Journal of Memory and Language 70 (2014) 1–12

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory?

Michael L. Mueller [a], John Dunlosky [a,*], Sarah K. Tauber [a], Matthew G. Rhodes [b]

[a] Department of Psychology, Kent State University, Kent, OH, USA
[b] Colorado State University, Fort Collins, CO, USA

ARTICLE INFO

Article history:
Received 6 May 2013
revision received 16 September 2013

Keywords:
Judgments of learning
Fluency
Beliefs
Theory of metamemory judgments

ABSTRACT

Evidence suggests that processing fluency affects many kinds of judgments. For instance, when words are presented either in large (48 point) or smaller (18 point) font sizes during study, people's judgments of learning (JOLs) are higher for the words presented in the larger font size. This *font-size effect* presumably arises because items presented in a larger font size are easier to process at study, which in turn leads to higher JOLs. In the present studies, we evaluated this fluency hypothesis against an alternative one that the font-size effect occurs because people believe that words printed in a large font size are better remembered. In Experiments 1 and 2, we measured differences in processing fluency during study to evaluate whether fluency could account for any of the relationship between font size and JOLs. In Experiments 3a and 3b, college students read about the font-size experiment and then predicted whether hypothetical particpiants would better remember the large or smaller words. In Experiment 4, we evaluated whether the effect occurred for prestudy JOLs, which are made prior to studying the to-be-learned words and hence cannot be affected by processing fluency. Surprisingly, the evidence across experiments supported the belief hypothesis and did not support the fluency hypothesis. Thus, the font-size effect does not exemplify the effect of fluency on JOLs, and more generally, these outcomes suggest that measuring processing fluency is essential for establishing its role in people's judgments and decision making.

Introduction

For over four decades, researchers have investigated how people make judgments of learning (i.e., predictions about the likelihood of remembering recently studied information), partly because people use their judgments of learning (JOLs) to make decisions about how to regulate their study time (Metcalfe & Finn, 2008; for reviews, see Dunlosky & Ariel, 2011; Son & Metcalfe, 2000). Unfortunately, people's JOLs can suffer from metacognitive illusions, such as when they are influenced by a cue that has no impact on memory, which in turn can lead to inappropriate regulation of study (e.g., Castel, McCabe, & Roediger, 2007; Rhodes & Castel, 2009). Researchers have identified many cues that lead to metacognitive illusions wherein a cue influences JOLs but does not influence memory performance. For instance, JOLs are greater for words presented in larger (vs. smaller) font sizes during study (McDonough & Gallo, 2012; Rhodes & Castel, 2008), words spoken more loudly (vs. softly) at study (Rhodes & Castel, 2009), and words that are associated with a greater (vs. lesser) physical weight during study (Alban & Kelley, 2013), to name a few. Despite the discovery of a host of cues that produce such metacognitive illusions (for a review, see Bjork, Dunlosky, & Kornell, 2013), little is known about why any particular cue influences JOLs and hence specifically why metacognitive illusions occur.

* Corresponding author. Address: Department of Psychology, Kent State University, Kent, OH 44242, USA.
E-mail address: jdunlosk@kent.edu (J. Dunlosky).

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.09.007

A main goal of the present research is to provide a more complete understanding of how people construct JOLs by empirically scrutinizing how one particular cue – font size – influences JOLs. This *font-size effect* was first reported by Rhodes and Castel (2008). In their research, participants studied individual words that were presented either in a large font size (48 point font) or in a smaller font size (18 point font). JOLs were higher for words presented in a large than smaller font size, even though font size had no impact on final recall performance. This metacognitive illusion is robust and has occurred when participants completed two study-test trials, when participants also studied related and unrelated word pairs as a more diagnostic cue during study, and when judgments were framed in terms of forgetting (vs. learning). The effect of font size on JOLs could potentially be mediated by two factors, processing fluency or beliefs about memory (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Matvey, 2001; Koriat, Bjork, Sheffer, & Bar, 2004). We consider each factor in turn and how they presumably mediate the effects of any given cue on JOLs, and then we discuss in detail how they may mediate the font-size effect.

The construct of fluency and its relation to memory judgments has been instantiated in different ways. Larry Jacoby and his colleagues (Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982; for a review, see Jacoby, Kelley, & Dywan, 1989) argued that fluency has a direct influence on people's memory judgments. For instance, when judging whether a word presented on a test had been presented during study, the fluent processing of the word leads to a subjective experience of familiarity that in turn is unconsciously attributed to the word having been presented at study. In this case, the absolute level of fluency is causal, with greater ease of processing leading to higher levels of familiarity that influences memory judgments. By contrast, Whittlesea and Leboe (2002; see also, Whittlesea, 2002; Whittlesea & Williams, 2000) proposed that fluency leads to a subjective experience of familiarity only when the fluency of processing of an item is inconsistent with one's expectations. According to this discrepancy-attribution account (Whittlesea, 2002), when processing fluency is discrepant with expectations, the discrepancy causes surprise that leads to the subjective experience of familiarity, which in turn influences people's judgments.

Although both instantiations of fluency have been empirically supported, in the present research we focus on the degree to which absolute changes in processing fluency can mediate cue effects on JOLs. Our focus was on absolute fluency for two reasons. First, theories of JOLs have exclusively viewed fluency as having a direct and unmediated influence on how people make JOLs (Koriat & Bjork, 2006). The claim is that fluently processing an item at study leads to the subjective experience of familiarity that is unconsciously attributed to memorability. And, as argued by Koriat et al. (2004), "JOLs are based predominantly—perhaps exclusively—on the subjective experience associated with processing fluency" (p. 653). Second, and as important, evidence indicates that processing fluency of individual items is related to JOLs, with more fluent processing leading to higher JOLs (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003; Koriat, 2008; Matvey, Dunlosky, & Guttentag, 2001; Undorf & Erdfelder, 2011). Thus, one of our key questions in the present research is, "Does the cue of font size produce differences in processing fluency across items, which then mediates its effect on JOLs?"

Font size may also influence JOLs because people have a belief about how this cue affects memory and it is this belief that mediates the font-size effect. Such a belief about memory may be developed prior to the experiment and then applied in the experimental context, or it may be developed on-line as people consider how the cue could potentially influence memory. For an instance of the latter case, explicit instructions to predict future memory performance may trigger an analytic problem-solving mode in which people attempt to identify cues that will allow them to reduce uncertainty in their predictions. Research using multiple-trial methods has firmly established that people's beliefs can mediate the effects of some cues on JOLs (e.g., Bieman-Copland & Charness, 1994; Dunlosky & Hertzog, 2000; Hertzog et al., 2009; Tauber & Rhodes, 2010). For instance, Tauber and Rhodes (2010) had participants study face-noun pairs and make a JOL for each one. Some faces were associated with a name (Mr. Baker) whereas others were associated with an occupation (baker). After studying and making JOLs, the faces were presented and participants had to recall the corresponding name or occupation for each one. During an initial study-test trial, this cue (name vs. occupation) did not influence JOLs, although recall was better for occupations than for names. On a second study-test trial with new face-noun pairs, JOLs were greater for face-occupation pairs than for face-name pairs. Thus, participants learned from task experience during the initial test trial that occupations were better remembered, and this new knowledge influenced JOLs on the second trial. That is, a belief was developed on-line during the experiment and subsequently mediated the effect of this cue on JOLs.

