word/non-word) might affect JOLs as well as the font size. As the aim of the experiment was specifically to explore whether the CID task is more sensitive to variations in perceptual fluency than the lexical decision task, omitting both the requirement for participants to make JOLs and the final memory test allowed us to compare the sensitivities of these two tasks to perceptual fluency while minimising influences from other task demands.

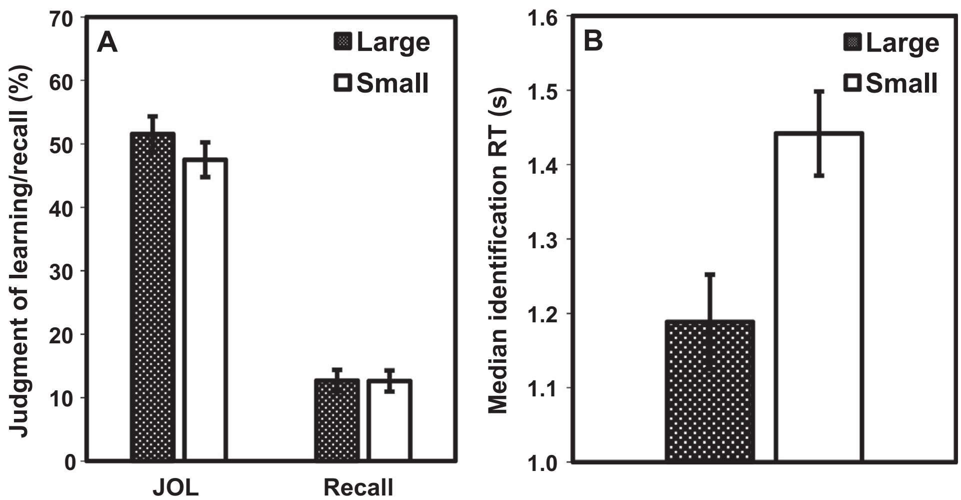

Fig. 2. Experiment 1. Panel A: Judgments of learning (JOLs) and recall for large and small words. Panel B: Median identification RTs for large and small words. Error bars represent ±1 standard error.

Results

In the CID task, there was no significant difference in identification accuracy between large and small words, difference = 0.5%, 95% CI = [−3.4%, 4.3%], $t(11) = 0.27$, $p = .80$, Cohen's $d = 0.08$ (see Table 1). In the lexical decision task, a repeated measures ANOVA, with word type (word/non-word) and font size as the within-subjects independent variables and judgment accuracy as the dependent variable, showed that words were judged more accurately than non-words, $F(1, 11) = 5.27$, mean square error (MSE) = 434.67, $p = .04$, $\eta_p^2 = 0.32$, but there was no main effect of font size, $F(1, 11) = 1.44$, $MSE = 23.15$, $p = .26$, $\eta_p^2 = 0.12$, and no interaction between font size and word type, $F(1, 11) = 0.85$, $MSE = 41.15$, $p = .38$, $\eta_p^2 = 0.07$ (see Table 1). All incorrectly identified trials in the CID task and incorrectly judged trials in the lexical decision task were removed from subsequent analyses.

Fig. 3 shows participants' median RTs in the CID and lexical decision tasks. In the CID task, participants identified large (M = 1.04 s, SD = 0.38) words faster than small (M = 1.23 s, SD = 0.41) words, difference = −0.19 s, 95% CI = [−0.28, −0.09], $t(11) = 4.40$, $p = .001$, Cohen's $d = -1.27$ (see the left pair of bars in Fig. 3). All participants responded faster on average to large than to small words, $\chi^2(1) = 12.00$, $p < .001$. RTs overall were slightly faster than in Experiment 1, probably caused by a combination of two factors: (i) The requirement to make a JOL on each trial in Experiment 1 may have induced participants to delay making their identification response while they formed their judgment; (ii) In Experiment 2 participants were free to choose which hand to use to make their response (in Experiment 1 they used their left hand).

For the lexical decision task, a repeated measures ANOVA, with word type (word/non-word) and font size as the within-subjects variables and RTs as the dependent variable, showed that participants responded faster to words than to non-words, $F(1, 11) = 22.19$, $MSE = 0.15$, $p = .001$, $\eta_p^2 = 0.67$, but there was no main effect of font size, $F(1, 11) = 1.06$, $MSE = 0.003$, $p = .33$, $\eta_p^2 = 0.09$, and no interaction between word type and font size, $F(1, 11) = 0.59$, $MSE = 0.001$, $p = .46$, $\eta_p^2 = 0.05$ (see the middle and right pairs of bars in Fig. 3). There was no difference in RTs between large (M = 0.63, SD = 0.11) and small (M = 0.60, SD = 0.07) words, difference = 0.02 s, 95% CI = [−0.02, 0.06], $t(11) = 1.24$, $p = .24$, Cohen's d = 0.36. Five participants responded faster to large words than to small words while seven showed the reverse pattern, $\chi^2(1) = 0.33$, $p = .57$. These results replicate Mueller et al.'s (2014) finding that there is no reliable difference in RTs between large and small words in a lexical decision task.

