Journal of Experimental Psychology:
Learning, Memory, and Cognition
2017, Vol. 43, No. 5, 680–693

© 2016 American Psychological Association
0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000332

More Than Just Beliefs: Experience and Beliefs Jointly Contribute to Volume Effects on Metacognitive Judgments

David J. Frank
The University of North Carolina at Greensboro and
Case Western Reserve University

Beatrice G. Kuhlmann
University of Mannheim

Experience-based cues, such as perceptual fluency, have long been thought to influence metacognitive judgments (Kelley & Jacoby, 1996; Koriat, 1997). Studies found that manipulations of perceptual fluency via changes in font and volume alter Judgments of Learning (JOLs) without influencing memory performance (Rhodes & Castel, 2008, 2009). Nonetheless, recent research (Mueller, Tauber, & Dunlosky, 2013; Mueller, Dunlosky, Tauber, & Rhodes, 2014, 2016) has challenged the notion that experience-based cues such as fluency are the primary basis for item-level JOLs, arguing instead that preexisting beliefs about these manipulations are responsible for these effects. For the first time, we compared global metacognitive judgments to item-level JOLs made during study to independently assess the contribution of beliefs and experience to volume-effects on JOLs. In 3 experiments, we found evidence for strong beliefs about volume-effects on memory, both before and after a study-test phase. However, these beliefs either did not account for the volume effect on JOLs (Experiment 3) or only partially accounted for the volume effect on JOLs (Experiments 1 and 2). Further, in Experiments 2 and 3 global performance estimates (before and after study) did not differ with respect to the volume dose whereas item-level JOLs generally varied with dose strength. Taken together, our findings suggest that both beliefs and experience-based cues contribute independently to the effects of volume on item-level JOLs, but that beliefs alone *cannot* fully account for the effects of volume on item-level JOLs.

Keywords: confidence judgments, fluency, judgments of learning, metacognition, metamemory

Understanding how well one has learned a piece of information is critical for deciding when to stop studying, when to continue studying, and recognizing when one's study strategy is not working (Ariel, Dunlosky, & Bailey, 2009; Hines, Touron, & Hertzog, 2009; Metcalfe & Kornell, 2005; Nelson & Narens, 1990; Pressley, Levin, & Ghatala, 1984). Likewise, memory confidence is important for deciding whether to change a test response or double check a piece of information before moving forward (e.g., checking the calendar or trusting that one knows when a meeting is scheduled). For these reasons, how people make judgments about their learning and memories is of great theoretical interest. Although such *metacognitive judgments* are typically better than chance at predicting memory performance (e.g., Nelson, 1996; Rhodes, 2015), they are far from perfect, suggesting that people do not have direct access to the strength of their memories (Nelson, 1996). Prominent metacognitive theories (Kelley & Jacoby, 1996; Koriat, 1997; Koriat, Bjork, Sheffer, & Bar, 2004) distinguish between two classes of information that can influence metacognitive judgments: experience-based cues, such as how quickly an item is processed or accessed, and beliefs/naïve theories about memory. Many factors known to affect metacognitive judgments, such as prior exposure, could operate through either experience or beliefs. For example, recently processed information is more quickly and easily processed the second time (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). Thus, subjective experiences (such as processing fluency) might contribute to item-level *Judgments of Learning* (JOLs; i.e., predictions of one's likelihood to remember a just studied item). Alternatively, one may simply suspect that studying an item twice will lead to better memory than only studying it once, and one may apply this belief when predicting how likely one is to remember a repeated word on a future test. Thus item-level JOLs may instead be belief-based. We will first review evidence for experience-based cues on metacognitive judgments, then explain the recent belief-based challenge to these results. Following, we will introduce a set of experiments designed to assess the contributions of both experience and beliefs on metacognitive judgments regarding volume.

Evidence for Influences of Experience-Based Cues (Fluency) on Metacognitive Judgments

Because factors like prior exposure (e.g., Begg et al., 1989) and relatedness (e.g., Undorf & Erdfelder, 2015) influence both fluency of processing and memory, processing fluency may be a valid experience-based cue on which many, if not most, metacognitive

This article was published Online First October 6, 2016.

David J. Frank, Department of Psychology, The University of North Carolina at Greensboro, and Department of Psychological Sciences, Case Western Reserve University; Beatrice G. Kuhlmann, Department of Psychology, School of Social Sciences, University of Mannheim.

We thank Christopher Dillon for the voice recording of the words and Dylan Brown, Kenzie Davis, Zijie "Tom" Fu, Ashkon Koucheki, Andrea Lau, Chuyue "Meredith" Gao, Madelina Ochenatu, Amanda Pupo, Natasha Rupani, Matt Shina, Srivatsan Uchani, Fizah Zafar, and Jiayuan "Michael" Zhang for assistance with data collection.

Correspondence concerning this article should be addressed to David J. Frank, Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH 44120. E-mail: djf94@case.edu

judgments are based (Koriat et al., 2004). Kelley and Jacoby (1996) even argued that, although beliefs and naïve theories play a role in metacognitive predictions, they are often trumped by experience-based cues, such as fluency, for item-level predictions. Quite intriguing evidence for experience-based effects on metacognitive judgments come from Rhodes and Castel (2008, 2009; see also Kornell, Rhodes, Castel, & Tauber, 2011) who manipulated the perceptual experience of study material by presenting words in either large or small font, standard or alternating capitalization (e.g., PiAnO), as well as loud or quiet volume. Although these perceptual manipulations did *not* influence memory (free recall) performance, they consistently affected JOLs (higher average JOLs for words studied in large font, standard capitalization, and loud volume). The authors interpreted their findings as indications that perceptual fluency is a cue often used to make metacognitive judgments (specifically JOLs). Similarly, the effects of many other manipulations on metacognitive judgments (mostly JOLs) have been interpreted to be mediated by the experience of fluency, including the effects of repetition (Begg et al., 1989), semantic relatedness (e.g., Undorf & Erdfelder, 2015), identicality of words in a pair (Castel, McCabe, & Roediger, 2007), and visual and auditory interference (Besken, 2016; Besken & Mulligan, 2013, 2014). Importantly, some studies have backed this fluency-account of metacognitive judgments with direct empirical evidence that objective measures of processing time are related to metacognitive judgments (e.g., Besken, 2016; Besken & Mulligan, 2014; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Undorf & Erdfelder, 2011, 2013, 2015).

Although in some cases processing fluency and performance go hand-in-hand, such as in the case of semantic relatedness of word pairs (Undorf & Erdfelder, 2015), fluency is not always indicative of better performance. Indeed, sometimes disfluency actually enhances memory (Yue, Castel, & Bjork, 2013) such that relying on fluency experiences for making memory predictions may lead to single or even double dissociations between metacognitive judgments and memory performance (Besken, 2016; Besken & Mulligan, 2013, 2014; Kornell et al., 2011; Rhodes & Castel, 2008, 2009). Such dissociations are important because they highlight that participants are using cues (presumably, fluency experience) to make their metacognitive judgments and do not somehow "know" what their memory performance is.

Most research on how processing experience influences metacognitive judgments has focused on fluency, specifically, but it should be noted that other experience-based cues may influence JOLs as well. For example, Alban and Kelley (2013) found that participants gave higher JOLs when the words they were studying were affixed to heavier material. Weight of study material should not affect word processing in itself, thus experience-based cues other than processing fluency presumably influenced JOLs in this embodied-cognition effect, and may also play a role in the effect of other manipulations (e.g., volume). Thus, various aspects of *in-the-moment experience* may influence metacognitive judgments.

Belief-Based Reinterpretation of Experience-Based Effects on Metacognitive Judgments

The fluency-experience account of JOLs has recently been challenged by researchers arguing that beliefs are the primary determinant of the JOL-effect for many of the manipulations commonly attributed to fluency (Dunlosky, Mueller, & Tauber, 2015; Mueller, Dunlosky, & Tauber, 2016; Mueller et al., 2014; Mueller, Tauber, & Dunlosky, 2013). Mueller and colleagues investigated whether beliefs might explain the effects of font size (2014), alternating capitalization (2013; 2016), semantic relatedness of word pairs (2013; 2015), and identicality of word pairs (2015) on JOLs. To do so, they measured participants' beliefs about the various manipulations using either vignettes or prestudy JOLs. The vignettes described the levels of the manipulation (including examples) and asked participants to estimate how many words/word-pairs participants of the described study remembered for each level. For prestudy JOLs (Castel, 2008), participants are told what type of word they are about to see (e.g., large or small font) and asked to estimate their memory *before* studying the word. Importantly, these metacognitive estimates are made in the absence of fluency information (they do not process the word or pair prior to making their judgments) and can thus only be influenced by beliefs. Across experiments, and for both the vignette and prestudy JOL measures, Mueller and colleagues found that participants estimated higher performance for the more "fluent" (large font, standard capitalization, related pairs, identical pairs) item type without/prior to processing it. That is, Mueller and colleagues demonstrated that participants hold preexisting beliefs about how these fluency manipulations affect memory, and suggested that these beliefs may be responsible for the effects on JOLs.