Despite the intuitive plausibility that these factors – fluency and beliefs – contribute to the effects of cues on JOLs, the degree to which they actually mediate any cue-JOL relations has rarely been directly investigated (for exceptions, see Besken & Mulligan, 2013; Castel, 2008; Mueller, Tauber, & Dunlosky, 2013), and no research has evaluated their contribution to the font-size effect. We were particularly interested in the font-size effect because the prevailing interpretation is that it is mediated by fluency (e.g., Alter & Oppenheimer, 2009; Bjork et al., 2013; Carpenter, Wilford, Kornell, & Mullaney, 2013; Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Greifeneder & Unkelbach, 2012; Kornell, Rhodes, Castel, & Tauber, 2011; Rhodes & Castel, 2008; Schwartz & Efklides, 2012). For example, Diemand-Yauman et al. (2011) claimed that "Rhodes and Castel (2008) demonstrated biases in metacognition by manipulating fluency via font size" (p. 114). And more recently, Bjork et al. (2013) note that "metacognitive judgments tend to be higher for items with greater perceptual fluency – that is, items that are subjectively easier to process at a perceptual level... [e.g.,] larger fonts have been incorrectly judged to be memorable" (p. 432). In some circumstances, researchers assumed that the font-size effect was driven by processing fluency so as to explore other aspects of how learners make JOLs. For instance, Miele, Finn, and Molden (2011, Experiment 2) hypothesized that individuals with an entity theory about intelligence (i.e., intelligence is fixed) would base JOLs on perceptual fluency, whereas individuals with an incremental theory (i.e., intelligence can be improved) would not. Most relevant here, their evaluation of this prediction assumes that the font-size effect is mediated by processing fluency.

Although the font-size effect has come to exemplify the contribution of processing fluency on JOLs, no direct evidence has been reported to support this conjecture. The available evidence is indirect and from the original paper by Rhodes and Castel (2008, Experiment 5). They hypothesized that if processing fluency was responsible for the font-size effect, then disrupting fluency should in turn minimize the size of the effect. To disrupt fluency, some words were presented in aLtErNaTiNg case, and as expected, font size did not significantly influence JOLs when the words were presented in the altered case. These data suggest the possibility that fluency may contribute to the font-size effect, but even so, beliefs about font size could also be playing a role. In fact, Rhodes and Castel's (2008) outcome could have been driven completely by belief. For example, participants may believe that font size has an influence on memory, which led them to give higher JOLs to large font-size items; when words are presented in alternating case, however, the use of such a belief could be disrupted (e.g., if participants applied a competing belief that alternating case dampens memory for all items). Of course, given that the influence of belief and fluency are not mutually exclusive, another possibility is that both factors will jointly mediate the font-size effect on JOLs.

Accordingly, we systematically investigated the degree to which fluency and beliefs mediate the font-size effect. To do so, we used a variety of methods to estimate whether each factor (or both) contributed to the effect. In Experiments 1 and 2, we directly evaluated the contribution of fluency by estimating the degree to which font size affects objective measures that presumably tap different aspects of processing fluency at encoding. This approach is critical for establishing whether fluency can mediate the effect of a cue on JOLs. In Experiments 3a, 3b, and 4, we examined whether people's beliefs about font size can mediate its effect on JOLs.

Experiment 1

In Experiment 1, we investigated whether large and smaller font-sized words would produce different levels of processing fluency as measured by response latency in a lexical decision task. This task measures the time it takes for a participant to decide whether a string of letters is a word or a non-word. In the present context, a large font size (vs. a smaller one) may enhance the speed with which information accumulates about whether the current item is a word or non-word, which in turn would lead to faster decision times (e.g., in terms of the diffusion model, Ratcliff, Gomez, & McKoon, 2004). Another possibility is that the larger font size will be more quickly perceived, so that conceptual processing begins earlier for larger font-size words and leads to faster lexical decision response times. In either case, if this aspect of processing fluency is driving the font-size effect, then lexical decision times will be faster for large than for smaller words. Alternatively, the differential size of 18-point vs. 48-point font may not impact processing fluency at all, because even the 18-point font is rather large compared to the font sizes used in standard printed text and hence may be easily processed. One might even expect the 48-point font, because it occupies more space, to disrupt fluency and hence result in slower lexical decision times.[1] Finally, in addition to a group who made lexical decisions for large and smaller font items, we had another group make JOLs to examine whether the font-size effect would remain in the presence of non-words.

Method

Design, participants, and materials

A 2 (font size: large or smaller) × 2 (group: immediate JOL or lexical decision) mixed factorial design was used, with font size as the within-participant factor and group as the between-participants factor. Sixty-four undergraduates were recruited from Kent State University to fulfill a partial course requirement and were randomly assigned to either the lexical decision group or the JOL group. Two participants were excluded from the lexical decision group for having performance below 80% for the words on the lexical decision task. Thus, 27 participants were in the immediate JOL group and 35 participants were in the lexical decision group (we ran more participants in the latter group expecting that more participants would be excluded due to not meeting the exclusion criterion).

Thirty-six words were taken from the lists used by Rhodes and Castel (2008), along with another set of 36 non-words (e.g., arage; sampled from the English Lexicon Project, Balota et al., 2007). Each word had a non-word counterpart that was matched for length and orthography. Half of the words and non-words were presented in an 18 point Arial font and half were presented in a 48 point Arial font. Font size was counterbalanced between participants, such that all words and non-words were presented equally often in both font sizes. Items were presented in a random order with the restriction that no more than three words or non-words of the same font size could appear consecutively. The order was randomized anew for each participant.

Procedure

Seventy-two items (36 words and 36 non-words) were presented for 5 s each. Participants were instructed to study all items for an upcoming test (although during the test they were instructed to recall only words). In the lexical decision group, participants were instructed that they would be completing a lexical decision task where they must decide as fast as possible whether the item presented

[1] We thank Dave Balota for suggesting this possibility.

