Magnitude and accuracy differences between judgements of remembering and forgetting

Michael J. Serra and Benjamin D. England
Department of Psychology, Texas Tech University, Lubbock, TX, USA

Metacognition researchers have recently begun to examine the effects of framing judgements of learning (JOLs) in terms of forgetting (rather than remembering) on the judgements’ magnitude and accuracy. Although a promising new direction for the study of metamemory, initial studies have yielded inconsistent results. To help resolve these inconsistencies, in four experiments we had college students (N=434) study paired associates and make JOLs framed in terms of either remembering or forgetting over two study–test trials. Our goals were to further document the effects of framing on the magnitude and accuracy of JOLs and to consider explanations for why specific patterns tend to emerge. The present experiments provide evidence that (a) judgements of forgetting are psychologically anchored at the midpoint of the JOL scale, whereas judgements of remembering are anchored at a lower point, (b) differences in absolute accuracy (calibration) by frame are largely artefactual and stem from differences in anchoring, (c) differences in JOL magnitude and absolute accuracy by frame do not obtain when memory cues are salient to participants, and (d) a forget frame impairs the relative accuracy (resolution) of JOLs across trials by reducing participants’ reliance on cues such as memory for past test performance.

Keywords: Metacognition; Metamemory; Judgements of learning; Accuracy; Framing.
memory items in terms of the likelihood of remembering (e.g., “How likely do you feel you will be to remember this item on the test?”). Although the accuracy of such judgements (i.e., how well the judgements predict future memory performance) is often above chance, a wealth of metacognition research has demonstrated that such judgements typically demonstrate only modest accuracy in most situations (Dunlosky & Metcalfe, 2009) and—perhaps worse—are prone to errors and illusions (Serra & Metcalfe, 2009). Recently, researchers have begun to examine whether changing the framing of JOLs from being about the likelihood of remembering to being about the likelihood of forgetting improves the accuracy of these judgements (e.g., Finn, 2008; Koriat, Bjork, Sheffer, & Bar, 2004; Kornell & Bjork, 2009; Rhodes & Castel, 2008; Tauber & Rhodes, in press-a; see also Halamish, McGillivray, & Castel, 2011, who only examined judgements of forgetting). In addition to increasing our understanding of how people make metacognitive judgements in general, there is also an applied promise of this research. Affecting the manner in which students make metacognitive self-evaluations of their ongoing learning by simply changing the phrasing of the judgement prompt might be a simple and effective way to improve the accuracy of students’ judgements and the efficacy of their study choices (Finn, 2008; see also Koriat et al., 2004). As we describe next, however, research into this question has thus far yielded inconsistent results, so it is not yet possible to recommend this method as a way to improve monitoring.

Past research on the framing of JOLs

Within the domain of judgement and decision making, framing often seems to change how people make judgements and—more importantly—decisions. For example, prospect theory (Kahneman & Tversky, 1979; for a review of findings see Kahneman & Tversky, 1984) suggests that people view potential gains and losses differently, with losses having larger psychological effects than gains of the same magnitude (e.g., losing 20 dollars affects someone psychologically more than does gaining 20 dollars). Because of this difference, people typically work to avoid losing resources (i.e., loss aversion). Furthermore, people will favour less risky bets when framed in terms of gain but will favour riskier bets when framed in terms of loss. For example, when given a choice to either for certain save one third of a population from a disease or the riskier option that could either save the entire population or none of it, people will opt for the certain option (for certain save one third versus possibly saving everyone) when framed in terms of saving lives, but will opt for the riskier option (possibly losing no one versus for certain losing two-thirds) when framed in terms of loss of lives (see Kahneman & Tversky, 1984). Although the choices are exactly the same mathematically regardless of how they are framed, the different frames focus people on two different aspects of the scenario (saving lives versus the loss of life), and this leads them to make different decisions under the two frames.

Although framing is a well-established topic of study within the decision-making literature, examining the effects of framing on metacognitive judgements such as JOLs is a fairly new endeavour. As we describe in the following sections, some studies to date have found that different frames (i.e., remembering versus forgetting) affect JOLs and restudy decisions, whereas other studies have found no differences by frame. To further complicate matters, within the set of studies demonstrating differences by frame, the effects have not been consistent across studies.