Aside from demonstrating that people hold beliefs about these manipulations, Mueller and colleagues also attempted to more directly measure the fluency effects of these manipulations. For font size, Mueller et al. (2014) did not find any evidence for faster processing for large font words when looking at either lexical-decision or self-paced study times, arguing against the hypothesis that font size affected JOLs via fluency experience. Consequently, lexical-decision times did not mediate the relationship between JOLs and font size. Although lexical-decision and study times were indeed shorter for related and identical word pairs, they also did not statistically mediate the relationship between JOLs and relatedness (Mueller et al., 2016, 2013). Lastly, Mueller et al. (2013) demonstrated that related and identical word pairs were given higher JOLs relative to unrelated pairs even when processing fluency was disrupted using alternating capitalization, again suggesting that fluency does not mediate the effects of these manipulations on JOLs. Overall, Mueller et al. (2013, 2014, 2016) conclude that JOLs derive from preexisting beliefs rather than fluency experience.

However, there is evidence that experience-based cues may influence JOLs in addition to or in spite of beliefs. For example, manipulations of alternating case, assumed to disrupt fluency, eliminated the font-size effect on JOLs (Rhodes & Castel, 2008) and decreased the effects of identical word pairs (Mueller et al., 2013), suggesting that fluency *does* play a role in these effects. Further, Undorf and Erdfelder (2015) argue that lexical-decision times are not a sensitive measure of processing fluency and provide evidence (based on study times) that processing fluency *does* mediate the influence of word-pair relatedness on JOLs. Thus, results seem to be highly dependent on the fluency measure and there may be different kinds of fluency (e.g., perceptual vs. conceptual fluency; cf. Alter & Oppenheimer, 2009; see also Susser & Mulligan, 2015, for the concept of motoric fluency). Importantly, Mueller and colleagues' (2013, 2014, 2016) studies did not directly test whether beliefs are necessary for these manipulations to

affect JOLs. That is, they do not report whether item-level JOL effects are shown by all participants, or only those who held an a priori belief. Likewise, they did not test whether beliefs mediate the effects on JOLs. Such analyses could speak to the role of experience for item-level JOLs without relying on a (debatable) fluency measure.

Overview of the Present Experiments

In the present experiments, we investigate the contribution of beliefs and/or experience to different metacognitive judgments. We focused on the volume manipulation (Rhodes & Castel, 2009), for which the question of preexisting beliefs has not yet been addressed. Most importantly, we applied a novel approach to answering this research question consisting of assessing and comparing various metacognitive measures at multiple time-points throughout the study-test cycle.

To establish whether participants have preexisting beliefs about volume and memory, we described the volume manipulation to participants (before they studied/experienced any words in these volumes) and asked them to estimate the percentage of loud and quiet words they would remember. Such ratings are commonly referred to as *global differentiated predictions* (GPREDs; Kornell et al., 2011), and they are similar to the predictions elicited after vignettes by Mueller and colleagues (2013, 2014, 2016). Importantly, because they are made before any experience of the volume manipulation, they tap into preexisting beliefs. We also used *poststudy* GPREDs (Experiment 3) to assess whether beliefs changed as a result of experiencing the volume manipulation (note that these are also referred to as *aggregate JOL*s in the literature, e.g., Besken, 2016; Besken & Mulligan, 2013, 2014; but we reserve JOL for item-level JOLs in the present manuscript).

During the study phase, we assessed traditional immediate *item-level JOLs* in all three experiments to capture (and replicate) in-the-moment effects of the volume manipulation. As described, Mueller and colleagues (2013, 2014, 2016) demonstrated preexisting beliefs about various manipulations and concluded that these explain item-level JOLs. By assessing both GPREDs and item-level JOLs, we can, for the first time, directly test this assumption: If item-level JOLs are based on beliefs only, with no contribution of in-the-moment experience, there should not be quantitative or qualitative differences between volume effects on (pre- and post-study) GPREDs and item-level JOLs. Additionally, JOLs for loud and quiet words should not differ after controlling for preexisting beliefs about volume and memory, or in participants who do not demonstrate beliefs about volume and memory.

Lastly, in Experiments 1 and 2, we assessed whether *global differentiated postdictions* (GPOSTs) of memory performance following the test are based on beliefs only (i.e., show equivalent volume effects as GPREDs), or if participants instead rely on their study experience and/or their ability to monitor and aggregate performance information during testing (i.e., show equivalent volume effects as either JOLs or memory performance). Thereby, the GPOST measure speaks to the persistence of metacognitive beliefs about volume (resulting from preexisting beliefs and/or experience, as to be tested) over time.

Thus, this novel combined assessment of different metacognitive measures at different points during the study-test cycle allows us to examine the dynamic interplay of beliefs and experience to item-level JOLs within each of our three experiments. This is in contrast to previous research which focused on either beliefs or experience within a given experiment (but see Besken, 2016, Experiment 3 for an exception).

Experiment 1

In addition to comparing volume effects on GPREDs, item-level JOLs, and GPOSTs, as just explained, Experiment 1 explored whether beliefs about volume also influence confidence judgments (CJs). Notably, Mueller and colleagues (2013, 2014, 2016) focused on the effects of beliefs on memory predictions, but recent research has begun to investigate the impact of beliefs on other types of metacognitive judgments. Specifically, McDonough and Gallo (2012) investigated the effect of beliefs about font size on recognition decisions and recognition-confidence judgments. In different tests, they asked their participants to exclusively recognize small or large font words only. If participants believe that large font words are easier to remember, then they should be reluctant to guess that a large font word was studied but forgotten—that is they should make fewer false alarms when asked to recognize only large font words compared with participants asked to recognize only small font words. Although McDonough and Gallo (2012) indeed observed fewer false alarms in the large font recognition condition, there was no difference in average confidence judgments across conditions. Thus, beliefs influenced participants' decision criteria, but not their recognition confidence. This suggests that CJs are not influenced by beliefs. However, the null effects of font-size on CJs in this study may have resulted either from uncertainty that a word was presented at all, or from uncertainty about the *font size* the word was presented in. To more clearly examine whether beliefs influence CJs, we used a standard old/new recognition paradigm, but provided participants with volume-information (accurate for old words) after their decision, so that there would not be any uncertainty about what volume a word was presented in. We can then examine whether participants are more confident in their recognition responses when they are told that a word was presented in a loud volume, consistent with their beliefs about volume effects, or if they instead rely primarily on other, experience-based cues, available in the test moment.

Method

Participants and materials. Fifty-two Introduction to Psychology students at the University of North Carolina at Greensboro participated in exchange for partial course credit. Loud words were three times as loud as quiet words. Stimuli were presented via E-Prime (Schneider, Eschman, & Zuccolotto, 2002), which also controlled changes in volume. Specifically, quiet words were set to 15.85 decibels quieter than the computer-set volume whereas loud words were played at the computer-set volume (Dell Optiplex 760 computers running Windows 7 with the volume set at 50%). Stimuli were played through Sennheiser HD 202 headphones.

One hundred English nouns from the Paivio, Yuille, and Madigan (1968) Word Pool were selected. Concreteness ratings (1–7; higher is more concrete) ranged from 1.18 to 4.87 (M = 2.79). Imageability (1–7; higher is more imageable) ranged from 1.93 to 4.80 (M = 3.41). Francis and Kuchera (1982) word frequency ranged from 1 to 100 (M = 34.30). We recorded each word spoken

by the same male voice. For each participant, half of the words were randomly assigned to the loud volume and the other half to the quiet volume. Half of the loud words and half of the quiet words were then assigned for study with the others serving as distractors during the recognition test. The order of the words during the study and test phases was randomized.

Procedure.

Global predictions and instructions. Participants were informed that they would study a list of words one-at-a-time, with half of the words presented loudly and half presented quietly. Participants only received a verbal description of the volumes, no sound files were played during these instructions. Participants were also explicitly told that they would *not* have to remember whether a word was presented at a loud or quiet volume. Participants were next asked to estimate what percentage of loud and quiet words (counterbalanced order) they would remember (GPREDs). To make their estimate, participants entered a value between 0 (labeled as "no loud/quiet words will be remembered") and 100 (labeled "all loud/quiet words will be remembered") via the keyboard.[1]

Study and judgments of learning. Participants were then instructed on the item-level JOL procedure. This procedure was computerized, but otherwise mimicked that of Rhodes and Castel (2009). Participants were given five seconds to study and enter their JOL beginning with the onset of the word audio, which was played once. The word did not appear anywhere on the screen. A countdown timer appeared on the screen when three or fewer seconds remained. The JOL screen remained until the full five seconds had expired. During this time participants could enter or change their JOL. Immediately after the five seconds the next word was played. Participants were given two practice trials with this procedure prior to the actual trials. During this practice phase, the words "idea" and "effort" were presented at medium volume (7.925 decibels less than the loud words). After the practice trials, participants had the opportunity to reread the instructions. If they chose to reread the instructions they were also prompted to redo the GPREDs and practice trials.

Participants then studied the 50 words, one-at-a-time, making JOLs for each word immediately following presentation. JOLs were typed using the keyboard and spanned the range of whole numbers between 0% ("definitely will not remember") and 100% ("definitely will remember"; see Footnote 1).