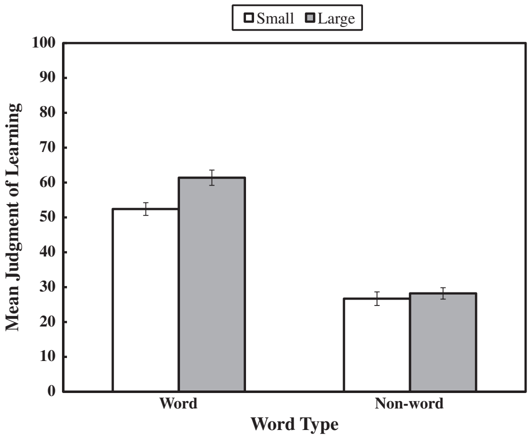

Fig. 1. Mean judgment of learning for each font size and each word type for the JOL group in Experiment 1. Error bars represent within-participant standard error.

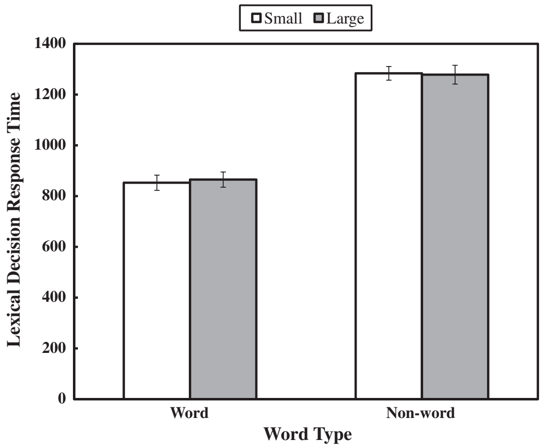

Fig. 2. Mean lexical decision response time in milliseconds for each font size and each word type for the lexical decision group in Experiment 1. Error bars represent within-participant standard error.

was a word or a non-word. Before each item, a screen appeared telling participants to get ready for the next item while a progress bar counted down 2 s. Then a fixation point appeared in the center of the screen for .5 s, followed by the item. When the item was displayed on the screen, participants made their decision by pressing the ''Z'' key to indicate the item on the screen was a word or by pressing the ''M'' key to indicate the item on the screen was a non-word. The word remained on the screen until the study time had elapsed. In the immediate JOL group, participants were not asked to make a lexical decision when the item was presented for study. After each item was presented, participants made a JOL on a scale from 0 to 100, with 100 indicating that they would definitely remember the word and 0 indicating that they would definitely not remember the word. After all 72 items were presented, participants had 4 min to recall as many of the 36 words that they had studied. In all experiments with free recall, responses in which the first three letters matched the first three letters of a study word were scored as correct.

Results and discussion

Mean JOLs for words and non-words are presented in Fig. 1.[2] A 2 (font size) × 2 (word type: word or non-word) within-participant ANOVA revealed that participants gave higher JOLs to words presented in a large font size than words presented in a smaller font size, $F(1,26) = 15.6$, $MSE = 744.27$, $p = .001$, $\eta_p^2 = .38$. This main effect, however, was qualified by a significant interaction: JOLs were higher for large than for smaller words, but JOLs did not differ as a function of font size for non-words, $F(1,26) = 12.23$, $MSE = 376.305$, $p = .002$, $\eta_p^2 = .32$. Also, JOLs were higher for words than for non-words, $F(1,26) = 73.96$, $MSE = 23418.92$, $p < .001$, $\eta_p^2 = .74$. Although less relevant to our current aims, the interaction (with no font-size effect for non-words) is reminiscent of how font size interacts with alternating case (Rhodes & Castel, 2008), which we reconsider in the General discussion.

Most important, mean response times for lexical decisions for each word type are presented in Fig. 2. (For each participant, an item was removed from this analysis if the participant responded to it incorrectly on the lexical decision trial; see Yap, Balota, & Tan, 2012). A 2 (font size) × 2 (word type) within-participant ANOVA indicated response times were faster for words than for non-words, $F(1,34) = 57.2$, $MSE = 6234$, $p < .001$, $\eta_p^2 = .63$. No significant difference was found between response times for large and smaller font-size words, $F(1,34) = .04$, $p = .843$, suggesting that font size had no impact on processing fluency.[3] The interaction was not significant, $F < 1$.

Mean recall performance for words is presented in Table 1 (note that participants were not instructed to recall non-words). No effects were significant, $F\text{s} < 1.7$.

Experiment 2

In Experiment 1, no differences were found in response times for large and smaller font-size words indicating that processing fluency, as measured by lexical decision, is not mediating the font-size effect. However, other kinds of processing fluency have been described in the literature (Alter & Oppenheimer, 2009) and hence font size may affect JOLs by producing differences in another aspect of processing fluency. Thus, in Experiment 2, we investigated the possibility that processing fluency, as measured by study time, mediates the relationship between font size and JOLs.

[2] The main comparison occurs for the within-participant manipulation of font size, so standard errors reported in this paper (both in the text and in figures) are based on within-participant computations (Franz & Loftus, 2012). Thus, the standard error for the contrasting values for large and smaller font sizes will be identical.

[3] In Experiments 1 and 2, we also conducted analyses using medians of each participant's response times. The outcomes were nearly identical in both experiments, except a small and nonsignificant trend occurred in the unexpected direction in Experiment 1, with response times being slightly (but not significantly) slower for large than smaller font-size words.

Table 1
Mean percentage of correct recall for Experiments 1, 2, and 4.

	Font size	
	Smaller	Large
Experiment 1		
Immediate JOL	16.0 (1.0)	17.7 (1.0)
Lexical decision	15.0 (1.0)	12.8 (1.0)
Experiment 2		
Limited study time	25.4 (1.2)	25.9 (1.2)
Self-paced	26.7 (1.5)	27.8 (1.5)
Experiment 4		
Immediate JOL	23.7 (1.2)	23.7 (1.2)
Prestudy JOL	23.1 (1.1)	23.7 (1.1)

Note: Values are mean percentage of correct recall for each font size and group. Within-participant standard errors of the mean are in parentheses.

Study time is inversely related to JOLs (Castel et al., 2007; Koriat, 2008; Undorf & Erdfelder, 2011) and hence is arguably a likely candidate for mediating the font-size effect vis-à-vis a fluency mechanism. For instance, Castel et al. (2007) examined whether differences in study time could account for why JOLs were higher for identical pairs (e.g., spoon – spoon) than for unrelated pairs (e.g., dog – spoon). Consistent with their prediction, study time was faster for identical pairs than for unrelated pairs, suggesting "fluency plays a critical role, since subjects overestimate recallability of identical pairs. Identical pairs may be perceived as fluent because reading the first word facilitates reading speed of the second word" (p. 111, Castel et al., 2007).