The critical question of interest is whether a significant interaction is obtained between task and font size in RTs. A repeated measures ANOVA, with task (CID/lexical decision) and font size as the within-subjects independent variables and RTs in the CID task and to words in the lexical decision task as the dependent variable, showed that participants responded faster to large words than to small words, $F(1, 11) = 10.60$, $MSE = 0.08$, $p = .008$, $\eta_p^2 = 0.49$, faster in the lexical decision task than in the CID task, $F(1, 11) = 27.19$, $MSE = 3.26$, $p < .001$, $\eta_p^2 = 0.71$, and there was a significant interaction between task and font size, $F(1, 11) = 24.97$, $MSE = 0.13$, $p < .001$, $\eta_p^2 = 0.69$.

Discussion

By employing the same participants, trials, and materials, we found a significant difference in RTs between large and small words in the CID task but not in the lexical decision task. These results clearly reveal that the CID task provides a more sensitive measure of perceptual fluency than the lexical decision task.

Experiment 3

In Experiment 1, we observed an inverse relationship between RTs and JOLs (i.e., the fluency effect on JOLs), and that fluency (RTs) partly mediates the font size effect size on JOLs. The first aim of Experiment 3 is to replicate these findings. The second aim is to explore how fluency affects JOLs. There are two possibilities. The first is that fluency affects JOLs directly: Fluency produces a feeling-of-knowing, which acts as a basis for JOLs. The second possibility is that fluency affects JOLs indirectly through people's beliefs about fluency: People believe that fluently processed items are easier to remember, and therefore they give higher JOLs to fluently processed items (for detailed discussion, see Dunlosky, Mueller, & Tauber, 2014; Mueller & Dunlosky, 2017). For example, Mueller and Dunlosky (2017) recently proposed that "a *belief about processing fluency* appears to produce the font-size effect (on JOLs) (Mueller et al., 2014) and not differential processing fluency per se." (p. 11). However, recent research has also provided evidence that beliefs about fluency cannot explain the fluency effect on JOLs (e.g., Undorf et al., 2017). Therefore, Experiment 3 aims to explore whether fluency directly affects JOLs or affects them indirectly through beliefs about fluency.

Table 2
Multilevel mediation analysis results in Experiments 1 and 3.

	b	*SE*	95% CI
Experiment 1: Font size-RTs-JOLs			
Effect of font size on RTs	−0.21	0.03	[−0.28, −0.15]
Effect of RTs on JOLs	−3.70	1.06	[−5.87, −1.69]
Total effect of font size on JOLs	4.11	1.02	[2.10, 6.12]
Direct effect of font size on JOLs	3.27	0.99	[1.34, 5.22]
Indirect effect of font size on JOLs through RTs	0.84	0.30	[0.31, 1.50]
Proportion of the total effect of font size on JOLs mediated by RTs	21%	15%	[8%, 42%]
Experiment 3: Font size-RTs-sJOLs			
Effect of font size on RTs	−0.20	0.04	[−0.28, −0.13]
Effect of RTs on sJOLs	−2.81	0.65	[−4.09, −1.51]
Total effect of font size on sJOLs	4.30	0.91	[2.51, 6.09]
Direct effect of font size on sJOLs	3.69	0.91	[1.88, 5.49]
Indirect effect of font size on sJOLs through RTs	0.60	0.18	[0.30, 0.99]
Proportion of the total effect of font size on sJOLs mediated by RTs	15%	6%	[7%, 28%]
Experiment 3: RTs-oJOLs-sJOLs			
Effect of RTs on oJOLs	1.40	2.92	[−4.29, 7.17]
Effect of oJOLs on sJOLs	0.07	0.02	[0.02, 0.11]
Total effect of RTs on sJOLs	−3.33	0.7	[−4.69, −1.98]
Direct effect of RTs on sJOLs	−3.21	0.66	[−4.49, −1.91]
Indirect effect of RTs on sJOLs through oJOLs	−0.12	0.38	[−0.94, 0.58]
Proportion of the total effect of RTs on sJOLs mediated by oJOLs	3%	12%	[−22%, 26%]
Experiment 3: Font size-oJOLs-sJOLs			
Effect of font size on oJOLs	17.23	3.02	[11.31, 23.24]
Effect of oJOLs on sJOLs	0.02	0.03	[−0.04, 0.07]
Total effect of font size on sJOLs	4.30	1.01	[2.32, 6.31]
Direct effect of font size on sJOLs	3.91	1.05	[1.88, 6.00]
Indirect effect of font size on sJOLs through oJOLs	0.39	0.57	[−0.65, 1.63]
Proportion of the total effect of font size on sJOLs mediated by oJOLs	9%	15%	[−17%, 38%]
Experiment 3: Font size-(RTs, oJOLs)-sJOLs			
Indirect effect of font size on sJOLs though RTs	0.62	0.17	[0.29, 0.95]
Indirect effect of font size on sJOLs though oJOLs	0.64	0.44	[−0.22, 1.50]
Difference between the indirect effect through RTs and through oJOLs	−0.02	0.45	[−0.90, 0.86]

Note: JOL = judgment of learning; sJOL = study phase judgment of learning; oJOL = observation phase judgment of learning.