Within the metamemory literature, Koriat et al. (2004) first demonstrated that framing memory predictions in terms of forgetting instead of remembering (e.g., “How likely do you feel you will be to forget this item?”) could affect the judgements. Participants in their experiments were asked to imagine that other students had studied 60 paired associates and to indicate how many items they thought that person would either remember or forget over a given time span (e.g., a week; a year). Surprisingly, participants did not factor retention intervals into their remember-framed predictions (Figure 1, Panel A) when retention intervals were varied between subjects (e.g., they...
judged that items would be equally likely to be remembered 10 minutes later as a year later). When the same retention intervals were asked about in terms of the likelihood of forgetting (again between subjects), however, participants predicted that items would be less likely to be forgotten 10 minutes later than one week later, and less likely to be forgotten one week later than one year later (Figure 1, Panel A). The authors concluded that framing the prediction in terms of forgetting caused participants to consider different—and presumably more predictive—memory cues when making their JOLs than when they were framed in terms of remembering. They dubbed this idea the forgetting-notion hypothesis.

Although Koriat et al.’s (2004) studies did not directly demonstrate that the frame of the judgement changed people’s cue utilization while making their predictions, this idea is supported indirectly by other work that demonstrates that asking participants to better consider the bases for their judgements can improve the accuracy of those judgements. For example, Koriat, Lichtenstein, and Fischhoff (1980) asked participants to consider reasons why their answers to general-information questions might be correct or incorrect. When participants considered why they might be wrong, the overconfidence of their judgements was reduced, and the judgements therefore became more accurate. According to the forgetting-notion hypothesis, the forget frame presumably causes participants to think about why they might forget an item rather than why they might remember it, and the former is presumably less prone to biases (e.g., confirmation bias) in monitoring than is the latter. This idea is also indirectly supported by the main tenets of prospect theory (Kahneman & Tversky, 1979) in that the remember (gain) and forget (loss) frames presumably change what information about memory people consider when making metamemory judgements and predictions.

Finn (2008) recently expanded upon Koriat et al.’s (2004) findings and examined the effect of framing on item-by-item JOLs using a more standard paired-associates learning method. In Finn’s experiments, judgements of forgetting (after reverse scoring) demonstrated better absolute accuracy (the difference of the mean JOLs and mean recall; sometimes referred to as calibration) than did judgements of remembering on a single study–test trial (Figure 2, Panel A; but see Finn, 2008, Experiment 3, in which a difference in this measure did not obtain even though judgement magnitude differed by frame). In Finn’s Experiments 1A and 2, immediate judgements of forgetting were typically less overconfident (i.e., more accurate) than were immediate judgements of remembering. This apparently occurred...
because, in all experiments, judgements of forgetting (after reverse scoring) were significantly lower than were judgements of remembering, and actual memory performance never differed by frame. Furthermore, in Finn’s Experiments 2 and 3, participants who made judgements of forgetting chose to restudy more items than did those who made judgements of remembering. This outcome is most in line with the ideas of prospect theory, as it demonstrates that the frame of the JOLs not only changed people’s judgements about their memories, but also changed the decisions they made about their memories. Presumably, the “loss” aspect of the forget frame made people more likely to restudy items than did the remember frame (gain) because people are naturally averse to losses such as forgetting or, more generally, negative outcomes such as doing poorly on a test (Kahneman & Tversky, 1979).

Although Finn (2008) did not directly demonstrate that the frame manipulation affected cue utilization, the data nevertheless led to the conclusion that the forget frame improves the absolute accuracy of immediate JOLs by focusing participants on memory cues that better predict future learning than do the cues participants attend to when making judgements of remembering (cf. Korniat et al., 2004; see also Korniat, 1997, 2008). Taken together, these results suggest that framing JOLs in terms of forgetting rather than remembering might be a simple way of improving the absolute accuracy of JOLs and increasing the efficacy of