In Experiment 2, we measured study time to further evaluate the contribution of processing fluency for the font-size effect. In particular, participants were given an unlimited amount of time to study words presented in large and smaller font sizes. Study time was operationalized as the time from when a word was presented for study to when the participant pressed the space bar to indicate he or she was done studying. Immediately after studying an item, the participant made a JOL. Although this procedure is standard to examine study time-JOL relationships (e.g., Koriat, 2008), it does not map directly onto the procedure used to demonstrate the font-size effect in which the presentation rate is fixed by the experimenter (typically at 5 s per item). In this case, the fluency hypothesis predicts that the fixed rate (or nominal time) is not always identical to the actual study time per item (Cooper & Pantle, 1967), with participants spending less time studying large than smaller words. Thus, we wanted to also explore whether differences in study time arise during a fixed presentation rate, which would be critical for explaining the font-size effect as it has been previously investigated. Accordingly, we included a group who indicated when they completed studying each word, but the study time was limited to 5 s. According to the fluency hypothesis, both groups will use less time studying words presented in a large than smaller font size, and such differences in study times across items will mediate the font-size effect on JOLs.

Method

Design, participants, and materials

A 2 (font size) × 2 (study time: limited study time or self-paced) mixed factorial design was used, with font size as the within-participant factor and study time as the between-participants factor. Sixty participants were recruited from Kent State University for a partial course requirement and were randomly assigned to either study time group. Only words from Experiment 1 were used with the same counterbalancing of font size between participants and constraints on the random presentation order.

Procedure

Participants in the limited study-time group had up to 5 s to study the word and were told to press the spacebar to indicate when they had finished studying the word. The word remained on the screen until the 5 s had elapsed. Participants in the self-paced group were told to press the spacebar when they had finished studying the word. For all participants, after a word was presented for study, a JOL was made for it. After participants had studied and made JOLs for all the words, they completed a free recall test for 4 min.

Results and discussion

As shown in Fig. 3, both groups provided higher JOLs for large compared with smaller words, $F(1,58) = 37.11$, $MSE = 2640.15$, $p < .001$, $\eta_p^2 = .39$. The main effect for group and the interaction were not significant, *F*s < 1. Mean study times for both groups are presented in Fig. 4. Participants spent more time studying the words when they had unlimited study time than when they had only 5 s, $F(1,58) = 7.83$, $MSE = 93.52$, $p = .007$, $\eta_p^2 = .12$. The main effect of font size and the interaction for study time were not significant, *F*s < 1, suggesting no differences in processing fluency for large and smaller font-size words. Mean recall performance is presented in Table 1, all *F*s < 1.

We also examined the relationship between study time and JOLs. Means across intra-individual correlations between study time and JOLs were not significantly different than 0 for either the limited study-time group (*Mean* $r = -.04$) or for the unlimited time group (*Mean* $r = .06$). Although speculative, we may not have found the expected negative relationship (e.g., Koriat, 2008) specifically because the cue of font size overshadowed the use of study time as a cue. Consistent with this possibility, Evans and Benjamin (2011) review evidence that "fluency cues appear to be more important under conditions in which other bases for the recognition decisions... were reduced" (pp. 20–21); that is, when other cues are available that participants view as relevant to the retrieval status of an item, fluency no longer influences their judgments. Further research will be needed to evaluate this hypothesis in the present context. Most important, however, the lack of an effect of font size on study time is inconsistent with the hypothesis that encoding fluency is responsible for the font-size effect on JOLs.

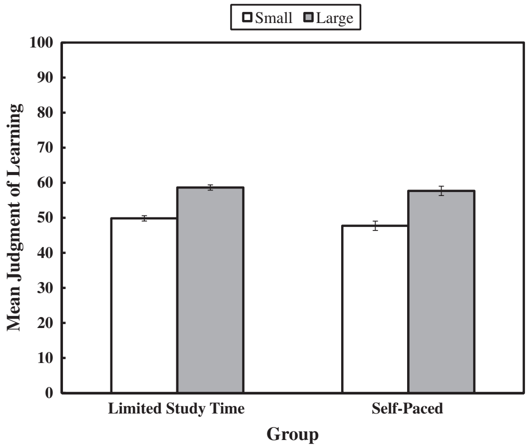

Fig. 3. Mean judgment of learning for each font size and for each study-time group for Experiment 2. Error bars represent within-participant standard error.

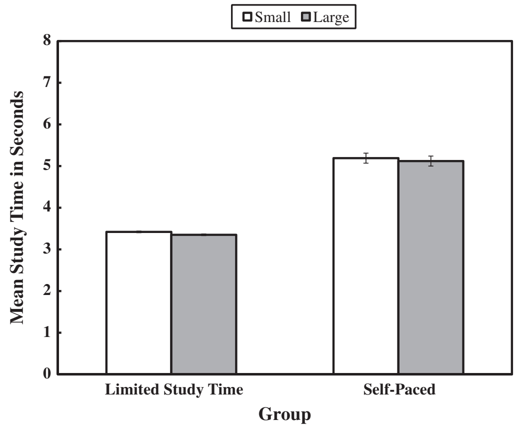

Fig. 4. Mean study time in seconds for each font size and for each study-time group for Experiment 2. Error bars represent within-participant standard error. Errors bars for the limited study time group are approximately the width of the figure lines.

Experiments 3a and 3b

Evidence from Experiments 1 and 2 suggests that processing fluency plays a minimal role in the font-size effect. Thus, in Experiments 3a and 3b, we changed our focus to examine whether people have a belief about font size impacting the memorability of individual words. To do so, we adapted a methodology used by Koriat et al. (2004) and Kornell et al. (2011), who administered a questionnaire that described an experiment and had participants estimate recall performance for different conditions. Most relevant, Kornell et al. (2011) described an experiment where students studied words presented in a large or smaller font size and later tried to recall them during a free recall test. After reading the scenario, participants estimated how many large and smaller font-size words would be recalled. Participants reported that recall would be greater for words presented in the larger font size, which is consistent with the belief hypothesis.

Experiments 3a and 3b provide a replication and extension of Kornell et al. (2011), which is critical given the importance of their outcome for evaluating the belief hypothesis and a general emphasis on replicating key experimental outcomes (e.g., Pashler & Harris, 2012). Moreover, participants from Kornell et al. (2011) were recruited through Amazon's Mechanical Turk and hence their sample (age = 35 years, age range from 18–81 years) may not entirely represent the college students who participated in the present studies. Thus, in Experiment 3a, we sampled college students from the same population used in the other experiments reported here; to foreshadow, like other research comparing outcomes from Mechanical Turk to those using a standard population of college students (Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013), outcomes from Experiment 3a replicated Kornell et al. (2011). Thus, Experiment 3b was used to extend these results in two ways. First, one questionnaire included an example of one word in a large font size and another in the small font size (as in Experiment 3a), whereas another version did not include word examples but instead presented black rectangles to illustrate the size of each font size (e.g., small = ■, but in 18 pt font). The rectangles allowed a relative comparison of the two font sizes but removed the possibility of experiencing the fluency of processing words of different font sizes. Second, and most important, after participants estimated performance they were asked to justify their estimates using a free-response format. These reports may give insight into the kinds of belief people develop about how font size may impact memory.