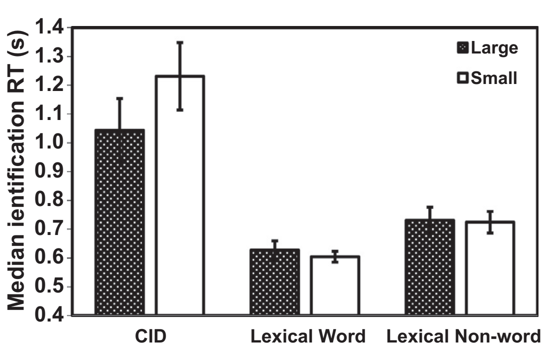

Fig. 3. Experiment 2. Median identification RTs in the CID task and median judgment RTs in the lexical decision task for the word and non-word trials. Error bars represent ±1 standard error.

The third aim of Experiment 3 is to test the *analytic processing* (Mueller & Dunlosky, 2017; Mueller et al., 2016) and *dual-basis* theories (Koriat, Bjork, Sheffer, & Bar, 2004; Mueller et al., 2016). Analytic processing theory proposes that people's beliefs play a dominant role in JOLs whereas fluency plays a much smaller or even no role. In contrast, dual-basis theory claims that both fluency and beliefs contribute importantly to JOLs (Koriat et al., 2004; Mueller et al., 2016). A few previous studies have tested these two theories, with inconclusive results (e.g., Mueller & Dunlosky, 2017; Mueller et al., 2014, 2016; Undorf et al., 2017; Witherby & Tauber, 2017).

To conceptually replicate Experiment 1's findings, in Experiment 3 we asked participants to perform the same study task as that in Experiment 1, in which they identified 100 words, half in small and half in large font sizes, and made item-by-item JOLs. To test whether fluency directly affects JOLs or affects them indirectly through beliefs, we need to measure the latter and explore to what extent beliefs can explain the fluency effect on JOLs (Undorf & Erdfelder, 2011; Undorf et al., 2017). We employed the learner-observer paradigm to measure beliefs about fluency, following Undorf and Erdfelder (2011). Immediately following the study task, participants were asked to perform an observation task, in which they were instructed to view identification responses purportedly from another participant and make item-by-item JOLs to predict the likelihood that that participant would remember the item. In the observation task, each word was replaced by a letter string (i.e., *abcde*), and the letter string and the mask were presented in the same font size and duration as a corresponding item in the study task (see below for details). Because in the observation task participants did not explicitly experience the identification process, JOLs can only be based on beliefs. This observation task can be regarded as a measure of both participants' beliefs about the font size effect on memory (that is, whether they believe that large words are more likely to be remembered than small words) and their beliefs about fluency (that is, whether they believe that more rapidly identified items are more likely to be remembered), because in the observation task they viewed each item's font size and identification speed.

In the data analysis, we conducted a multilevel mediation analysis to explore whether beliefs mediate the fluency effect on JOLs. To test the analytic processing and dual-basis theories, we asked whether beliefs (both about the font size effect on memory and about processing fluency) can explain a greater proportion of the font size effect on JOLs than fluency, or *vice versa.*

Previous studies showed that participants may adjust their beliefs across a study phase (e.g., Susser et al., 2017; Undorf & Erdfelder, 2011). Putting the observation task after the study task for all participants allows us to measure the beliefs that they developed and applied in the study task.

Method

Participants

Given that the first aim of Experiment 3 is to conceptually replicate Experiment 1's findings, we planned the same sample size as in Experiment 1. In total, 30 participants were recruited from the UCL participant pool. One participant's data were not recorded due to computer failure, leaving a final sample of 29 participants, with a mean age of 20.72 (SD = 2.45) years, 21 females. They reported normal or corrected-to-normal vision, and received £5 or course credit as compensation.

Materials, design, and procedure

The same stimuli were employed as in Experiment 1. Experiment 3 consisted of three tasks: study, observation, and test. The study task was same as in Experiment 1: Participants identified 100 words with the CID procedure (Fig. 1), half in large and half in small font sizes, and made item-by-item JOLs. Following the study phase, participants were given the following instructions for the observation task:

> *You will observe the responses of another participant who had undergone the same learning task. However, instead of seeing the exact words which the participant identified, you will see the letter string "abcde" in place of all the words. On each trial, the mask and the letter string will be displayed to you in the same FORMAT as in the learning phase, and for the same DURATION that the participant took to identify the word. Please CAREFULLY observe the participant's identification process, put yourself in his or her perspective, and judge the likelihood that he or she would remember that word at a later test.*

Although we told participants that they would observe another participant's identification trials, in fact they observed their own study phase trials replayed without the word information. Ten practice trials were presented in the same font size and duration as the practice trials in the study task, but in a new random order. In the main observation phase, they observed their own identification trials in a new random order. On each trial, the letter string (i.e., *abcde*) and mask were alternately presented. No response was required to terminate the identification part of the trial. Following the presentation of the letter string and mask, participants made a JOL to predict the likelihood that the "other participant" would remember that word later. Then they pressed the ENTER key to trigger the next trial.

To summarise, participants observed their own identification trials during the observation task, but we informed them that they were observing another participant's trials. In addition, we presented all items in a new random order. The aim was to prevent participants from realizing that they were observing their own identification trials and then explicitly recalling their JOLs from the study task.