Recognition test and confidence judgments. Following the study phase, participants completed a filler task involving math verification problems (e.g., 4 − 3 = 2, True/False?) for five minutes. Participants were then instructed on and completed the recognition test and CJs. Participants were told that they would see a series of words, half of which were from the list they studied, and half from a list studied by another participant. They first indicated whether a word was "old" (from their studied list) or "new" (not from their studied list). We were additionally interested in whether loud words would be given higher confidence ratings. To avoid the possibility that participants may not recall which words were loud or quiet, we provided the volume information beneath the word during the CJ but not during the old/new judgment. That is, after making the old/new judgment, the word "loud" or "quiet" appeared beneath the word. If the word had been from the study list, this volume information always corresponded to the actual volume the word had been played at; for the distractor words (presumably from another participant's study list to give meaning to the volume information), a random half received a loud cue and the other half received a quiet cue. To measure CJs on the same scale as the other metacognitive measure, participants were then asked to rate their confidence in their old/new response from 0% (*not at all confident*) to 100% (*absolutely confident*). As an additional label, 50% was classified as *somewhat confident*, to avoid participants thinking that 50% means guessing (see Footnote 1). Recognition and confidence judgments were self-paced. Note that because the words loud and quiet were presented visually (and in the same size font), the volume was not reexperienced during the recognition test. Thus, any effects of volume could only be transferred to CJs through beliefs about volume.

Global postdictions. Following the recognition test and CJs, participants were asked to predict what percentage of loud and quiet words they actually got correct on the recognition test (GPOST) on the same scale as the other metacognitive judgments. Again the order of these predictions was counterbalanced. Lastly, participants were asked a series of questions regarding their motivation and task experience.

Results

For all analyses, alpha was set to .05.

Global predictions. Even before studying any words, participants predicted that they would remember more loud words than quiet words on average (see Figure 1), $t(51) = 6.34$, $p < .001$, $d_z = 0.88$. These predictions were made before any fluency information was presented and thus can only be driven by beliefs. This pattern held for most participants, with 75% giving a higher GPRED to loud than quiet words, 17% giving equal ratings for both volumes, and only 8% indicating that quiet words would be remembered better than loud words.

Judgments of learning. JOLs greater than 100 were considered typos and eliminated prior to analysis (0.2% of all JOLs).[2] As expected, loud words received higher average JOLs relative to quiet words (see Table 1), $t(51) = 4.94$, $p < .001$, $d_z = 0.68$. This pattern occurred for 81% of participants; the remaining 19% predicted better memory for quiet words (no participant had mean

[1] With base-rate appropriate old-new guessing, chance performance is 50 rather than 0 in this recognition test. However, there are individual differences in response criterion (e.g., Kantner & Lindsay, 2012) and recognition hits below 50% are possible for conservative participants. For this reason, and for the purpose of keeping maximal comparability to Rhodes and Castel's (2009) study and our recall conditions (Experiment 2), we opted to use a 0 to 100 scale rather than 50 to 100. As indicated, we provided clear verbal labels for our participants, asking them to predict word memorability, not recognition hits. If participants considered the 50% guessing probability, rendering values below 50 little likely, they would have to consider this for both quiet and loud words equally, thus not affecting the relative comparison of the predictions we were most interested in. As evident from the mean values though, participants used the full scale as instructed.

[2] In each of the three experiments, missing JOLs constituted fewer than 1% of all JOLs. JOLs greater than 100 were considered typos and eliminated prior to analysis (less than 0.5% of all JOLs). JOLs less than 10 (less than 5% of all JOLs) may also indicate a failure to enter a JOL in time (e.g., wanting to type 80, but only entering the 8 prior to the response deadline) but because many of them could also be intentionally low entries all of them were retained for analysis. In all three experiments, separate analyses removing these low JOLs did not change the pattern of results.

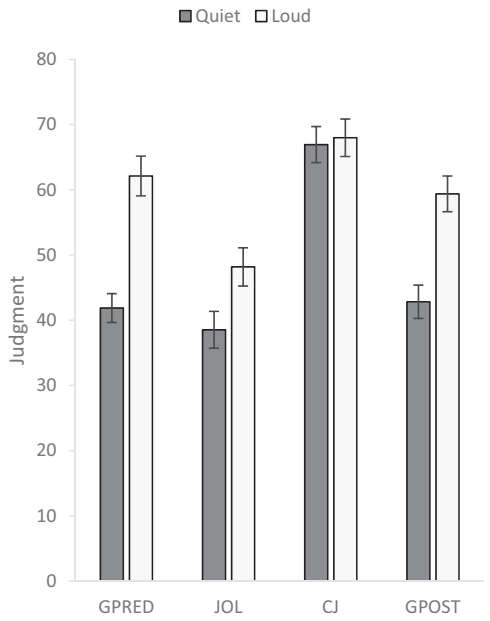


Figure 1. Experiment 1 metacognitive predictions. GPRED = Global differentiated prediction; JOL = Judgment of learning; CJ = Confidence judgment in old/new response; GPOST = Global differentiated postdiction. Error bars display the standard error of the mean.

JOLs exactly equal for quiet and loud words). Within-participant gamma correlations indicated that JOLs were reasonably predictive of later recognition performance, $M = .21$, $SE = .04$, $p < .001$.

Recognition performance. Because lures did not have volume information associated with them during the recognition judgment (only during the following confidence judgment), we focus our analyses on hit rates. Hits were slightly higher for loud words relative to quiet words (see Table 1), $t(51) = 2.73$, $p = .009$, $d_z = 0.41$.[3]

Confidence judgments. If participants reference their beliefs about volume and memory when making CJs, then it is conceivable, that the volume information presented after the recognition judgment would influence CJs (regardless of whether the word was studied or a lure). However, CJs did not differ by volume information when analyzing all responses (hits, missed, correct rejections, and false alarms; Table 1), $t(51) = 1.14$, $p = .259$, or when restricting the analysis to hits (loud: $M = 82.41$, $SE = 2.26$, quiet: $M = 82.60$, $SE = 2.13$), $t(51) = 0.14$, $p = .894$. These results suggest that participants did not use the visually provided volume information when making CJs. Within-participant gamma correlations indicated that CJs were reasonably predictive of recognition performance, $M = .37$, $SE = .04$, $p < .001$.

Global postdictions. Following the recognition test, participants continued to believe they had performed better on loud than quiet words (see Figure 1), $t(51) = 7.12$, $p < .001$, $d_z = 0.99$. As with GPREDs, most participants (77%) indicated that loud words were remembered better than quiet words, with only 17% of participants indicating equal memory for both volumes, and only 6% indicating better memory for quiet words. When classifying participants as predicting better memory for loud words, quiet words, or no difference, 63% of the participants were classified the same in both their GPREDs and GPOSTs.

Joint contributions of beliefs and experience-based factors. As indicated earlier, the majority of participants had a preexisting belief that loud words would be remembered better than quiet words. Notably, of the 13 "non-believers" (participants who did *not* believe that loud words would be better remembered than quiet words), 10 (77%) subsequently gave higher JOLs to loud words. Thus, preexisting beliefs do not appear necessary for volume to affect JOLs.

To formally test whether beliefs about volume fully account for the effect of volume on JOLs, we subtracted each participant's GPRED for quiet words from their GPRED for loud words. This provides a measure of how much better (positive difference scores) or worse (negative difference scores) each participant (prestudy) believes they will perform on loud compared to quiet words—hereafter referred to as "beliefs." We then entered beliefs and presentation volume into a MIXED model predicting JOLs (see Table 2; SAS Proc MIXED; Littell, Milliken, Stroup, & Wolfinger, 2000). For volume, quiet words were the reference group (0 = quiet, 1 = loud). Thus, the intercept is the average JOL for quiet words when a participant predicts equivalent memory for both volumes (belief = 0).

The significant main effect of volume indicates that average JOLs were higher for loud relative to quiet words even when participants did not believe that volume affected memory (belief = 0). There was no main effect of belief, indicating that average JOLs for quiet words did not significantly differ with the magnitude of beliefs. Importantly, the significant Belief × Volume interaction indicates that there was a significant positive slope between the magnitude of one's beliefs and the magnitude of one's volume effect on JOLs. The more strongly people believed (prestudy) that loud words would be remembered better, the larger JOLs they gave to loud compared to quiet words during study. That is, beliefs inflated JOL for the loud words, but did not significantly affect the quiet words. Importantly, this model revealed that both in-the-moment experience-based factors (main effect of volume) and beliefs (Belief × Volume interaction) independently contribute to volume-effects on item-level JOLs.

Discussion

Participants demonstrated evidence of preexisting beliefs about volume similar to Mueller and colleagues' (2013, 2014, 2016) conclusions regarding relatedness, font size, and the identicality effect. However, even participants who did not indicate (via GPREDs) an initial belief that loud words would be more memorable than quiet words gave higher in-the-moment JOLs to loud

[3] We also computed d' for the recognition data from Experiments 1 and 2 (see Table 1). Note that in Experiment 1 we used the false alarm rate across both loud and quiet distracters for computing d' for both loud and quiet words because volume information was provided *after* the recognition decision. To correct for participants with false alarm or hits rates of zero, we followed Hautus' (1995) recommendation and added 0.5 to each participant's total hits and false alarms. These values were then divided by 26 (one more than the number of old/new items) to compute corrected hit and false alarm rates. These corrected rates were then used to compute d' scores. Analyses on d' yielded the same pattern of results as the reported analyses on hit rates.