Figure 2. Participants’ mean recall performance and mean JOLs (judgements of learning) from previously published studies. Panel A: immediate-JOL data from Finn (2008), Experiment 1A. Panel B: delayed-JOL data from Finn (2008), Experiment 1B. Panel C: immediate-JOL data from the first trials of Kornell and Bjork (2009), Experiments 6 (remember frame) and 7 (forget frame). Panel D: immediate-JOL data for items of different font sizes from Rhodes and Castel (2008), Experiments 1 (remember frame) and 5 (forget frame). Forget-framed JOLs were reverse scored.
study. That said, the relative accuracy (i.e., how well the values of the judgements discriminated learned from unlearned items regardless of the actual magnitude of the judgement; sometimes referred to as resolution) of the JOLs in Finn’s (2008) studies never differed significantly by frame. A difference would have been expected if participants consulted more predictive cues to make one type of judgement than they consulted to make the other. In addition, relative-accuracy correlations were most often numerically lower for forget-framed JOLs than for remember-framed JOLs, which is the opposite of what would have been expected if people used more predictive cues to make judgements of forgetting than judgements of remembering.

Although Finn (2008) is the only set of published studies designed specifically to study the effects of framing on item-by-item JOLs, at least two other papers (Kornell & Bjork, 2009; Rhodes & Castel, 2008) contain comparisons of memory judgements framed in terms of either remembering or forgetting for other purposes (cf. Koriat et al., 2004). For example, Kornell and Bjork (2009) conducted 12 studies that demonstrate that people do not predict additional learning to occur with repeated study of the same materials. Although these studies were typical of other multtrial learning (and metamemory) studies in that they involved participants studying and testing over the same items over multiple trials, Kornell and Bjork’s studies differed somewhat from most otherwise similar metamemory studies in that the authors asked their participants to make learning predictions for each item for all four trials after they studied each item on the first study trial. When asked to predict future learning in this way, participants predicted that their learning would stay constant across the study trials, a pattern that Kornell and Bjork referred to as “the stability bias in human memory”. Notably, Kornell and Bjork demonstrated that this occurred even when they asked participants to make predictions of their future forgetting rather than remembering (i.e., their Experiment 7), a pattern that indirectly contradicts the general findings of Koriat et al. (2004, also Experiment 7) and the basic idea behind the forgetting-notion hypothesis (i.e., Finn, 2008; Kornell et al., 2004). To aid in the comparison of results across studies, we have included some of Kornell and Bjork’s Trial 1 results in our comparison figure (Figure 2, Panel C), as they are the most relevant to the design of the present experiments.

Examining a different question, Rhodes and Castel (2008) conducted six experiments examining the influence of perceptual factors on JOLs. Most notably, they considered how presenting memory items on a computer screen in either a relatively larger (48 pt) or a smaller (18 pt) font size would affect participants’ JOLs. Although font size did not affect memory for the items, participants judged items shown in the larger font size to be more memorable than items shown in the smaller font size (Figure 2, Panel D). Importantly, this outcome obtained regardless of whether they asked participants to make JOLs in terms of remembering (their Experiment 1) or forgetting (their Experiment 5). This presumably occurred because the font size cue was salient to the participants. Although font size did not actually affect the participants’ memory for the items, the experience of studying the items in the larger font size (which was easier to read than the smaller font size) led them to judge that they were more likely to remember the larger items than the small items. In much the same way, Finn (2008) demonstrated that the magnitude and absolute accuracy of delayed JOLs (i.e., JOLs made at a delay from the study of each item, see Rhodes & Tauber, 2011, for a recent review) were not affected by framing, presumably because participants used the same cue—covert retrieval—to inform their JOLs (cf. Nelson, Narens, & Dunlosky, 2004) regardless of the frame (Figure 2, Panel B).