Method

In Experiment 3a, forty-eight participants were recruited from Kent State University and received extra credit for their participation. Participants read the following description of an experiment:

> In a previous experiment that we conducted, students were presented with a list of 36 words one after the other. Critically, half of the words (i.e., 18) were presented in a large font size and half of the words (i.e., 18) were presented in a small font size.
> Each of the words was presented for 5 s. The students' task was to study these words so that they would remember as many words as possible on a memory test. This memory test took place immediately after studying all the words and students were asked to recall as many words that they had previously studied.

On the questionnaire, an example of the font size was provided, such that the word "large" was presented in 48 point Arial font and the word "small" was presented in 18 point Arial font. Also, the order of the description of each font size and the order of estimates for each font size were counterbalanced across participants. After reading the scenario, participants estimated the number of words (out of 18) of each font size that the students in the experiment would recall.

In Experiment 3b, 100 participants were recruited using Mechanical Turk and were paid $1.00 for participating in the study. Participants read the same description as in Experiment 3a, but half were randomly assigned to the questionnaire including word examples and half were assigned to the one including rectangle examples. Order of descriptions and ratings were counterbalanced as in Experiment 3a. After making their estimates for recall, participants answered the following question: "We are interested in why you thought large [small] words would be better remembered than small [large] words. Please explain in as much detail as possible below." (Note that the prompt for participant explanations was customized based on whether they estimated higher recall for large words or for small words.) Their typed answers were saved for subsequent scoring.

Results and discussion

We replicated Kornell et al. (2011), with estimates being higher for large than smaller font sizes (Table 2) both for college students (Experiment 3a), $t(47) = 12.94$, $p < .001$, Cohen's $d = 1.92$, and for participants recruited from Mechanical Turk (Experiment 3b), $F(1,98) = 149.67$, $MSE = 1.87$, $p < .001$, $\eta_p^2 = .60$. Of less importance, the main effect of group in Experiment 3b was marginally significant, $F(1,98) = 3.18$, $MSE = .1$, $p = .078$, and was qualified by a significant interaction. Namely, higher estimates were given for large words when word examples were presented ($M = 63\%$, $SE = 2.0$) than when rectangle examples were presented ($M = 55\%$, $SE = 2.4$), whereas both groups gave similar estimates to smaller words ($M = 40\%$, $SE = 2.0$; $M = 39\%$, $SE = 2.4$, respectively), $F(1,98) = 4.45$, $MSE = .06$, $p = .037$, $\eta_p^2 = .04$. This interaction is small, needs to be replicated, and can be explained in multiple ways, such as that participants were unintentionally influenced by the specific words in the example (e.g., the word "large" is believed to be more memorable) or that presenting the filled-in rectangles does not adequately illustrate the difference in font size for individual letters. We leave evaluation of these and other explanations for future research, but critically, the main effect (Table 2) indicates that people have a belief that font size influences memory.

The primary focus of Experiment 3b was to evaluate participants' explanations for their estimates on the questionnaire. We focused on participants ($n = 88$) who had estimated higher recall for larger font-size words, and their explanations were independently classified by two raters into these categories: large words (a) are easier to remember (a circular explanation for the estimates), (b) are more important, (c) are easier to learn, (d) are easier to read, (e) are more distinctive, (f) would better capture attention, (g) are easier to see, or (h) are the words that the researcher wants me to say will be better remembered (i.e., demand characteristic). Rater agreement was high (97% consistency), and the few disagreements were handled via discussion.[4] Categories that included more than 5% of participant responses are included in Table 3, along with the percent use and an example response. The most prevalent responses were that the larger words were more distinct or would draw more attention; in both cases, participants are drawing on a reasonable psychological mechanism to explain their higher estimates, even though larger (vs. smaller) font sizes are not more distinctive and do not draw more attention in a manner that improves memory performance.

Although few participants ($n = 10$) responded that small words would be better remembered than large words, their responses provide similar justifications in favor of smaller words. For instance, participants responded that "you would have to strain your eyes more to see [the smaller words] and focus more so maybe you would remember better," and "[the smaller words] are easier to read." Given that few participants indicated that smaller words would be better remembered, however, we do not consider these outcomes further.

Table 2
Mean estimates of percent recall in Experiments 3a and 3b.

	Font size	
	Large	Smaller
Experiment 3a	68 (.96)	43 (.96)
Experiment 3b	59 (1.6)	40 (1.6)

Note: Values are mean percentage of recall estimates. Within-participant standard errors of the mean are in parentheses. Estimates were collapsed across groups in Experiment 3b.

Experiment 4

Experiments 3a and 3b provided evidence that most people believe that large (vs. smaller) font-sized words are better remembered. In Experiment 4, we used prestudy JOLs to investigate whether this belief would be applied when making item-by-item judgments. *Prestudy JOLs* are made before studying each item and were introduced by Castel (2008); they have been used to understand the influence of serial position, associative relatedness, and memory-for-past test performance on people's metacognitive judgments (Ariel & Dunlosky, 2011; Castel, 2008; Mueller et al., 2013). Prestudy JOLs typically include a prompt about the kind of item that will be presented for study. For example, a prompt for the current experiment would be, "You are about to study a small word, please rate how likely you are to remember it." Because participants do not study the item prior to making a prestudy JOL, processing fluency per se cannot influence judgments. Thus, if font size influences prestudy JOLs, it would indicate that fluency alone cannot account for the font-size effect and would further implicate the role of beliefs.

Method

Design, participants, and materials

A 2 (font size) × 2 (judgment group: immediate or prestudy JOL) mixed factorial design was used, with font size as the within-participant factor and judgment group as the between-participants factor. Sixty undergraduates were recruited from Kent State University to fulfill a course

[4] For 7% of the responses, neither rater understood the participant's explanation, and these responses were omitted from this analysis.

Table 3
Percentage of participant explanations for why they predicted that words would be more memorable when presented in larger font sizes (Experiment 3b).

	Percent of responses	Example response
Easier to remember	25	"The large words would be more memorable"
Easier to read	10	"Larger words are easier to read, and I feel that they make a stronger imprint in a person's mind. Larger visuals are usually easier to remember."
Distinctive	39	"To me words in larger, bold font is similar to someone screaming a word at me. I would think that if someone were to scream a word at me, it would be embedded into my memory a lot more than if someone were to whisper it to me, which would be like reading a word in a small font."
Attention	23	"Because I believe the larger words would be more memorable. They would be focusing on the larger words more and that means they would remember the larger words more."
Easier to see	17	"Because they would be easy to see making them easier to remember."