Following the observation task, all participants completed a free recall test, which was the same as in Experiment 1. We also measured how much effort participants put into the study and observation tasks. After completing each of these phases, participants reported how much effort they had exerted on a scale ranging from 1 (no effort at all) to 7 (full effort).

Results and discussion

Participants' effort ratings were greater than the median of the rating scale (4) in both the study (M = 5.10, SD = 0.94) and observation (M = 4.93, SD = 1.16) tasks. There was no difference in effort ratings between the two tasks, difference = 0.17, 95% CI = [−0.21, 0.55], $t(28)$ = 0.93, p = .36, Cohen's d = 0.17. These results suggest that participants engaged in both tasks to an approximately equal extent.

In the study task there was no significant difference in identification accuracy between large and small words, difference = 1.1%, 95% CI = [−0.1%, 2.3%], $t(28)$ = 1.86, p = .07, Cohen's d = 0.35 (see Table 1). All data from incorrectly identified trials were removed from the following analyses.

There was no difference in recall accuracy between large (M = 14.6%, SD = 8.31%) and small words (M = 14.1%, SD = 9.6%), difference = 0.5%, 95% CI = [−2.2%, 3.2%], $t(28)$ = 0.36, p = .72, Cohen's d = 0.07 (see the right pair of bars in Fig. 4A). In contrast, in the study task participants gave significantly higher sJOLs (i.e., JOLs in the study task) to large (M = 50.63, SD = 11.38) than to small words (M = 46.40, SD = 10.10), difference = 4.23, 95% CI = [2.47, 5.99], $t(28)$ = 4.92, p < .001, Cohen's d = 0.91 (see the left pair of bars in Fig. 4A), reflecting a font size effect on sJOLs. In the observation task, participants gave significantly higher oJOLs (i.e., JOLs in the observation task) to large (M = 52.84, SD = 14.32) than to small words (M = 35.62, SD = 14.16), difference = 17.22, 95% CI = [11.35, 21.09], $t(28)$ = 6.01, p < .001, Cohen's d = 1.12 (see the middle pair of bars in Fig. 4A). As can be seen in Fig. 4B, participants' median identification RTs were significantly faster for large (M = 1.17 s, SD = 0.38) than for small words (M = 1.43 s, SD = 0.56), difference = −0.26 s, 95% CI = [−0.36, −0.16], $t(28)$ = −5.42, p < .001, Cohen's d = −1.00. Twenty-five participants responded faster on average to large than to small words while only four showed the reverse pattern, $\chi^2(1)$ = 15.21, p < .001.

Does fluency contribute to the font size effect on JOLs?

We conducted a multilevel mediation analysis with font size as the independent variable, fluency (RTs) as the mediator, and sJOLs as the dependent variable, using the R *bmlm* package, to explore whether fluency mediates the font size effect on JOLs (see Table 2 for detailed results). The total effect of font size on sJOLs was 4.30, 95% CI = [2.51, 6.09]. The indirect effect of font size on sJOLs through RTs was 0.60, 95% CI = [0.30, 0.99], slightly smaller than in Experiment 1 but nonetheless again indicating that large fonts increase JOLs indirectly by increasing perceptual fluency. Fluency (RTs) explained 15%, 95% CI = [7%, 28%], of the font size effect on sJOLs. The direct effect of font size on sJOLs was 3.69, 95% CI = [1.88, 5.49]. These results suggest successful replication of Experiment 1's findings: Font size affects JOLs (at least partially) through perceptual fluency.

Does fluency affect JOLs through beliefs about fluency?

In the following analyses, we explored whether fluency affects JOLs through beliefs about fluency. We first conducted a multilevel regression of RTs on sJOLs to quantify the fluency effect on sJOLs. The results showed an inverse relationship between RTs and sJOLs, fixed effect = −3.34, 95% CI = [−4.49, −2.17], indicating that every decrease of 1 s in RTs increases sJOLs by 3.34. Then we conducted a multilevel regression of RTs on oJOLs to explore people's beliefs about fluency. This analysis found no significant relationship between RTs and oJOLs, fixed effect = 1.41, 95% CI = [−4.41, 7.22], hence revealing a dissociation between the fluency effect on sJOLs and beliefs about fluency, contradicting the claim that fluency affects JOLs through beliefs. Thus while the identification RT for a word in the learning task predicts the sJOL given to it, it does not predict the oJOL given to the letter string *abcde* when the latter is presented in the observation phase for the same duration as the word had been in the learning phase.

Although we observed no relationship between RTs and oJOLs, we also conducted a multilevel mediation analysis to explore whether beliefs mediate the fluency effect on sJOLs. This multilevel