Table 1
Means for Memory Performance in Experiments 1 and 2

	Experiment 1		Experiment 2							
	Recognition		Recognition				Recall			
	Large dose		Small dose		Large dose		Small dose		Large dose	
Variable	Quiet	Loud	Quiet	Loud	Quiet	Loud	Quiet	Loud	Quiet	Loud
d′	1.58 (.09)	1.74 (.09)	1.91 (.11)	1.92 (.11)	1.83 (.13)	2.00 (.13)	—	—	—	—
Hits/Recall	74.31 (1.73)	78.54 (1.60)	79.89 (2.11)	80.34 (1.90)	74.88 (2.62)	80.00 (1.98)	24.2 (4.30)	23.5 (4.02)	19.9 (3.14)	22.6 (3.69)
FA/Com	19.46 (1.17)		17.89 (1.51)		16.25 (2.14)		1.31 (.24)		.82 (.18)	

Note. Values in parentheses indicate standard errors of the mean. Hits/Recall = Percentage of items correctly endorsed as "old" in the recognition condition and percentage of items correctly recalled in the recall condition. FA = False alarms, percentage of items incorrectly endorsed as "old" in the recognition condition. Com = Commissions, number of items recalled that were not presented in the recall condition. Note that false alarms and commissions do not have volume information associated with them.

words than quiet words, and both beliefs and volume contributed independent variance in JOLs.

Unlike Rhodes and Castel (2009), who used a free recall test, we did find a small recognition memory benefit for louder words. This finding suggests that beliefs about volume and memory are not entirely inaccurate. Notably, Foster and Sahakyan (2012) previously documented a recall-benefit for loud words but only following instructions to forget, not standard encoding instructions as in our study. Comparably, McDonough and Gallo (2012) documented a recognition-benefit for large-font words but only when separately testing for large and small font words, not in a standard recognition test like ours. To foreshadow, Experiment 2 also provided evidence for a small volume effect on recognition memory. We discuss volume-effects on memory further in the General Discussion.

Notably, beliefs about volume did not influence CJs. These results are in line with McDonough and Gallo's (2012) study that found little evidence for effects of font-size related beliefs on confidence judgments. Experience-based cues available during the test, such as recollection (Robinson, Johnson, & Robertson, 2000), retrieval fluency (Kelley & Lindsay, 1993; Robinson et al., 2000), familiarity (Yonelinas, 1994), or fluency of processing the recognition cue (Chua, Hannula, & Ranganath, 2012), and/or other beliefs might have influenced CJs. The lack of volume effects on CJs is informative in that it shows that beliefs about cues (here: volume), even when quite pronounced, are not always referenced when making metacognitive judgments. It could be argued that CJs are not affected by volume manipulations because participants realize during the test that their beliefs about volume were inaccurate. However, the continued effect of volume on GPOSTs argues against this explanation and instead suggests that beliefs are not referenced for all metacognitive judgments.

Although the results of Experiment 1 established that people have preexisting beliefs regarding volume and memory, they also suggest that these preexisting beliefs cannot *fully* account for item-level JOLs (and not at all for item-level CJs). To our knowledge, this is the first study to relate a belief-measure to item-level JOLs. Previous experiments either measured beliefs (via vignettes or prestudy JOLs) or item-level JOLs separately and then speculated about their relationship (e.g., Mueller et al., 2013, 2014, 2016) without formally assessing it. In the following experiments, we thus strove to replicate the important finding that beliefs alone cannot fully account for item-level JOLs. Further, we added a manipulation of the volume experience to provide stronger evidence for experience-based effects.

Experiment 2

In Experiment 2, we aimed to further demonstrate the influence of experience on JOLs. We created two doses of the volume manipulation by reducing the perceptual difference between loud and quiet stimuli by 50% in one condition, and maintaining the large perceptual difference in the other. If beliefs about volume are absolute (i.e., louder words are always remembered better than quiet words, no matter the dosage) and item-level JOLs are solely belief-based, then volume-differences in GPREDs, JOLs, and GPOSTs should be comparable in both dose conditions. However, if JOLs are based on in-the-moment experiences, then the difference between JOLs for loud and quiet words should be smaller in the low dose condition. Of course, it is also possible that beliefs about volume and memory are graded, such that a larger volume difference is believed to produce a larger effect on memory. In this case, GPOSTs, made after study experience, should be sensitive to the dose manipulation as well. Additionally, if graded beliefs about volume and memory exist, then they should account for variance in the volume-JOL relationship under both dose conditions.

We also sought to further investigate the effects of volume on memory. Therefore, we manipulated test format to (a) replicate Rhodes and Castel's (2009) finding that volume does not impact free recall performance, and (b) test whether the influence of volume on recognition memory is dose sensitive.

Lastly, neither Experiment 1 nor the prior research on JOLs and beliefs (Mueller et al., 2013, 2014, 2016), address the possibility that participants may have only weak, if any, prior beliefs regarding volume and memory, but may give different predictions based on demand characteristics. That is, when given only information about volume, participants may assume that they are supposed to use this information when making their GPREDs despite the absence of *strong* beliefs regarding the relationship between the two. Indeed, the relationship may have never occurred to the participant prior to the study (cf. Dunlosky et al., 2015). In an attempt to address this question, we added second-order confidence judgments for the GPREDs. Specifically, we asked participants to rate their confidence in how many words they would remember on average and

Table 2
Mixed Model Results on the Joint Contributions of Beliefs and Experiential Effects

Experiment	Estimate	*SE*	*df*	*t* value	*p* value
Experiment 1					
Intercept	37.19	3.61	50	10.30	<.001
Volume	7.21	1.24	51	5.81	<.001
Belief	.07	.12	50	.56	.578
Belief × Volume	.12	.04	2520	2.96	.003
Experiment 2					
Intercept	54.64	3.25	132	16.83	<.001
Volume	.65	1.06	134	.61	.540
Dose	−.11	.10	132	−1.15	.252
Belief	−.08	.14	132	−.55	.585
Volume × Dose	5.41	1.52	134	3.56	<.001
Belief × Dose	−.06	.19	132	−.31	.755
Belief × Volume	.18	.05	6577	3.91	<.001
Belief × Volume × Dose	.12	.06	6577	1.96	.050
Experiment 3					
Intercept	64.45	3.38	83	19.06	<.001
Volume	4.55	1.23	85	3.69	<.001
Dose	−17.46	5.80	83	−3.01	.004
Belief	−.29	.14	83	−2.11	.038
Volume × Dose	1.30	2.11	85	.61	.541
Belief × Dose	.78	.26	83	2.96	.004
Belief × Volume	−.10	.05	4190	−2.01	.045
Belief × Volume × Dose	.07	.10	4190	.72	.469

Note. Regression weights were computed using SAS Proc MIXED with maximum likelihood estimation and a compound symmetry covariance structure. For volume, 0 = quiet volume, 1 = loud volume. For dose, 0 = small dose, 1 = large dose. Beliefs indicate the advantage for loud words over quiet words (loud GPRED – quiet GPRED).

then asked them to make another prediction based on how confident they were in any volume differences they predicted. Thus, participants could indicate an effect of volume (large, small, or null), but indicate that this was essentially a guess in which they have very little confidence or a strong belief they are confident in. We also assessed confidence in GPOSTs to track potential changes in confidence in beliefs about volume after study-test experience.

Method

Participants. One hundred thirty-six Introduction to Psychology students at the University of North Carolina at Greensboro participated in exchange for partial course credit. One participant was excluded from analysis because they gave JOLs of 100 for every word, which more likely indicates goal-setting as opposed to actual performance estimation (Connor, Dunlosky, & Hertzog, 1997).

Design and procedure. Experiment 2 used a 2 (Dose: small, large) × 2 (Test: recall, recognition) × 2 (Volume: loud, quiet) design, with volume as a within-subjects factor and random assignment of participants to the between-subjects factors test and dose. The procedure and stimuli were identical to Experiment 1 unless noted in the additions and changes described next.

Global prediction confidence ratings. Following their GPREDs, participants were asked to make two confidence judgments regarding their GPREDs. The Participant's GPREDs were summarized on the screen as follows with the higher prediction first, "You indicated that you would remember __% of the [quiet/loud] words and __% of [loud/quiet] words." Participants were then asked, "How confident are you that you will remember more [quiet/loud] words than [loud/quiet] words [or: the same number of loud and quiet words]?" Participants were next asked, "How confident are you that you will remember [average of the two GPREDs]% of the words overall?" This latter question was added to allow participants to indicate their confidence in the magnitude of their overall memory performance and discourage participants from incorporating this information into the prediction pertaining to the volume effect. Because this latter confidence judgment did not address the questions of interest we do not report on these data. Participants then completed the practice stimuli (both at a medium volume) and were given the opportunity to reread the instructions and redo the GPREDs as in Experiment 1.