Note: This table shows the percentages based on the subset of participants that responded that recall would be higher for large words than small words. Participants could have been classified into multiple categories depending on their responses, so the percentages do not add to 100.

requirement and were randomly assigned to the immediate or prestudy JOL group. Participants were also randomly assigned to view an example of the two font sizes in their instructions; however, these instructions had no bearing on the results, so we collapsed across groups for our analysis. The same set of materials was used as in Experiment 2 with the same counterbalancing of font size between participants and constraints on the random presentation order.

Procedure

Participants were encouraged to do their best to study the words for an upcoming test. Each word was presented individually for 5 s for study. Participants in the prestudy JOL group made a JOL before studying each word and were shown the prompt: "You are about to study a small [large] word, please rate how likely you are to remember it." Participants in the immediate JOL group made a JOL using a similar prompt after they had studied each word. After studying and judging all the words, participants were given 4 min to complete a free recall test.

Results and discussion

As evident from Fig. 5, JOLs in both the prestudy JOL group and immediate JOL group were significantly higher for large than smaller words, $F(1,58) = 22.06$, $MSE = 924.59$, $p < .001$, $\eta_p^2 = .28$. The main effect of judgment type and the interaction were not significant, $F(1,58) = 1.8$, $p = .185$ and $F(1,58) = .70$, $p = .41$, respectively. These results support the hypothesis that people use a belief about font size when making JOLs.

To further explore these effects, we examined metamemory serial position curves (e.g., Castel, 2008; Dunlosky & Matvey, 2001; Tauber & Dunlosky, 2012), which can reveal the degree to which the font-size effect arises with task experience. If it does, then the font-size effect may initially be small and increase in size across serial positions. So as to reduce noise, the 36 words were grouped into 9 serial position bins (serial positions 1–4, 5–8, et cetera) as a function of font size. Fig. 6 shows that, regardless of judgment type, participants made higher judgments for large than smaller font-size words, $F(1,58) = 25.96$, $MSE = 15896.33$, $p < .001$, $\eta_p^2 = .31$, and JOLs tended to decrease across serial position bins, $F(8,464) = 11.32$, $MSE = 4608.95$, $p < .001$, $\eta_p^2 = .16$. No other effects were significant, $Fs < 1.52$. The general consistency in the effect of font size on judgments across serial positions supports the hypothesis that this effect is partly due to participants' beliefs that large (vs. smaller) words are better remembered. Moreover, these outcomes are relevant to evaluating one explanation for how participants develop beliefs about how font size affects memory. In particular, if larger words were more fluently processed (although that does not appear to be the case), then experiencing this differential processing fluency could inform their beliefs, and hence more exposure to differential fluency across serial positions would be expected to increase the size of the font-size effect. This possibility is ruled out, however, by the consistency in the font-size effect across serial positions.

Mean recall performance is presented in Table 1. No effects were significant, $Fs < 1$, which replicates previous research (Rhodes & Castel, 2008; see also Kornell et al., 2011) and establishes that this metacognitive illusion occurs both for prestudy JOLs and for JOLs made immediately after study.

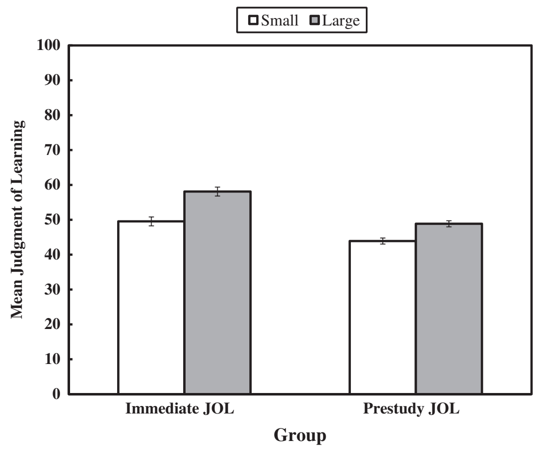

Fig. 5. Mean judgment of learning for each font size and for each judgment group for Experiment 4. Error bars represent within-participant standard error.

General discussion

Processing fluency presumably affects people's judgments and decisions, including judgments of truth, product preference, the quality of responses provided for trivia questions, and the intelligence of authors, to name only a few (Alter & Oppenheimer, 2009; Kelley & Rhodes, 2002; Schwarz, 2004). Given the prominence of fluency as a construct in theory of judgments across multiple domains, it perhaps is not surprising that manipulations of perceptual features – e.g., font size – are presumed to influence people's metamemory judgments vis-à-vis processing fluency. Nevertheless, an alternative hypothesis is that people's beliefs or theories about these manipulations are largely mediating their effects. The former fluency-based effects allegedly arise unconsciously and are mediated by the subjective experience from differential processing at study (e.g., Koriat & Bjork, 2006). By contrast, the latter belief-based effects arise from inferences derived from a plausible theory about how a manipulation influences memory. Despite that both factors – fluency and belief – can produce metacognitive illusions and are central to theorizing about metacognitive judgments (Dunlosky & Hertzog, 2000; Kelley & Jacoby, 1996), rarely have investigators directly evaluated their contribution to how people use cues when making JOLs.

Theory of judgments of learning: fluency, beliefs, and the font-size effect

Evidence from Experiments 3a, 3b and 4 suggest that beliefs in part – if not solely – mediate the font-size effect on JOLs. One possibility is that people do not have any beliefs about the influence of font size on memory prior to participating in these experiments. However, when they are instructed to predict later memory performance and font size is manipulated, they construct a theory about font size that seems plausible in the current experimental context. This proposal is inspired in part by Jacoby and Kelley (1987), who distinguished between memory used as a tool vs. as an object. When memory is used as a tool, such as to make a perceptual identification judgment or to judge a name as famous, then memory influences those judgments via an unconscious attribution about processing fluency. However, when memory is the object of a task, people "monitor performance for effects of a particular prior event" (p. 333, Jacoby & Kelley, 1987). The idea here is that focusing people on evaluating memory per se triggers an analytic mode of processing where people search for cues that would indicate whether a particular item was encountered during a prior study episode. For JOL experiments, when participants are instructed to judge their memory, such explicit instructions to predict future memory trigger an analytic problem-solving mode where participants search for cues that will reduce their uncertainty about the likelihood of subsequent recall.