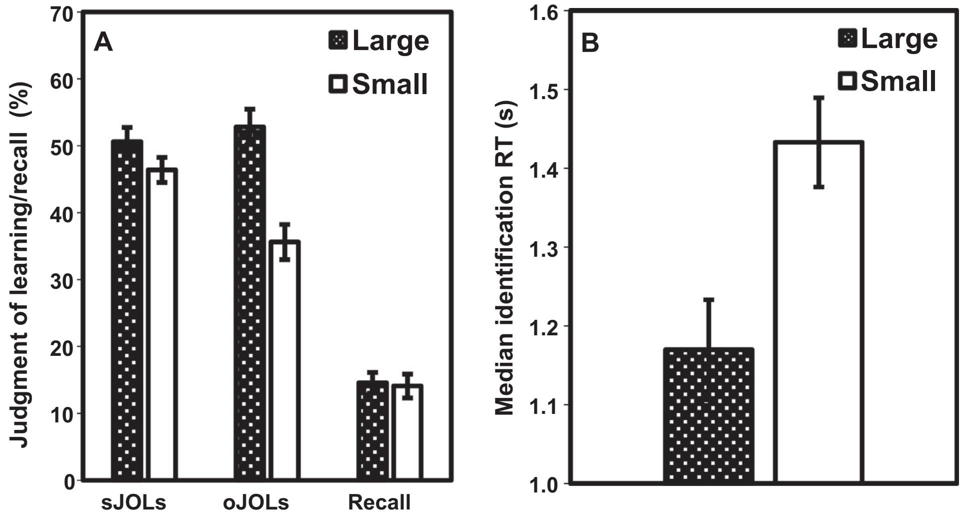

Fig. 4. Experiment 3. Panel A: Study judgments of learning (sJOLs), observation JOLs (oJOLs), and recall for large and small words. Panel B: Median identification RTs for large and small words. Error bars represent ±1 standard error.

mediation analysis was conducted using the R *bmlm* package, with fluency (RTs) as the independent variable, beliefs (oJOLs) as the mediator, and sJOLs as the dependent variable (see Table 2 for detailed results). The results show that the indirect effect of fluency (RTs) on sJOLs through beliefs (oJOLs) was −0.12, 95% CI = [−0.94, 0.58], which is non-significant (because the 95% CI includes 0) and again counter to the claim that fluency affects JOLs via beliefs.

Do beliefs play a more important role than fluency in the font size effect on JOLs, or vice versa?

We next explored whether beliefs play a more important role than fluency in the font size effect on JOLs, or vice versa. We first conducted a multilevel mediation analysis to determine whether beliefs (oJOLs) mediate the font size effect on sJOLs. This mediation analysis was conducted using the R *bmlm* package, with font size as the independent variable, beliefs (oJOLs) as the mediator, and sJOLs as the dependent variable (for detailed results, see Table 2). The results show that the total effect of font size on sJOLs was 4.30, 95% CI = [2.32, 6.31] and the indirect effect of font size on sJOLs through beliefs was 0.39, 95% CI = [−0.65, 1.63]. The proportion of the effect of font size on sJOLs mediated by beliefs (oJOLs) was 9%, 95% CI = [−17%, 38%]. The direct effect of font size on sJOLs was 3.91, 95% CI = [1.88, 6.00]. Overall, this multilevel mediation analysis shows little evidence supporting the claim that font size affects JOLs via beliefs.

We also conducted a multilevel mediation analysis to explore whether fluency plays a more important role than beliefs in the font size effect on JOLs, or vice versa. In this analysis, font size was taken as the independent variable, fluency (RTs) and beliefs (oJOLs) as two mediators, and sJOLs as the dependent variable. The analysis was conducted using the M*plus* program (Muthén & Muthén, 1998–2010; Preacher, Zyphur, & Zhang, 2010).[2] Table 2 reports the detailed results. The indirect effect of font size on sJOLs through fluency (RTs) was 0.62, 95% CI = [0.29, 0.95], again indicating that font size affects JOLs at least partially through fluency. The indirect effect of font size on sJOLs though beliefs (oJOLs) was 0.64, 95% CI = [−0.22, 1.50], which again shows little evidence that font size affects JOLs through beliefs. The difference between the indirect effect of font size on sJOLs via fluency (RTs) and via beliefs (oJOLs) was −0.02, 95% CI = [−0.90, 0.86], indicating no difference between the indirect effects. Overall, these results are inconsistent with the claim of analytic processing theory that beliefs dominate fluency, but are in line with the dual-basis theory.

[2] We switched to M*plus* because the R *bmlm* package is not yet applicable to multilevel mediation analyses with multiple mediators (we thank Matti Vuorre for confirming this). We also conducted multilevel mediation analyses using M*plus* to replicate the ones reported above conducted with the *bmlm* package. All the results showed the same patterns. We report results from the R *bmlm* package in the Results section because it provides Bayesian estimation (Vuorre, 2017). All data for these analyses are available at OSF.

General discussion

Until recently the font size effect on JOLs was widely taken as direct evidence that perceptual fluency can affect JOLs. However, Mueller et al. (2014) found no difference in fluency between large and small words when assessed by means of lexical decision and self-regulated study time allocation and hence suggested that the fluency theory is unlikely to provide an adequate account for the font size effect on JOLs. Subsequently, many researchers began to question the role that perceptual fluency plays in the font size effect on JOLs. For instance, Mueller and Dunlosky (2017) interpreted font size experiments as revealing that JOLs are mainly based on the deliberate application of people's beliefs. We suspected that the null result in RTs in the lexical decision task in Mueller et al.'s (2014) Experiment 1 might be caused by task insensitivity to variations in perceptual fluency. In addition, the null result in the self-regulated study time allocation task in Mueller et al.'s (2014) Experiment 2 might have resulted from the fact that this dependent measure can be affected by many other possible factors besides fluency.