More importantly for the present purposes, the effect of framing on JOL magnitude—and therefore, indirectly, absolute accuracy—has so far been inconsistent across studies. For example, judgements of remembering were significantly higher after reverse scoring than were judgements of forgetting in Finn (2008, Experiment 1A), but the opposite pattern obtained in Kornell and Bjork (2009, Experiments 6 and 7; see Figure 2, Panel C). Similarly, the mean predicted recall performance
(10-minute retention interval in Koriat et al. 2004, Experiments 4B and 7) demonstrated a pattern almost identical to that of Kornell and Bjork (2009): Forget-framed judgements were higher in magnitude (after reverse scoring) than were judgements of remembering (Figure 1, Panel B). Furthermore, people’s memory predictions in the Koriat et al. (2004) and Kornell and Bjork (2009) studies are quite similar: Participants expected to remember about 50% of the items when researchers framed the question in terms of forgetting (and reverse-scored the predictions), but only expected to remember about 30% of the items when researchers framed the question in terms of remembering. A mean of 30% is in accord with earlier proposals that participants tend to anchor immediate, remember-framed JOLs at or around the 30% point on the JOL scale, which causes the mean of such JOLs to typically fall around 30% (England & Serra, in press; Scheck, Meeter, & Nelson, 2004; Scheck & Nelson, 2005). As previously noted, JOLs in Rhodes and Castel (2008; see Figure 2, Panel D) and delayed JOLs in Finn (2008; see Figure 2, Panel B) did not differ in magnitude or absolute accuracy by frame. Further, in Finn’s (2008) Experiment 3, participants studied Spanish–English translations of varying difficulties, and differences in absolute accuracy by frame also did not obtain in this experiment, presumably because the item difficulty (i.e., memorability) was salient to participants. As such, addressing these inconsistencies across studies was one purpose of the present experiments.

The present experiments

Given the existing data and the authors’ interpretations of them, there are two possibilities for how framing affects the absolute accuracy of JOLs. First, the forgetting-notion hypothesis (Koriat et al., 2004; see also Finn, 2008) suggests that a forget frame causes participants to rely on better (i.e., more predictive) cues to make their JOLs than does a remember frame, which causes judgements of forgetting to exhibit better absolute accuracy than do judgements of remembering. Although there are some data that indirectly support this idea (Koriat et al., 2004; immediate JOLs in Finn, 2008), to date no studies have demonstrated direct evidence for this possibility (e.g., evidence indicating that participants’ utilization of any specific cues differed by frame). More problematically, there are also other data that clearly do not support the forgetting-notion hypothesis, either because no differences in absolute accuracy obtained by frame (i.e., delayed JOLs in Finn, 2008, Experiment 1b; immediate JOLs in Finn, 2008, Experiment 3; immediate JOLs in Rhodes & Castel, 2008) or because judgements of remembering demonstrated better absolute accuracy than did judgements of forgetting (i.e., Kornell & Bjork, 2009). Further, differences in cue usage by frame should also produce differences in relative accuracy, an outcome that to our knowledge has only been examined in one set of studies (Finn, 2008) and did not obtain.

In contrast, a second possibility is that framing does not actually affect the absolute accuracy of JOLs. Rather, a more parsimonious explanation might be that the absolute accuracy of JOLs is largely artefactual in nature and that apparent differences in absolute accuracy by frame stem mostly from participants’ basing their JOLs around different psychological “anchor points” on the JOL scale (cf. England & Serra, in press; Meeter & Nelson, 2003; Scheck et al., 2004; Scheck & Nelson, 2005). When people make JOLs, they presumably do so by adjusting the JOL for a given item up or down from a preset psychological anchor based on whether they have accumulated evidence that they will or will not remember that item. Given that adjustment from the anchor is often insufficient (Scheck & Nelson, 2005), JOLs will often exhibit poor absolute accuracy (unless memory performance happens to fall near the anchor). Anchoring and framing could potentially interact to affect the absolute accuracy of JOLs if (a) the different frames cause people to adopt different anchors on the JOL scale or (b) the different frames cause people to factor different information into their adjustment from an anchor or to make larger/smaller adjustments from an anchor. Either way, when framing produces a
difference in JOL magnitude (but memory is not affected), whichever type of JOL happens to have a mean judgement magnitude closer to the mean memory-performance level will appear to be the more accurate frame. Although this pattern by definition indicates greater absolute accuracy, this possibility nevertheless suggests that any apparent differences in absolute accuracy by frame should be carefully considered before a strong conclusion is made about the effect of framing on the “accuracy” of JOLs (see also Scheck & Nelson, 2005). If a forget frame actually improves the absolute accuracy of JOLs compared to a remember frame, then the accuracy advantage for forget-framed JOLs over remember-framed JOLs should persist across changes in—or manipulations of—actual memory performance.