Study and judgments of learning. The study and JOL procedures were identical to those in Experiment 1. For the large dose condition, loud words were three times louder than quiet words as in Experiment 1. For the small dose condition, the loud words were two times (7.925 decibels) louder than the quiet words. Quiet words in the small dose condition were played at 10.567 decibels quieter than the computer-set volume whereas loud words were played at 2.642 decibels quieter than the computer-set volume.[4] This produced the same average volume in both the small and large dose conditions, with the words in the small dose condition all being closer to the volume midpoint than in the large dose condition.

Memory tests. Following the study phase, participants completed the same filler task as in Experiment 1. The same recognition memory procedure was used in Experiment 2, except that CJs were not collected. For the recall condition, participants typed as many of the studied words as they could remember (self-paced).

Global postdictions and second-order confidence judgments. Following the test phase, participants were asked to estimate how many words of each volume they actually remembered correctly. As with GPREDs, GPOSTs were then summarized on the screen, and participants were asked to make two second-order confidence judgments.

Results

Data were analyzed using 2 (Dose: small, large) × 2 (Test: recall, recognition) × 2 (Volume: quiet, loud) mixed ANOVAs, except when noted otherwise. Alpha was set to .05.

Global predictions. As with Experiment 1, a main effect of volume on GPREDs indicated that participants had pronounced preexisting beliefs that loud words will be easier to remember than quiet words (see Figure 2), $F(1, 131) = 105.39$, $p < .001$, $d_z = 0.89$. There were no main effects of dose or test type, Fs < 1. There were no Test × Dose, $F(1, 131) = 3.64$, $p = .059$, Test × Volume, $F < 1$, Dose × Volume, $F < 1$, or three-way interactions, $F < 1$. These null effects and null interactions are expected as there were no condition differences prior to the GPREDs. The majority of participants (67%) gave higher GPREDs to loud than to quiet words, 30% gave equal ratings for both volumes, and only 2% indicated that quiet words would be remembered better than loud words.

[4] Informal pilot testing with student assistants confirmed a noticeable difference in volume dose between the two conditions.

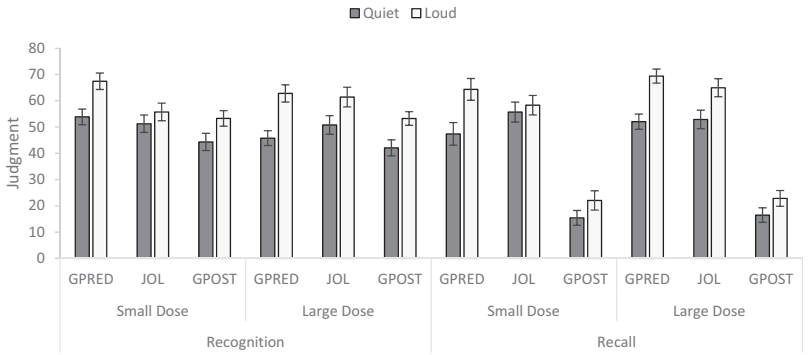


Figure 2. Experiment 2 metacognitive predictions. GPRED = Global differentiated prediction; JOL = Judgment of learning; GPOST = Global differentiated postdiction. Error bars display the standard error of the mean.

Confidence. To assess the strength of participants' preexisting beliefs about volume we first categorized participants as either believing loud words were more memorable or believing loud words were no more memorable. Because only three participants indicated that they would remember more quiet than loud words, this group was excluded from analysis. We analyzed confidence in the direction of their prediction using a 2 (Dose: small, large) × 2 (Test: recall, recognition) × 2 (Belief Direction: loud, equal) between-subjects ANOVA. Confidence in these predictions was generally high (see Table 3). A main effect of test, $F(1, 131) = 4.82$, $p = .030$, $d = 0.45$, resulted from participants in the recognition condition being somewhat more confident in their predictions regarding volume than in the recall condition. Confidence did not differ by dose or belief direction and there were no significant interactions, all Fs < 1. Overall these results suggest that preexisting beliefs about volume, even if generated on the spot, are held in relatively high confidence.

Judgments of learning. JOLs greater than 100 were considered typos and eliminated prior to analysis (less than 0.1% of all JOLs). Louder words received higher average JOLs relative to quiet words (see Figure 3), a main effect of volume, $F(1, 131) = 61.77$, $p < .001$, $d_z = 0.63$. There were no main effects of dose or test, Fs < 1. However, there was a Volume × Dose interaction, $F(1, 131) = 17.93$, $p < .001$, $\eta^2 = .14$, resulting from a greater volume effect on JOLs in the large dose condition. Importantly, the effect of volume on JOLs was significant in both the large, $t(65) = 7.99$, $p < .001$, $d_z = 0.98$, and small dose conditions, $t(68) = 2.81$, $p = .007$, $d_z = 0.34$. As expected there were no interactions involving Test, Fs < 1 (participants did not yet know which type of test they would receive). Within-participant gamma correlations indicated that JOLs were, at best, only moderately predictive of recognition performance in the large, $M = .13$, $SE = .07$, $p = .059$, and small dose conditions, $M = .18$, $SE = .06$, $p = .004$. By contrast, JOLs were somewhat better predictors of recall performance in both the large, $M = .40$, $SE = .83$, $p < .001$, and small dose conditions, $M = .32$, $SE = .06$, $p < .001$.

Memory performance. We compared hit rates and free recall proportions separately for their respective conditions via 2 (Dose: small, large) × 2 (Volume: quiet, loud) mixed ANOVAs.

Recognition performance. There was no main effect of dose, $F < 1$. The main effect of volume, $F(1, 65) = 5.81$, $p = .019$, $d_z = 0.23$, resulted from slightly higher hit rates for loud words (see Table 1). The Dose × Volume interaction, $F(1, 65) = 4.06$, $p = .048$, $\eta^2 = .06$, was attributable to a significant effect of volume on recognition hits (loud > quiet) in the large dose condition, $t(31) = 2.88$, $p = .007$, $d_z = 0.51$, but not in the small dose condition, $t(34) = 0.30$, $p = .763$.

Recall performance. Because the first four letters were necessary to uniquely distinguish all study words, we scored as correct any response that matched the first four letters of a studied word. The remaining commissions were then examined and any obvious misspellings were manually recoded as their correctly spelled counterparts (2.6% of all recall responses). There was no main effect of volume on recall performance, $F < 1$. Likewise, there was no main effect of dose or Dose X Volume interaction, Fs < 1. Thus, we replicate Rhodes and Castel's (2009) finding that volume does not influence recall.

Global postdictions. Not surprisingly, participants in the recall condition estimated that they scored lower on the memory test than participants in the recognition condition (i.e., main effect of test; Figure 1), $F(1, 131) = 109.07$, $p < .001$, $d = 1.77$. A main effect of volume indicated that participants continued to believe that loud words were remembered better than quiet words, $F(1, 131) = 48.42$, $p < .001$, $d_z = 0.60$. The main effect of dose was not significant, $F < 1$. The Test × Volume interaction was not significant, $F(1, 131) = 2.24$, $p = .137$. No other interactions were significant, all Fs < 1. Thus, participants exited the experiment believing that they generally performed better on loud items relative to quiet items and this belief was independent of the magnitude of the volume difference. Most participants (58%) indicated that they remembered loud words better than quiet words, with 30% of participants indicating equal memory for both volumes, and only 12% indicating better memory for quiet words. When classifying participants as predicting better memory for loud words, quiet words, or no difference, 58% of the participants were classified the same in both their GPREDs and GPOSTs.

Confidence. As with GPREDs, we assessed the strength of participants' beliefs about volume at the end of the study by first categorizing participants as either believing loud words were more memorable or believing loud words were no more memorable. Again we excluded those who believed that quiet words would be remembered more often than loud words due to inadequate sample

Table 3
Confidence in Directional Predictions for GPREDs and GPOSTs in Experiments 2 and 3

	Experiment 2								Experiment 3			
	Recognition				Recall				No Test			
	Small dose		Large dose		Small dose		Large dose		Small dose		Large dose	
Variable	*n*	*M*	*n*	*M*	*n*	*M*	*n*	*M*	*n*	*M*	*n*	*M*
GPRED Confidence												
Loud > Quiet	22	71.23 (5.18)	20	66.75 (6.03)	24	79.58 (3.72)	25	78.36 (3.66)	31	70.77 (4.46)	37	72.64 (3.27)
Quiet > Loud	0	—	1	35.00 (—)	2	37.50 (12.5)	0	—	3	74.00 (13.01)	1	75.00 (—)
Quiet = Loud	13	61.15 (6.84)	11	72.73 (8.10)	8	77.50 (8.18)	9	74.88 (8.58)	10	78.00 (9.10)	5	63.00 (12.81)
GPOST/GPRED 2 Confidence												
Loud > Quiet	22	57.45 (7.39)	18	66.89 (6.86)	19	45.74 (9.06)	19	43.68 (6.91)	33	72.91 (3.25)	31	63.61 (3.19)
Quiet > Loud	4	56.50 (16.40)	5	61.00 (9.27)	5	44.40 (19.15)	2	52.50 (37.50)	1	65.00 (−)	4	60.00 (7.07)
Quiet = Loud	9	69.44 (9.15)	9	60.00 (7.77)	10	50.50 (12.17)	13	50.54 (10.03)	10	58.00 (8.63)	8	71.75 (9.49)

Note. Confidence judgments for directional predictions were made from 0 to 100%. Values in parentheses indicate standard error of the mean. GPRED = Global differentiated prediction; GPOST = Global differentiated postdiction; GPRED 2 = Post-study global differentiated prediction. Loud > Quiet: participants predicting better memory for loud words. Quiet > Loud: participants predicting better memory for quiet words. Quiet = Loud: participants indicating equal memory for loud and quiet words.

size (see Table 3). There was a main effect of test, $F(1, 131) = 7.41$, $p = .007$, $d = 0.47$, with participants in the recognition condition being more confident in their postdictions regarding volume than in the recall condition. Confidence did not differ by belief direction or dose, Fs < 1, and there were no significant interactions, Fs < 1.