In the current research, we directly tested the fluency theory by employing a CID task. In Experiments 1 and 3, we found a substantial font size effect on JOLs as reflected by a significant difference in JOLs between large and small words, while font size had no effect on actual recall. These results replicate the classic font size effect on JOLs (Hu et al., 2015, 2016; Mueller et al., 2014; Rhodes & Castel, 2008). Our results show that large words are processed with greater perceptual fluency than small words, as revealed by a significant difference in identification RTs. There was a significant fluency effect on JOLs, supported by an inverse relationship between RTs and JOLs. More importantly, we also found that large font size increases JOLs indirectly by increasing perceptual fluency, as reflected by a significant mediation of RTs in the font size effect on JOLs. These results bring the fluency theory back to the foreground as an account for the font size effect on JOLs. Going beyond Undorf et al. (2017), these results also provide direct evidence that perceptual fluency partly causes the stimulus size effect on JOLs.

Experiments 1 and 3 contradict Susser et al.'s (2016) proposal that effects of perceptual fluency on JOLs do not exist. Our findings, corroborating those of Undorf et al. (2017), support the conclusion that perceptual fluency can affect JOLs. The differences in perceptual fluency in our Experiments 1 and 3 and those in Undorf et al.'s (2017) Experiments 1–3 were greater than in Susser et al.'s (2016) Experiment 2. The perceptual fluency effects on JOLs might have been too small to be detected in Susser et al.'s (2016) study which had only 36 trials, compared to 100 trials in our Experiments 1 and 3 and 64 trials in Undorf et al.'s (2017) Experiments 1–3. Therefore, lack of power resulting from the small number of trials might have contributed to the null result in Susser et al.'s study.

Consistent with Undorf et al.'s (2017) findings, our Experiments 1 and 3 also challenge the proposal that beliefs play a dominant role in the formation of JOLs (Mueller & Dunlosky, 2017; Mueller et al., 2013, 2014). Experiment 1 supports the dual-basis theory, which proposes that JOLs are based on both beliefs and fluency (Koriat, 1997; Undorf et al., 2017). Furthermore, Experiment 3 directly compared the contributions of fluency and beliefs (both beliefs about the effect of font size on memory and about fluency) to the font size effect on JOLs. The results revealed no difference in the roles (importance) between fluency and beliefs in the font size effect on JOLs, which is inconsistent with the analytic processing theory but in line with the dual-basis theory. However, it is important to note that we do not reject the analytic processing theory (see below for detailed discussion).

In Experiment 3, we also explored whether fluency affects JOLs directly or indirectly through beliefs about fluency. We observed an inverse relationship between RTs and sJOLs but no relationship between RTs and oJOLs, indicating a dissociation between the fluency effect on JOLs and beliefs about fluency. In addition, the multilevel mediation analysis found no evidence that beliefs about fluency mediate the fluency effect on JOLs. There are two potential explanations of these results. The first possibility is that participants in Experiment 3 simply had no beliefs about fluency. The second possibility is that they had such beliefs but did not apply them when forming their oJOLs in the observation task (Koriat et al., 2004; Kornell & Hausman, in press; Kornell et al., 2011). Participants in the observation task might regard font size as a more salient cue than identification speed, and therefore base their oJOLs on font size rather on identification speed. If they did not apply beliefs about fluency to form their oJOLs, then there is little reason to expect that they applied beliefs about fluency when forming their sJOLs in the study task, because they experienced the difference in font sizes in both the study and observation tasks. Therefore, regardless of whether participants had no beliefs about fluency or had such beliefs but did not apply them, we propose that, at least in the current research, beliefs about fluency play no role in the fluency effect on JOLs.

In recent years, the roles of fluency and beliefs in JOLs has received a great deal of attention among researchers (e.g., Dunlosky et al., 2014; Frank & Kuhlmann, 2016; Mueller & Dunlosky, 2017; Mueller et al., 2014, 2016; Undorf & Ackerman, 2017; Undorf & Erdfelder, 2011, 2015; Yang et al., 2017b). How to measure and compare the roles (importance) of fluency and beliefs in JOLs has been a key concern. Experiment 3 provides an illustration of how to achieve this using the same participants and the same items.

In Experiment 2, we directly compared the CID and lexical decision tasks by employing the same participants, trials, and materials. We found a significant difference in identification RTs between large and small words in the CID task, but no difference was found in judgment RTs in the lexical decision task. These results are consistent with previous studies' findings (Ferrand et al., 2011; Grainger & Segui, 1990) and clearly indicate that the CID task is more sensitive to variations in perceptual fluency than the lexical decision task. Although the principal implications of the results concern the effects of fluency on metacognitive judgments, they also bear on the theoretical analysis of perceptual identification and lexical decisions. It is well-established that variables can have effects of very different magnitude on naming (identification) and lexical decision. For instance, word frequency has a much larger impact on lexical decision than on naming latencies (Schilling, Rayner, & Chumbley, 1998). The difference between these tasks is usually conceptualized in terms of the additional decision stage required to judge whether a lexical item is a word or a nonword. Models of lexical decision (e.g., Balota & Chumbley, 1984; Ratcliff, Gomez, & McKoon, 2004), which tend to focus on nonperceptual variables such as word frequency and concreteness, could be extended to incorporate variables such as size, color, or font which have not traditionally been considered.