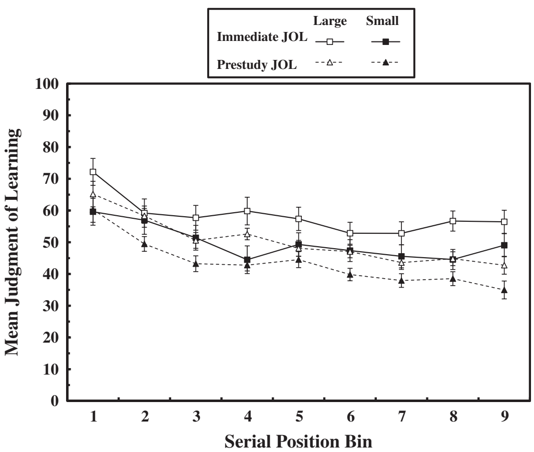

Fig. 6. Metamemory serial position curves for Experiment 4. Error bars represent within-participant standard error.

In the present experiments, an obvious cue is font size, and participants presumably develop a belief about why this cue may influence memory. Of course, the beliefs people generate about font size (or any cue) may vary considerably. One possibility was originally proposed by Rhodes and Castel (2008): "It may be that perceived importance (cf. Castel, 2007) also influences JOLs... [because] size may reflect the importance of to-be-learned information; large items may be considered more important than small items and thus be given higher JOLs" (p. 624). In contrast to this possibility, few participants (approximately 3%) in Experiment 3b claimed that larger font sizes are easier to remember because they are important. However, participants did provide other reasonable theories for why larger font sizes are more memorable, such as that words presented in a larger font size are more distinctive (Table 3). Research will be needed to evaluate the degree to which participants strategically search for cues when making JOLs so as to reduce their uncertainty about future recall.

More surprising, however, were the outcomes from Experiments 1 and 2, which did not provide any support for the fluency hypothesis. Presenting words in the larger font size simply did not lead to more fluent processing of items, as measured by lexical decision times and study time. Of course, perhaps some other – currently unknown – kind of processing fluency does differentiate between large (48-point) and smaller (18-point) font items, and we encourage researchers to reveal its contribution. Moreover, although our evidence suggests that absolute levels of fluency do not mediate the font-size effect, perhaps a discrepancy between expectations and fluency mediate a portion of this effect. We suspect that participants have few expectations for processing fluency within this task, but perhaps the discrepancy-attribution account (Whittlesea, 2002) could be modified to make testable predictions about how discrepancy contributes. At this point, however, evidence from the current experiments indicate that people's beliefs do contribute, so any currently undiscovered contribution of fluency (whether absolute or relative to expectations) would likely explain only a small portion of the font-size effect on JOLs.

The present results also raise questions with respect to interpreting prior evidence about the font-size effect. For instance, Rhodes and Castel (2008) reported that the font-size effect occurred when items were presented in a standard format but did not occur when items were presented in aLtErNaTiNg case. One explanation is that this interaction is due to the trade-off between the influence of two beliefs – that larger fonts yield better memory and that alternating case reduces memory (for supporting evidence, see Dunlosky, Mueller, & Tauber, 2013). In this case, the novelty of the alternating case may capture participants' attention making them less likely to consider the contribution of font size. This explanation appears consistent with the current interaction between font size and word status in Experiment 1, where the font-size effect was absent for non-words (Fig. 1), assuming that participants believe that non-words will be difficult to remember. Alternatively, note that alternating case would produce variability in physical appearance of letters within both large and smaller font-size items, which may serve to mask the presence of the font-size manipulation. Evaluating these possibilities in future research provides an important challenge because they concern how people combine multiple cues when constructing metacognitive judgments.

Given that processing fluency plays a minor role in producing the font-size effect, one may also want to reinterpret evidence from prior research that assumed fluency was responsible for it. For instance, Miele et al. (2011) found that individuals with an entity theory about intelligence demonstrated the font-size effect, whereas those with an incremental theory did not demonstrate it. Perhaps as compared to entity theorists, incremental theorists may be less likely to develop beliefs about factors that are less relevant to memory per se, and hence in general are less likely to demonstrate metacognitive illusions. Although this reinterpretation is provocative, research is needed to replicate Miele et al.'s (2011) original effect and to establish that individuals with different theories of intelligence also develop different beliefs about the effects of various manipulations on memory.

Beliefs, fluency, and font size for retrospective judgments of memory

Of particular relevance for the font-size effect, McDonough and Gallo (2012) reported that people's belief – or expectation – that larger font-size words are easier to remember can influence their retrospective memory judgments. After studying words presented in large and small font sizes, participants were given two recognition tests in which all words were presented in a neutral-sized font (so recognition decisions could not arise from differential processing fluency at test). One test involved identifying words previously presented in a large font (e.g., for each word, say "yes" if it had been presented in a large font size), and the other involved identifying words previously presented in a small font. Participants had lower false alarms for the test oriented towards identifying word previously presented in the large font size than for the test oriented towards smaller font-size words, demonstrating that participants' belief that large words are better remembered led to the use of a conservative response criteria on the former test. Using the same methods, when participants were oriented to a cue that they presumably did not believe would influence memory (i.e., font color), then the cue did not influence recognition decisions (Experiment 2, McDonough & Gallo, 2012). Thus, as in the present research focusing on prospective judgments, people's beliefs about memory – and, in particular, how font size should affect memory – impact their retrospective judgments of memory as well.

Even the effects of fluency on people's retrospective judgments appear to be (at least in part) mediated by people's beliefs. For instance, Westerman, Miller, and Lloyd (2003, Experiment 2) found that expectations can moderate the use of a fluency heuristic (i.e., familiarity) on tests of recognition memory. Two groups of participants were told they were to study either words or pictures that were being subliminally presented. However, during study, all participants were presented with just visual noise – no words or pictures were presented. After this phase, participants received a recognition test. Critically, each test word was primed with a word that either matched or mismatched it, so as to increase the familiarity of the target via enhancing processing fluency. Participants were asked to identify whether the test word had previously been presented by responding "yes" if the target "seemed familiar" (p. 623, Westerman et al., 2003). For the matched-prime test words, substantially more *yes* responses were reported by participants who were told they had previously studied words than by participants who were told they had previously studied pictures. This outcome indicates that when participants believed that familiarity was useful for identifying a previously presented word (i.e., when the modality of study and test was the same), they were more likely to use familiarity as a cue for recognition decisions. Put differently, only when participants believed that fluency was a diagnostic cue did they explicitly use it to make decisions about recognition memory (see also, Miller, Lloyd, & Westerman, 2008; Westerman, Lloyd, & Miller, 2002; and for a similar argument for fluency and JOLs, see Matvey et al., 2001).