Joint contributions of beliefs and experience-based factors. Although only 67% of all participants initially believed that loud words would be remembered better, 92% of participants in the large dose condition gave higher average JOLs to loud words relative to quiet words. By contrast, only 61% of participants in the small dose condition gave higher JOLs to loud words. Of the 21 "non-believers" in the large dose condition, 19 (90%) subsequently gave higher JOLs to loud words. By contrast, of the 23 "non-believers" in the small dose condition, only 11 (48%) gave higher JOLs to loud words. Thus, preexisting beliefs about volume were not necessary for volume to affect JOLs—particularly when the volume manipulation was strong.

As in Experiment 1, we further assessed whether beliefs could fully account for the effect of volume on JOLs using a MIXED model approach (see Table 2). However, for this experiment, we added dosage (small = 0, large = 1) and all possible interactions with dosage into the model as well. We did not include test as a factor because test was manipulated after GPREDs and item-level JOLs had been given and, consequently, did not influence those predictions (see prior analyses).

It is possible that the Volume × Dose interaction described earlier (ANOVA results) might partly result from graded beliefs about volume-effects on memory. That is, participants may have given more similar JOLs to loud and quiet words in the small dose condition, because they *believe* that minor differences in volume produce only minor differences in memory, not because of the reduced volume experience. Our MIXED model analyses allowed us to test for such a volume-graded belief-effect (i.e., a Volume × Dose × Belief interaction).

The main effect of volume was not significant, indicating that volume did not influence JOLs (on average) in the small dose condition (dose = 0) when participants did not believe that volume affected memory (belief = 0). The main effect of dose was negative but not significant, thus JOLs were slightly but not significantly lower for quiet words in the small dose condition (relative to the large dose condition, in the absence of beliefs about volume and memory; beliefs = 0). However, the Volume × Dose interaction was significant, indicating that JOLs were higher for loud relative to quiet words in the large dose condition, *even* when participants did not believe that volume affected memory (belief = 0). The main effect of belief and the Belief × Dose interaction were not significant. Thus, how much memory benefit one believed loud words would have relative to quiet words, had no impact on JOLs for quiet words in either the small or large dose conditions. However, the Belief × Volume interaction was positive and significant, indicating a positive slope between the magnitude of one's beliefs and the magnitude of the volume effect on one's JOLs. That is, the more strongly people believed (prestudy) that loud words would be remembered better, the larger JOLs they gave to loud compared with quiet words during study. A positive three-way interaction was just shy of significance. Thus, we cannot confirm that beliefs have a graded influence on JOLs, although the trend suggests that this may be the case. Specifically, the slope between beliefs and the volume effect on JOLs was marginally but not significantly steeper in the large dose condition compared with the small dose condition. Note that although the three-way interaction trend indicates that beliefs *may* be dose sensitive, the significant Volume × Dose interaction—in the presence of this three-way interaction trend—indicates that volume experience influenced JOLs in the large dose condition beyond what was explainable by beliefs alone.

Discussion

Experiment 2 strongly suggests that both beliefs and experience-based cues influence JOLs. As in Experiment 1, the GPREDs made before study suggested preexisting beliefs about volume and memory. Notably, this is the first study to collect second-order confidence estimates for such preexperience predictions, addressing a concern already expressed by Mueller et al. (2014, see also

Dunlosky et al., 2015) that eliciting cue-differentiated metacognitive judgments from participants may induce them to generate ad hoc guesses rather than measuring confidently held preexisting beliefs. Interestingly, the confidence measure suggests that people's confidence in their preexisting beliefs regarding volume and memory is fairly high ($M = 72.37$), arguing against the possibility that participants merely guess on these metacognitive predictions, at least for beliefs about volume. Additionally participants' beliefs sometimes changed following the study and test phase, and confidence in their beliefs declined, particularly in the recall condition.

Despite how confidently people initially held their beliefs about volume and memory, these beliefs only partly accounted for the effect of volume on JOLs. Beliefs in larger volume effects on memory produced larger volume differences in JOLs. Nonetheless, presentation volume during study exhibited independent effects, beyond those of beliefs, on JOLs, but only in the large dose condition. Thus, it appears that stronger perceptual manipulations are more likely to influence JOLs in the absence of preexisting beliefs. However, we caution against overinterpreting the nonsignificant main effect of volume in the mixed model, as 11 participants in the small dose condition *did* give higher JOLs to loud words despite not believing that loud words would be better remembered (out of 23 nonbelievers in this condition).

Experiment 2 found that volume influenced recognition memory such that loud words were better recognized than quiet words, but only in the large dose condition. The equivalent recall performance for loud and quiet words argues against the idea that participants could not hear the quiet words. Importantly, this subtle difference in recognition memory performance cannot explain the persistent belief that loud words are remembered better than quiet words, as GPOSTs demonstrate this continued (albeit less confident) belief even following recall tests where loud words were not remembered better than quiet words. This demonstrates both a persistent and largely exaggerated belief regarding volume and memory, as well as a general lack of metacognitive updating and performance aggregation (see also Dunlosky & Hertzog, 2000; Hertzog et al., 2009).

Whereas our dose manipulation exhibited a clear influence on JOLs, it did not affect GPOSTs. It could be that GPOSTs are instead based on an accurate monitoring and aggregation of performance during the test. However, if this were the case, then GPOSTs should have been more similar for loud and quiet words in accordance with actual memory performance. By contrast, if GPOSTs were based on an accurate aggregation of JOLs, then they should have been similarly affected by the dose manipulation—they were not. It is possible that participants lose access to the volume manipulation, such that they forget just how great or small the volume difference was between words by the end of the filler task and memory test. We investigate this issue further in Experiment 3.

Experiment 3

To better measure poststudy beliefs, Experiment 3 collected a second set of GPREDs immediately following the study phase. If beliefs are adjusted during the study phase, then these poststudy GPREDs may be more sensitive to the dose manipulation of volume.

Method

Participants and materials. Eighty-eight Introduction to Psychology students at Case Western Reserve University participated in exchange for partial course credit. One participant was excluded from analysis because they gave JOLs of 100 for every word, which may indicate goal-setting as opposed to actual performance estimation (Connor et al., 1997).

Stimuli were presented via E-Prime on Dell Optiplex 9030 computers running Windows 7. The computers and headphones used for Experiment 3 required that the volume be set at 25% to approximate the volumes used in the two dose conditions of Experiment 2. Stimuli were played through Sennheiser HD 280 pro headphones and volume was controlled via E-Prime.

Procedure. The procedure was identical to that of Experiment 2 except that a second set of GPREDs were collected immediately after the study phase. For practical reasons we eliminated the test phase and, consequently, the GPOSTs.

Results

Data were analyzed using 2 (Dose: small, large) × 2 (Volume: quiet, loud) mixed ANOVAs, except when noted otherwise. Alpha was set to .05.

Global predictions. As with Experiments 1 and 2, a main effect of volume on GPREDs indicated that participants had pronounced preexisting beliefs that loud words will be easier to remember than quiet words (see Figure 3), $F(1, 85) = 107.58$, $p < .001$, $d_z = 1.12$. There was no main effects of dose, $F < 1$, or Dose × Volume interaction, $F < 1$. These null effects and null interactions are expected as there were no condition differences prior to the GPREDs. The majority of participants (78%) gave higher GPREDs to loud than to quiet words, 17% gave equal

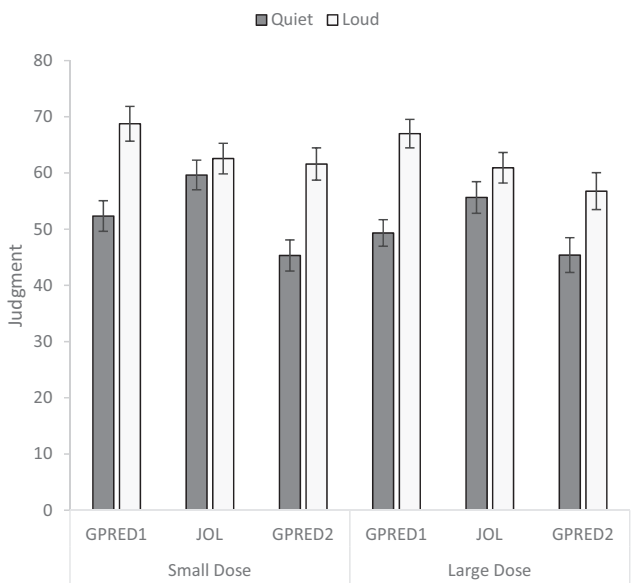


Figure 3. Experiment 3 metacognitive predictions. GPRED 1 = Prestudy global differentiated prediction. JOL = Judgment of learning. GPRED 2 = Post-study GPRED Global differentiated prediction. Error bars display the standard error of the mean.

ratings for both volumes, and only 4% indicated that quiet words would be remembered better than loud words.