The lexical decision and self-regulated study time allocation tasks have both commonly been used in previous studies examining the role of fluency in metamemory. For instance, Jia et al. (2015) explored whether fluency underlies the word frequency effect on JOLs (i.e., higher JOLs to high frequency words than to low frequency words) by employing a self-regulated study time allocation task. They found no difference in study times allocated to high versus low frequency words. Mueller et al. (2016) explored whether fluency underlies the identity effect on JOLs (i.e., higher JOLs to identical word pairs, e.g., *dog-dog*, than to related pairs, e.g., *dog-cat*) by employing a self-regulated study time allocation task. Mueller et al. (2016) found that study times were shorter for identical pairs than for related pairs, but study times did not mediate the identity effect on JOLs. Witherby and Tauber (2017) investigated whether fluency underlies the concreteness effect on JOLs (i.e., higher JOLs to concrete words than to abstract words). By employing a lexical decision task, Witherby and Tauber (2017) found that judgment RTs were shorter for concrete words than for abstract words, but RTs did not mediate the concreteness effect on JOLs. Using a self-regulated study time allocation task, Witherby and Tauber (2017) found no difference in study times between concrete and abstract words.

All of the aforementioned studies employed lexical decision or self-regulated study time allocation tasks to explore the role of fluency in some metamemory phenomena, failed to either find a significant difference in fluency or a significant mediation of fluency, and then concluded that fluency plays no role in these metamemory phenomena. We suggest future research to re-examine these metamemory phenomena by employing the CID task.

Limitations

There are two limitations of the current research. The first limitation is that in both Experiments 1 and 3, participants made a JOL immediately following each correct identification. Such a procedure might draw participants' attention to fluency and inflate its influence on JOLs. Drawing participants' attention to fluency might also contribute to the null difference between the indirect effect through fluency and that through beliefs in Experiment 3. Another limitation is that in Experiment 3, fluency (RTs) and sJOLs were collected in the study task, but beliefs (oJOLs) were measured in the observation task, which might contribute to the null difference in the indirect effects. Therefore, we reiterate that we do not reject the analytic processing theory. Future research is encouraged to develop more elegant procedures to avoid drawing participants' attention to fluency (or to measure fluency less overtly) while measuring the role of fluency in JOLs. In addition, future research is encouraged to develop new methods to measure the roles of fluency and beliefs in JOLs simultaneously (in contrast to the different tasks in Experiment 3), allowing researchers to compare the roles of fluency and beliefs more accurately.

Conclusion

Perceptual fluency can affect JOLs. Large font size increases JOLs at least in part through increasing perceptual fluency, which implies that perceptual fluency contributes to the stimulus size effect on JOLs. The current study found little evidence that beliefs about fluency play a role in the fluency effect on JOLs. The results support the dual-basis theory (Koriat, 1997), but we reiterate that we do not reject the analytic processing theory (Mueller & Dunlosky, 2017). Lastly, the CID task provides a more sensitive measure of perceptual fluency than the lexical decision task.

Acknowledgements

This research was supported by a China Scholarship Council (CSC) award to Chunliang Yang. We thank Jiawen Huang for his help in data collection, and John Dunlosky and two anonymous reviewers for their constructive suggestions.

Author note

All data have been made publicly available via the Open Science Framework at https://osf.io/2zfye/.

Appendix A

As Morey (2016) showed, effect sizes will change as a function of the number of trials. In Experiment 2, we decreased the number of trials to 36 compared to 100 in Experiment 1. To determine the required sample size for Experiment 2, we re-analyzed the RT data from Experiment 1. In Experiment 1, participants successfully identified about 94% of words, therefore we expected that participants in Experiment 2 would each successfully identify about 17 ($94\% \times 18$) large and small words. Based on this estimated number, we calculated the median RTs for the first 17 large and small words which were correctly identified by each participant in Experiment 1. Then we conducted a paired-sample t test, which showed that participants responded faster to large than to small words on these restricted sets, difference = -0.31 s, 95% CI = [-0.43, -0.19], Cohen's $d = -1.04$. Consistent with Morey's analysis, this is appreciably smaller than the effect size ($d = 1.25$) computed across all trials. Using this effect size, we calculated that Experiment 2 requires about 12–13 participants to detect a significant ($\alpha = 0.05$) difference in RTs between large and small words in the CID task at power = 0.9. Finally, we determined the sample size at 12.

It is interesting to note that the obtained effect size for the CID task in Experiment 2 ($d = 1.27$) is in fact slightly larger than that observed in Experiment 1. This does not invalidate Morey's argument, which is based on a statistical necessity. Instead it arises, presumably, because of other uncontrolled task or sample differences between the two experiments.

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*(4), 336–348. https://doi.org/10.1037/xap0000023.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. (2012). Models of recognition, repetition priming, and fluency: Exploring a new framework. *Psychological Review, 119*(1), 40–79. https://doi.org/10.1037/a0025464.

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 Cognition, 42*(9), 1417–1433. https://doi.org/10.1037/xlm0000246.

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

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823.

Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. *Psychology of Learning and Motivation, 48*, 225–270. https://doi.org/10.1016/s0079-7421(07)48006-9.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition, 12*(6), 590–606. https://doi.org/10.3758/BF03213348.