Towards these ends, the present experiments had three main purposes. First, given the inconsistent effect of framing on JOL magnitude and absolute accuracy in previous studies, we sought to further document the effect of this manipulation on these measures. Second, we sought to compare the forgetting-notion hypothesis (Koriat et al., 2004; see also Finn, 2008) with the possibility that any effect of framing on the absolute accuracy of JOLs is largely artefactual (i.e., that the differences are largely the result of differences in anchoring—rather than accuracy—by frame). Third, we wished to further examine the idea that the frame changes participants’ cue utilization by comparing the relative accuracy of judgements of remembering and forgetting over multiple study–test trials. Our general approach for examining these issues was to have participants study paired-associates items over two study–test trials, which should produce an increase in memory-test performance across the trials. Based on the forgetting-notion hypothesis, if a forget frame truly improves absolute accuracy compared to a remember frame, then this accuracy advantage should persist across trials even as the level of memory performance changes. If, however, absolute accuracy differences by frame are largely artefactual and stem from differences in the anchoring (or adjustment) of JOLs, then any obtained pattern should switch across trials as performance increases (i.e., any group with higher magnitude JOLs should be highly overconfident on Trial 1 and only slightly underconfident on Trial 2, whereas any group with lower magnitude JOLs should be only slightly overconfident on Trial 1 and highly underconfident on Trial 2). We predict this pattern because memory judgements such as JOLs tend to exhibit a stability bias in that they are fairly “flat” across trials even though memory typically improves with repeated study (e.g., Kornell & Bjork, 2009).

Using two study–test trials will also allow us to further consider the basic tenet of the forgetting-notion hypothesis, which is that the forget frame leads people to consult more predictive cues to make their JOLs than does the remember frame. Relative accuracy typically improves across study–test trials (in particular, from Trial 1 to Trial 2) because people change the type of cue they use to inform their JOLs (cf. Koriat 1997, 2008). One specific cue that becomes available on a second trial—and that people tend to use naturally—is memory for past test (MPT) performance (Ariel & Dunlosky, 2011; England & Serra, in press; Finn & Metcalfe, 2007, 2008; King, Zechmeister, & Shaughnessy, 1980). Given that memory performance for the items on Trial 1 is highly predictive of memory performance on Trial 2, when people base their Trial 2 JOL for an item on whether or not they remembered that item on the Trial 1 test, they typically achieve very good relative accuracy on Trial 2. If people base forget-framed JOLs on better-predictive cues than they base remember-framed JOLs, the relative accuracy of forget-framed JOLs should exceed—or at least match—that of remember-framed JOLs across trials.

To preview, in Experiment 1 we examined the effect of framing on the magnitude, absolute accuracy, and relative accuracy of both immediate and delayed JOLs across two study–test trials. In Experiments 2 and 3, we examined whether participants use different anchors on the JOL scale to make immediate JOLs framed in terms of either remembering or forgetting and considered how this might lead to differences in absolute accuracy by frame. In Experiment 4, we examined whether making a cue for JOLs salient to the
participants eliminated differences in the magnitude, absolute accuracy, and relative accuracy of immediate JOLs due to framing.

EXPERIMENT 1

The purpose of Experiment 1 was (a) to collect additional data using immediate and delayed JOLs framed in terms of either remembering or forgetting and (b) to consider the magnitude and accuracy of JOLs framed in terms of either remembering or forgetting over two study–test trials. To do so, we had participants in Experiment 1 study 64 paired associates over two study–test trials. Groups made their JOLs in terms of either the likelihood of remembering or the likelihood of forgetting and either immediately after studying each item (immediate-JOL groups) or at an eight-item delay (delayed-JOL groups). Given previous results with delayed JOLs (i.e., Finn, 2008, see also Rhodes & Tauber, 2011), we did not expect either the absolute or relative accuracy of delayed JOLs to differ by frame, nor did we expect frame to interact with trial. The reason for both of these predictions is that we did not expect the cues that participants typically use to make delayed JOLs (i.e., covert retrieval during the delayed JOL) to change by frame or across trials. In contrast, the results for immediate JOLs are more difficult to predict because of the inconsistent effects of framing on the magnitude and accuracy of immediate JOLs in previously published studies. In general, however, if it is true that people make forget-framed immediate JOLs using more predictive cues than they use to make remember-framed immediate JOLs (i.e., Finn, 2008; Koriat et al., 2004), then we expected absolute accuracy to change as memory performance improved across trials (i.e., regardless of frame, JOLs will exhibit overconfidence when memory is low and underconfidence when it is high regardless of the frame).