Confidence. To assess the strength of participants' preexisting beliefs about volume we first categorized participants as either believing loud words were more memorable or believing loud words were no more memorable. Because only three participants indicated that they would remember more quiet than loud words, this group was excluded from analysis. We analyzed confidence in the direction of their prediction using a 2 (Dose: small, large) × 2 (Belief Direction: loud, equal) between-subjects ANOVA. Confidence in these predictions was generally high (see Table 3). Confidence did not differ by belief direction or dose and there were no significant interactions, all Fs < 1. Overall these results again suggest that preexisting beliefs about volume, even if generated on the spot, are held in relatively high confidence.

Judgments of learning. JOLs greater than 100 were considered typos and eliminated prior to analysis (less than 0.1% of all JOLs). Louder words received higher average JOLs relative to quiet words (see Figure 3), a main effect of volume, $F(1, 85) = 26.63$, $p < .001$, $d_z = 0.55$. There was no main effect of dose, $F < 1$. As in Experiment 2, the effect of volume was descriptively greater in the large dose condition, however this Volume × Dose interaction did not reach statistical significance in Experiment 3, $F(1, 85) = 2.21$, $p = .141$, The effect of volume on JOLs was significant in both the large, $t(42) = 4.79$, $p < .001$, $d_z = 0.73$, and small dose conditions, $t(43) = 2.55$, $p = .014$, $d_z = 0.38$.

Poststudy global predictions. A main effect of volume indicated that participants continued to believe that loud words would be remembered better than quiet words, $F(1, 85) = 67.44$, $p < .001$, $d_z = 0.88$. The main effect of dose was not significant, $F < 1$. In contrast to the analysis for JOLs, the volume effect on poststudy GPREDs was descriptively larger in the *small dose* condition. However, like the effect on JOLs, this Volume × Dose interaction was not significant, $F(1, 85) = 2.12$, $p = .149$. Importantly, an analysis including Time of Predictions (JOL, poststudy GPRED) as a further within-subjects factor, yielded a significant three-way interaction of Volume × Dose × Time of Prediction (JOL, poststudy GPRED), $F(1, 85) = 4.12$, $p = .046$, demonstrating that the effects of dose and volume differed significantly between these two measurements (albeit the nonsignificant Volume × Dose interactions on either measurement). Again, this is attributable to a numerically stronger volume effect on JOLs in the large dose condition whereas the volume effect on GPREDs was numerically stronger in the small dose condition. This finding is inconsistent with the idea that the differences between Prestudy GPREDs and JOLs are the result of beliefs changing during the study phase and then being used to make JOLs for the remainder of this phase.

After studying, most participants (74%) predicted that they would remember loud words better than quiet words. Only 20% of participants predicted equal memory for both volumes, and only 6% predicted better memory for quiet words. When classifying participants as predicting better memory for loud words, quiet words, or no difference, 70% of the participants were classified the same in both their pre- and poststudy GPREDs. Again, this suggests that the study experience had only minimal effects on beliefs about the effects of volume on memory.

Confidence. As with Prestudy GPREDs, we assessed the strength of participants' beliefs about volume at the end of the study phase by first categorizing participants as either believing loud words were more memorable or believing loud words were no more memorable. Again we excluded those who believed that quiet words would be remembered better than loud words due to insufficient sample size (see Table 3). Confidence did not differ by direction, $F < 1$, or dose, $F(1, 85) = 1.11$, $p = .300$. Confidence was descriptively somewhat higher for participants in the large dose condition predicting equal memory for loud and quiet words relative to those in the small dose condition making the same predictions, but this Dose × Direction interaction was not significant, $F(2, 85) = 2.27$, $p = .110$.

Joint contributions of beliefs and experience-based factors. Although only 78% of all participants initially believed that loud words would be remembered better, 84% of participants in the large dose condition gave higher average JOLs to loud words relative to quiet words. By contrast, only 59% of participants in the small dose condition gave higher JOLs to loud words. Of the six "non-believers" in the large dose condition, all six subsequently gave higher JOLs to loud words, yet only two of them (33%) subsequently predicted better memory for loud words in their poststudy GPREDs. Of the 13 "non-believers" in the small dose condition, 10 (77%) gave higher JOLs to loud words and nine of them (69%) subsequently predicted better memory for loud words in their poststudy GPREDs. Thus, preexisting beliefs about volume were not necessary for volume to affect JOLs. Likewise, JOL effects on volume did not necessitate the development of beliefs about volume either. Of the 62 participants who gave higher JOLs to loud words, only 44 (71%) subsequently predicted better memory for loud words in their poststudy GPREDs.

As in Experiment 2, we further assessed whether beliefs could fully account for the effect of volume on JOLs using a MIXED model approach (see Table 2). The main effect of volume was significant, indicating that loud words received higher JOLs *even* when participants did not believe that volume affected memory (belief = 0). Because the Volume × Dose and three-way interactions were not significant, we cannot conclude that the effect of the volume manipulation was substantially larger in the large dose condition in the absence of beliefs about volume and memory. However, the main effects and interactions involving beliefs suggest a *negative* relationship between the volume effect on JOLs and beliefs. Specifically, the main effect of dose was significant, as were the main effect of beliefs and the Belief × Dose and Belief × Volume interactions. Taken together, the negative main effect of dose and the nonsignificant Dose × Volume interaction indicate that JOLs for all words (quiet and loud) were higher in the small dose condition relative to the large dose condition when participants did not believe that volume affected memory (belief = 0). The negative main effect of belief indicates that in the small dose condition, lower JOLs were given to quiet words by participants who predicted larger memory benefits for loud words. The significant Belief × Dose interaction indicates that the negative main effect of dose was less pronounced among participants predicting larger memory benefits for loud words. Unlike Experiments 1 and 2, the Belief × Volume interaction was *negative*, indicating a negative slope between the magnitude of one's beliefs and the magnitude of the volume effect on one's JOLs. Note that the effect of volume was still generally positive for all participants, but in contrast to Experiments 1 and 2, this effect was significantly smaller, not larger, among participants predicting larger memory

benefits for loud words. Nonetheless, consistent with the prior experiments, the results from Experiment 3 continue to show that presentation volume influences item-level JOLs beyond the effects of any prior beliefs—that is, prior beliefs about volume are not necessary for presentation volume effects on item-level JOLs.

Discussion

Experiment 3 did not find evidence that beliefs about volume and memory changed following the study phase. Although dose did not have strong effects in this experiment on item-level JOLs, dose affected JOLs and poststudy GPREDs differently. Specifically, whereas the volume effect on JOLs was numerically stronger in the large compared with the small dose condition, replicating Experiment 2, the dose effect was numerically in the opposite direction for poststudy GPREDS (i.e., numerically larger volume effect in the small dose condition). Experiment 3 replicated the findings of both preexisting beliefs about volume and memory and the effects of volume on JOLs. The dose effect was replicated in principle, however was not as pronounced as in Experiment 2. Finally, Experiment 3 did not provide evidence for a partial mediation of volume effects on JOLs through beliefs (as found in both Experiment 1 and 2); instead, Experiment 3 provides strong evidence for an independent effect of presentation volume experience during study on JOLs, which can operate even in contrast to preexisting beliefs. Keep in mind that most "non-believers" assumed no effect of volume; perhaps, these participants were then particularly surprised by the strong perceptual effect of the volume manipulation. In line with the conclusions from Experiment 1 and 2, Experiment 3 strongly suggests that the volume effect on item-level JOLs does not necessitate preexisting beliefs about volume and memory.

General Discussion

In three experiments, we found evidence for strong preexisting beliefs regarding volume and memory. These beliefs are similar in nature to Mueller et al.'s (2013, 2014, 2016) finding that participants have preexisting beliefs regarding relatedness, font size, and identicality, and may indicate a general preexisting belief regarding fluency and memory. Specifically, we found that the majority of participants believed that loud words would be more memorable than quiet words. Importantly, in Experiments 2 and 3, we found that participants' confidence in their preexisting beliefs regarding volume and memory is quite high.