Coltheart, M. (2007). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497–505. https://doi.org/10.1080/14640748108400805.

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

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.

Dunlosky, J., Mueller, M., & Tauber, S. K. (2014). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger, III (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honour of Larry L. Jacoby* (pp. 46–64). New York, NY: Psychology Press.

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.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Meot, A., ... Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from chronolex. *Frontiers in Psychology, 2*, 306. https://doi.org/10.3389/fpsyg.2011.00306.

Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review, 27*(4), 567–586. https://doi.org/10.1007/s10648-015-9313-7.

Frank, D. J., & Kuhlmann, B. G. (2016). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory & Cognition*. https://doi.org/10.1037/xlm0000332.

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Attention, Perception, & Psychophysics, 47*(2), 191–198. https://doi.org/10.3758/BF03205983.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 22–34. https://doi.org/10.1037/0278-7393.29.1.22.

Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PLoS One, 10*(11), e0142351. https://doi.org/10.1371/journal.pone.0142351.

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.

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2015). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*, 1995. https://doi.org/10.3389/fpsyg.2015.01995.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. *Perception 36 ECVP Abstract Supplement*.

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.

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

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

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

Kornell, N., & Hausman, H. (in press). Performance bias: Why judgments of learning are not affected by learning. *Memory & Cognition*. https://doi.org/10.3758/s13421-017-0740-1 (in press).

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

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition, 44*(5), 696–705. https://doi.org/10.3758/s13421-016-0598-7.

Li, F., Xie, R., Li, X., & Li, W. (2015). The influence of perceptual information on control processes involved in self-regulated learning: Evidence from item selection. *Psychonomic Bulletin & Review, 22*(4), 1007–1013. https://doi.org/10.3758/s13423-014-0762-7.

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*(1), 35–56. https://doi.org/10.1007/s11409-015-9147-1.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174–179. https://doi.org/10.3758/pbr.15.1.174.

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

Morey, R. (2016). Averaging can produce misleading standardized effect sizes Retrieved from <http://www.statsblogs.com/2016/01/07/averaging-can-produce-misleading-standardized-effect-sizes/>.

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. https://doi.org/10.1016/j.jml.2016.10.008.

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 memory? *The Quarterly Journal of Experimental Psychology, 69*(4), 781–799. https://doi.org/10.1080/17470218.2015.1058404.

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.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide (Version 6)*. Los Angeles: Muthén & Muthén.

Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multi-level induction of categories: Venomous snakes hijack the learning of lower category levels. *Psychological Science, 25*(8), 1592–1599. https://doi.org/10.1177/0956797614535938.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*(3), 209–213. https://doi.org/10.1037/a0020141.supp.

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.

Price, J., McElroy, K., & Martin, N. J. (2016). The role of font size and font style in younger and older adults' predicted and actual recall performance. *Aging, Neuropsychology, and Cognition, 23*(3), 366–388. https://doi.org/10.1080/13825585.2015.1102194.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159.

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.

Sanborn, A. N., Malmberg, K. J., & Shiffrin, R. M. (2004). High-level effects of masking on perceptual identification. *Vision Research, 44*(12), 1427–1436. https://doi.org/10.1016/j.visres.2004.01.004.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*(6), 1270–1281. https://doi.org/10.3758/BF03201199.

Stark, C. E., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 945–972. https://doi.org/10.10371/0278-7393.26.4.945.

Susser, J. A., Jin, A., & Mulligan, N. W. (2016). Identity priming consistently affects perceptual fluency but only affects metamemory when primes are obvious. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*(4), 657–662. https://doi.org/10.1037/xlm0000189.

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.

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language, 95*, 116–123. https://doi.org/10.1016/j.jml.2017.03.002.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review, 20*(2), 356–363. https://doi.org/10.3758/s13423-012-0319-6.

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1261-4. Advance online publication.

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*(5), 1264–1269. https://doi.org/10.1037/a0023719.

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition, 43*(5), 647–658. https://doi.org/10.3758/s13421-014-0479-x.

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.

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23*(1), 87–102. https://doi.org/10.3758/s13423-015-0892-6.

Vuorre, M. (2017). *bmlm: Bayesian multilevel mediation. R package version 1.3.0.* Retrieved from <https://cran.r-project.org/package=bmlm>.

Ward, E. V., Berry, C. J., & Shanks, D. R. (2013). An effect of age on implicit memory that is not due to explicit contamination: Implications for single and multiple-systems theories. *Psychology & Aging, 28*(2), 429–442. https://doi.org/10.1037/a0031888.

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition, 45*(4), 639–650. https://doi.org/10.3758/s13421-016-0681-0.

Yang, C., Potts, R., & Shanks, D. R. (2017a). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied, 23*(3), 263–277. https://doi.org/10.1037/xap0000122.

Yang, C., Potts, R., & Shanks, D. R. (2017b). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1073–1092. https://doi.org/10.1037/xlm0000363.

Yang, C., Sun, B., & Shanks, D. R. (2017). The anchoring effect in metamemory monitoring. *Memory & Cognition*. https://doi.org/10.3758/s13421-017-0772-6. Advance online publication.

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*(3), 597–613. https://doi.org/10.1037/xlm0000064.