Closing remarks

In their paper *Uniting the Tribes of Fluency to Form a Metacognitive Nation*, Alter and Oppenheimer (2009) conclude that "fluency is a ubiquitous metacognitive cue in reasoning and social judgment" (p. 219), which was based on a review of evidence showing that a variety of manipulations that presumably affect processing fluency also affect people's judgments. We suspect that fluency is rather pervasive and does mediate the effects of some manipulations on judgments. Indeed, some experiments discussed in their review included measures of fluency (which are essential to directly evaluating the hypothesis that processing fluency mediates the relationship between a focal manipulation and people's judgments and decision making), and Besken and Mulligan (2013) recently demonstrated that processing fluency mediates the effects of perceptual interference (presenting intact words that are easily read vs. briefly presenting them followed by a backwards mask) on people's judgments. Nevertheless, we are concerned that the effects of fluency may not be as widespread as assumed, especially given that processing fluency is often not measured and that the fluency hypothesis is rarely evaluated against competing hypotheses. For metamemory judgments such as JOLs, the belief hypothesis provides a plausible alternative and has fared well with respect to explaining the font-size effect as well as the effect of associative relatedness on JOLs (Mueller et al., 2013; for evidence that fluency may also contribute to the latter effect, see Castel et al., 2007). Thus, we encourage others to further explore the contribution of fluency by using strong inferences (Platt, 1964) that involve measures of processing fluency and evaluations of competitive hypotheses.

Acknowledgments

Many thanks to Deanne Westerman for comments on this paper, and we would like to thank Kianna Blakeney, Melissa Bishop, Lauren Bottoms, Caitlin Ciralsky, Sarah Cook, Tosha Jones, Tabatha Pizanie, and Caitlin Widuck and for their assistance with data collection.

This research was supported by the James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind, and Behavior Collaborative Award.

References

Alban, M. W., & Kelley, C. M. (2013). Embodiment meets metamemory: Weight as a cue for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1628–1634. http://dx.doi.org/10.1037/a0032420.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*, 219–235. http://dx.doi.org/10.1177/1088868310367640.

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition, 39*, 171–184. http://dx.doi.org/10.3758/s13421-010-0002-y.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*, 610–632.

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

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

Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging, 9*, 287–302. http://dx.doi.org/10.1037/0882-7974.9.2.287.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5. http://dx.doi.org/10.1177/1745691610393980.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*. http://dx.doi.org/10.3758/s13423-013-0442-z.

Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.). *The psychology of learning and motivation* (Vol. 48, pp. 225–270). London: Academic Press.

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

Castel, A. D., McCabe, D. P., & Roediger, H. L. III, (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review, 14*, 107–111. http://dx.doi.org/10.1177/ 1088868309341564.

Cooper, E. C., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin, 68*, 221–234.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical turk as a tool for experimental behavioral research. *PLoS One, 8*. http://dx.doi.org/10.1371/journal.pone.0057410.

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*, 111–115.

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. Ross (Ed.), *Psychology of Learning and Motivation, 54*, 103–140.

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2013). The contribution of processing fluency (and beliefs) to people's judgments of learning. In S. Lindsay (Ed.), *Remembering: Attributions, processes, and control in human memory*. Psychology Press (in press).

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about strategy effectiveness: A componential analysis of learning about strategy effectiveness from task experience. *Psychology & Aging, 15*, 462–474.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1180–1191. http://dx.doi.org/10.1037/0278-7393.27.5.1180.

Evans, K. M., & Benjamin, A. S. (2011). Fluency and familiarity: How memory for perceptual detail influences the remembering events. In P. A. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition* (pp. 17–27). Great Britain: Palgrave Macmillan.

Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review, 19*, 395–404. http://dx.doi.org/10.3758/s13423-012-0230-1.

Greifeneder, R., & Unkelbach, C. (2012). Experiencing thinking. In R. Greifeneder & C. Unkelbach (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behavior* (pp. 1–6). New York, NY: Psychology Press.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 22–34. http://dx.doi.org/10.1037/0278-7393.29.1.22.

Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S., & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology, 62*, 155–173. http://dx.doi.org/10.1080/17470210701855520.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3*, 306–340.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin, 13*, 314–336. http://dx.doi.org/10.1080/17470218.2012.739183.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36*, 300–324. http://dx.doi.org/10.1037/h0080638.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35*, 157–175.

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 293–320). New York: Academic Press.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36*, 416–428.

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*, 1133–1145.

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*, 643–656. http://dx.doi.org/10.1037/0096-3445.133.4.643.

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

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition, 29*, 222–233. http://dx.doi.org/10.3758/BF03194916.

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.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*, 174–179. http://dx.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*, 320–324. http://dx.doi.org/10.1177/0956797610397954.

Miller, J. K., Lloyd, M. E., & Westerman, D. L. (2008). When does modality matter? Perceptual versus conceptual fluency-based illusions in recognition memory. *Journal of Memory and Language, 58*, 1080–1094. http://dx.doi.org/10.1016/j.jml.2007.12.006.

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*, 378–384. http://dx.doi.org/10.3758/s13423-012-0343-6.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531–536.

Platt, J. R. (1964). Strong inference. *Science, 146*, 347–353.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.

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

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

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition, 1*, 145–151. http://dx.doi.org/10.1016/j.jarmac.2012.06.002.

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*, 332–348.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221. http://dx.doi.org/10.1037/0278-7393.26.1.204.

Tauber, S. K., & Dunlosky, J. (2012). Can older adults judge their learning of emotional information? *Psychology and Aging, 27*, 924–933.

Tauber, S. K., & Rhodes, M. G. (2010). Metacognitive errors contribute to the difficulty in remembering proper names. *Memory, 18*, 522–532. http://dx.doi.org/10.1080/09658211.2010.481818.

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. http://dx.doi.org/10.1037/a0023719.

Westerman, D. L., Lloyd, M. E., & Miller, J. K. (2002). On the attribution of perceptual fluency in recognition memory: The role of expectation. *Journal of Memory and Language, 47*, 607–617. http://dx.doi.org/10.1016/S0749-596X(02)00022-0.

Westerman, D. L., Miller, J. K., & Lloyd, M. E. (2003). Change in perceptual form attenuates the use of the fluency heuristic in recognition. *Memory & Cognition, 31*, 619–629. http://dx.doi.org/10.3758/BF03196102.

Whittlesea, B. W. A. (2002). Two routes to remembering (and another to remembering not). *Journal of Experimental Psychology: General, 131*, 325–348.

Whittlesea, B. W. A., & Leboe, J. P. (2002). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49*, 62–79. http://dx.doi.org/10.1016/S0749-596X(03)00009-3.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 547–565.

Yap, M. J., Balota, D. A., & Tan, S. E. (2012). Additive and interactive effects in semantic priming: Isolating lexical and decision processes in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 140–158. http://dx.doi.org/10.1037/a0028520.