Whereas Mueller and colleagues (2013, 2014, 2016) concluded that preexisting beliefs also explain item-level JOL effects, our novel approach of comparing global differentiated pre-/postdictions to in-the-moment JOLs made during the study phase revealed systematic differences. First, in all three experiments, the majority of participants gave higher JOLs to loud words regardless of whether they had expressed a preexisting belief that loud volume improves memory (as measured by GPREDs) or not. If preexisting beliefs about volume are not necessary for volume effects on JOLs, then it cannot be argued that beliefs are the only factor driving volume effects on JOLs. Second, in Experiment 2, in-the-moment JOLs were influenced by the dose of the volume manipulation but not later postdictions. Third, and most importantly, merely demonstrating preexisting beliefs about a cue in a pre-JOL or vignette-survey study is not sufficient for the conclusion that a cue's effect on item-level JOLs is based on beliefs alone. Instead individual difference in beliefs must be shown to fully predict individual differences in the effect of such perceptual manipulations on JOLs. In Experiments 1 and 2, both beliefs and volume information contributed unique variance to the volume effect in JOLs. Based on these findings, we argue that beliefs can only partially account for the effects of volume on item-level JOLs, which must further incorporate experience-based cues from the volume manipulation (in the case of Experiment 3 even against the initial belief). Notably, Mueller and colleagues (2013) also found that relatedness of the two words in a pair produced more robust effects on item-level JOLs compared with prestudy JOLs (used to assess beliefs), suggesting the former were influenced by experience-based cues (see Undorf & Erdfelder, 2015, for evidence that processing fluency indeed partially mediates the relatedness-effect on item-level JOLs).

Recently, Besken (2016) used a subliminal priming procedure that reduced response times for naming images, but had no effect on JOLs. By contrast, perceptually masking some of the images with a checkerboard grid influenced both response times and JOLs (masked items were given lower JOLs than unmasked items). Besken concluded that fluency alone did not influence JOLs except through beliefs. By contrast, we found that perceptual experience influenced JOLs above and beyond what could be explained by beliefs and that these experience effects may be magnified as the perceptual manipulation is magnified (i.e., dose manipulation in Experiment 2; but see Experiment 3 for weaker effects of dose). However, for experience-based cues to be used in JOLs in the absence of beliefs, the difference in fluency may need to be fairly salient. In Besken (2016) the effect of the mask on fluency was roughly $d_z = 3.80$. By contrast the effect of the subliminal prime on fluency was only $d_z = 0.76$. Although the prime influenced response times, the increase in fluency may not have been substantial enough to override other cues such as beliefs or idiosyncratic item-level features, or to even be *experienced* as more fluent. By contrast, when fluency was more obviously disrupted by the mask, it may have been sufficient to affect perceived fluency and thus JOLs in addition to, or in spite of, other cues.

Given that fluency effects on memory can vary between recall and recognition tasks (Besken & Mulligan, 2014; Yue et al., 2013), the current study extended beyond the recall-based research by Rhodes and Castel (2009) by employing both recall and recognition tests. There was evidence for a weak volume effect on recognition memory such that recognition was slightly better for loud than for quiet words, at least when volume differences were large. For recall, we replicated Rhodes and Castle's finding of no recall difference between loud and quiet words. Our findings may appear somewhat surprising given that fluency is often associated with worse rather than better memory performance and that fluency effects appear to be more prominent in recall as opposed to recognition performance (Besken & Mulligan, 2013, 2014; Yue et al., 2013). However, they do fit with other evidence suggesting that fluency can improve memory performance (Glass, 2007; Merritt, Cook, Wang, & Lyle, 2014). Nonetheless, even with a small potential positive effect of volume on recognition memory, at the least, participants grossly overestimate the effect of volume on memory and it remains a mystery why participants believe (from

the onset) that loud words will be much better remembered than quiet words.

Despite the strong preexisting beliefs about volume and memory, our results argue against the hypothesis that these beliefs are fully responsible for volume effects seen in *all* metacognitive judgments. Indeed, in all three experiments, volume effects occurred in JOLs independent of (Experiments 1 and 2) or even in contrast to (Experiment 3) what participants *a priori* believed about volume effects on memory. Additionally, in Experiment 1, CJs were *not* influenced by volume information. Taken together, these findings support the view that participants use different cues for different metacognitive judgments (Koriat, 1997). Specifically, experience-based cues are used for item-level metacognitive judgments (here: item-level JOLs and item-level confidence judgments), but not for global-differentiated metacognitive judgments even if these are made after experience with the item material. Instead, more general beliefs appear to be referenced when making poststudy GPREDs and (posttest) GPOSTs. This is akin to a student estimating her exam performance based on her beliefs about different question types (e.g., essay questions vs. multiple choice questions) as opposed to mentally tallying up the proportion of each question type that she believes she got correct. Experience-based information such as fluency, encoding strategies used, item imageability, and recollection of the study experience (Koriat, 1997), may not be referenced during global memory predictions because they cease to be available or because they are not recalled and aggregated accurately (Dunlosky & Hertzog, 2000; Hertzog et al., 2009).

It remains an open question as to whether these experience-based cues contributing to the volume effect involve processing fluency, embodied cognition, or something else (maybe in combinations). Future research should also investigate the generalizability of our findings to other manipulations known to affect item-level JOLs such as font size, identicality, word-pair relatedness, and standard versus alternating capitalization. If other preexisting beliefs (e.g., regarding font size or relatedness) are even more pronounced or held with even higher confidence, they may override experience-based cues during JOLs. However, our data suggest that experience-based cues are often difficult to fully disregard (see also Kelley & Jacoby, 1996), even in the presence of strong beliefs. Finally, future research should explore the impact of beliefs and experience-based cues, such as fluency and embodied cognition, on other metacognitive judgments, including not just performance predictions but also confidence-related judgments to better understand the basis for making various metacognitive judgments. We believe that our methodology of assessing different metacognitive judgments at different time-points of the study-test-cycle will be valuable for such future endeavors. Most importantly, we hope that future studies on the belief versus experience basis of metacognitive effects will strive to assess and relate measures of both factors as opposed to focusing on just one.

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/10.1177/1088868309341564

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138,* 432–447. http://dx.doi.org/10.1037/a0015928

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. http://dx.doi.org/10.1016/0749-596X(89)90016-8

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,* 1417–1433. http://dx.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,* 897–903. http://dx.doi.org/10.3758/s13421-013-0307-8

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

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

Castel, A. D., McCabe, D. P., & Roediger, H. L. (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/10.3758/BF03194036

Chua, E. F., Hannula, D. E., & Ranganath, C. (2012). Distinguishing highly confident accurate and inaccurate memory: Insights about relevant and irrelevant influences on memory confidence. *Memory, 20,* 48–62. http://dx.doi.org/10.1080/09658211.2011.633919

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12,* 50–71. http://www.ncbi.nlm.nih.gov/pubmed/9100268. http://dx.doi.org/10.1037/0882-7974.12.1.50

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462–474. http://dx.doi.org/10.1037/0882-7974.15.3.462

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2015). 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 (Eds.), *Remembering attributions, processes, and control in human memory* (pp. 46–64). New York, NY: Psychology Press.

Foster, N. L., & Sahakyan, L. (2012). Metacognition influences item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1309–1324. http://dx.doi.org/10.1037/a0027868

Francis, W. N., & Kuchera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

Glass, J. M. (2007). Visual function and cognitive aging: Differential role of contrast sensitivity in verbal versus spatial tasks. *Psychology and Aging, 22,* 233–238. http://dx.doi.org/10.1037/0882-7974.22.2.233

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers, 27,* 46–51. http://dx.doi.org/10.3758/BF03203619

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,* 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 (2006), 62,* 155–73. http://doi.org/10.1080/17470210701855520

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging, 24,* 462–475. http://dx.doi.org/10.1037/a0014417

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40,* 1163–1177. http://dx.doi.org/10.3758/s13421-012-0226-0

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35,* 157–175. http://dx.doi.org/10.1006/jmla.1996.0009

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24. http://dx.doi.org/10.1006/jmla.1993.1001

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

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 the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science: A Journal of the American Psychological Society/APS, 22,* 787–794. http://dx.doi.org/10.1177/0956797611407929

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models* (4th ed.). Cary, NC: SAS Institute.

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. http://dx.doi.org/10.1037/a0025548

Merritt, P., Cook, G. I., Wang, M., & Lyle, K. (2014, March). List composition affects metacognitive monitoring of emotional valence. In Paul Merritt, *List composition affects metacognitive monitoring of emotional valence*. North Carolina Cognition Conference, Durham, NC.

Metcalfe, J., & Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *Journal of Memory and Language, 52,* 463–477. http://dx.doi.org/10.1016/j.jml.2004.12.001

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? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 69,* 781–799. http://dx.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. http://dx.doi.org/10.1016/j.jml.2013.09.007

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

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51,* 102–116. http://dx.doi.org/10.1037/0003-066X.51.2.102

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation: Advances in Research and Theory, 26,* 125–173. http://dx.doi.org/10.1016/S0079-7421(08)60053-5

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25.

Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning & Verbal Behavior, 23,* 270–288. http://dx.doi.org/10.1016/S0022-5371(84)90181-6

Rhodes, M. G. (2015). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 90–117). New York, NY: Oxford University Press.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 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. http://dx.doi.org/10.3758/PBR.16.3.550

Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied, 6,* 207–221. http://dx.doi.org/10.1037/1076-898X.6.3.207

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22,* 1014–1019. http://dx.doi.org/10.3758/s13423-014-0768-1

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

Undorf, M., & Erdfelder, E. (2013). Separation of encoding fluency and item difficulty effects on Judgements of Learning. *The Quarterly Journal of Experimental Psychology, 66,* 1–31. http://dx.doi.org/10.1080/17470218.2013.777751

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

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.20.6.1341

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

Received July 22, 2015
Revision received May 25, 2016
Accepted July 22, 2016 ■