Method

Participants, materials, and design

The participants were 192 undergraduates enrolled in General Psychology at Texas Tech University. They took part in this experiment for course credit. The study materials were the 66 paired associates from the Appendix of Serra, Dunlosky, and Hertzog (2008). The design was a 2 (framing: remember vs. forget) × 2 (type of JOL: immediate vs. delayed) between-participants factorial design.1

Procedure

The researchers randomly assigned each participant to one of the four groups when they entered the laboratory, with the restriction that an equal number of participants were eventually assigned to the four groups. The participants first read detailed instructions describing the task and the judgement that they would be making for each item. The instructions for each participant were framed in the same way as the JOLs would later be framed for that participant (variations by framing condition are in square brackets):

After you study each pair of words, you will be asked to predict how likely you think you will be to [remember/forget] the item on the upcoming test. You will make this judgement on a scale from 0 to 100. A “0” would indicate that you feel there is a 0% chance you will [remember/forget] the item (i.e., you feel you definitely WILL NOT [remember/forget] it) and a “100” would indicate that you feel there is a 100% chance you will [remember/forget] the item (i.e., you feel you definitely WILL [remember/forget] it).
The instructions did not mention the potential delay of the JOLs but, to foreshadow, typical delayed-JOL effects obtained (i.e., good absolute and relative accuracy), so we have no reason to believe that omitting this aspect of the procedure from the instructions affected the JOLs.

Personal computers running a custom computer program then presented the materials to the participants and collected participants’ responses. The computer program first randomly selected 64 of the 66 pairs to serve as each participant’s study list on both trials. The program then randomized the order of the 64 items for each participant and presented each item individually for a fixed duration of 4 s/item. In the immediate-JOL groups, participants made a JOL for each item immediately after studying it until they studied and judged all 64 items in that way. No time was inserted between the offset of one item’s JOL and the onset of the next item’s presentation. In the delayed-JOL groups, participants studied 8 items and then made the JOLs for those 8 items in turn. Given that the study and judgement making of other items occurred between the study and judgement making for any given item, an average of 31.5 s passed between the study and judgement of any individual item. The program then studied and judged additional sets of 8 items in this way until they studied and judged all 64 items. Regardless of the type of JOL, we manipulated the framing of the JOL prompt in terms of remembering or forgetting between participants as in Finn (2008). To obtain JOLs, the computer program displayed the stimulus word (the first word in each pair) and asked the participant, “How likely do you think you will be to [remember/forget] the second word of this pair approximately 5 minutes from now?” Participants made JOLs on a scale from 0 (0% chance of remembering or forgetting) to 100 (100% chance of remembering or forgetting). The computer program accepted all whole-number responses between 0 and 100. The making of the JOLs was participant paced.

After participants had studied and judged all the items, the program again randomized the order of the items and presented each stimulus word individually for paired-associate recall. The instructions encouraged participants to type the response corresponding to each stimulus word, but allowed them to omit responses if they chose. The entire procedure then repeated for a second trial during which participants studied the same 64 items in a newly randomized order (again for 4 s/item), judged them in the same way as on Trial 1, and tested over them as before.

Results

Recall

We scored the response words recalled by the participants as either correct or incorrect; we did not award partial credit because the majority of errors were omission errors. Figure 3 shows the mean recall for the immediate-JOL groups and Figure 4 shows the mean recall for the delayed-JOL groups. Recall increased from Trial 1 to Trial 2 for the immediate-JOL groups, $F(1, 94) = 472.2$, $MSE = 94.6$, $\rho < .001$, $\eta_p^2 = .8$, and for the delayed-JOL groups, $F(1, 94) = 430.9$, $MSE = 88.1$, $\rho < .001$, $\eta_p^2 = .8$. Framing did not affect recall for either the immediate- or the delayed-JOL groups, nor did frame ever interact with trial.

JOL magnitude

We reverse-scored forget-framed JOLs (i.e., 100 – JOL) as in Finn (2008) and Koriat et al. (2004) so we could directly compare them to remember-framed JOLs on the same scale. We first calculated the mean JOL for each participant on both trials and then calculated the overall means for each of the two groups. Figure 3 shows the mean JOL for the immediate-JOL groups, and Figure 4 shows the mean JOL for the delayed-JOL groups. JOLs increased from Trial 1 to Trial 2 for the immediate-JOL groups, $F(1, 94) = 10.7$, $MSE = 633.2$, $\rho = .001$, $\eta_p^2 = .1$, but delayed JOLs did not differ by frame. Framing did not interact with trial for either the immediate- or the delayed-JOL groups.
Absolute accuracy
We operationalized absolute accuracy as the mean difference score between participants' JOLs and recall. For the immediate-JOL groups, absolute accuracy changed across trials, $F(1, 94) = 237.4$, $MSE = 133.4$, $p < .001$, $\eta_p^2 = .7$, and framing affected absolute accuracy, $F(1, 94) = 7.7$, $MSE = 778.8$, $p = .007$, $\eta_p^2 = .08$. Immediate remember-framed JOLs were somewhat overconfident on Trial 1, $F(1, 47) = 5.6$, $MSE = 205.2$, $p = .022$, $\eta_p^2 = .1$, and highly underconfident on Trial 2, $F(1, 47) = 23.3$, $MSE = 325.9$, $p < .001$, $\eta_p^2 = .3$, whereas immediate forget-framed JOLs were highly overconfident on Trial 1, $F(1, 47) = 46.6$, $MSE = 186.9$, $p < .001$, $\eta_p^2 = .5$, and only somewhat underconfident on Trial 2, $F(1, 47) = 7.1$, $MSE = 194.2$, $p = .01$, $\eta_p^2 = .1$ (Figure 3). The two factors, however, did
not interact significantly. For the delayed-JOL groups, absolute accuracy changed across trials, \( F (1, 94) = 73.3, \text{MSE} = 103.4, p < .001, \eta_p^2 = .4, \) as JOLs in both groups demonstrated overconfidence on Trial 1. That said, although the overconfidence was highly significant for the forget-framed group, \( F(1, 47) = 34.2, \text{MSE} = 150.5, p < .001, \eta_p^2 = .4, \) the difference for the remember-framed group was marginal, \( F(1, 47) = 3.5, \text{MSE} = 266.6, p = .07, \eta_p^2 = .07. \) Both groups were accurate on Trial 2 (Figure 4). Neither the main effect of framing nor the interaction of framing with trial was significant for delayed JOLs.

We also considered the groups’ overall absolute accuracy by calculating a related measure, participants’ shift to underconfidence (Meeter & Nelson, 2003). This measure is basically a total bias score and can be calculated by subtracting participants’ absolute accuracy scores on Trial 2 from their absolute accuracy scores on Trial 1. These total scores did not differ by frame for either the immediate- or the delayed-JOL groups in Experiment 1, further indicating that there were no actual differences in absolute accuracy by frame in either case.

### Slopes

As in Scheck et al. (2004), we considered participants’ anchoring by calculating the changes (i.e., slopes) in participants’ JOLs and recall across trials (Table 1). Although all values were positive, slopes were greater for changes in recall than for changes in JOLs for both immediate JOLs, \( F(1, 94) = 237.4, \text{MSE} = 133.4, p < .001, \eta_p^2 = .7, \) and delayed JOLs, \( F(1, 94) = 73.3, \text{MSE} = 103.4, p < .001, \eta_p^2 = .4. \) Slopes were not affected by the frame of the JOL. No interactions were significant. These results suggest that participants’ reliance on anchors to make their JOLs impaired their ability to fully capture increases in recall across trials (cf. Scheck et al., 2004). We explore this possibility further in Experiments 2 and 3.

### Relative accuracy

We operationalized relative accuracy as a gamma correlation between participants’ JOLs and memory performance on a given trial (Table 2).² Overall, the relative accuracy for the immediate JOLs improved from Trial 1 to Trial 2, \( F(1, 90) = 10.4, \text{MSE} = 0.2, p = .002, \eta_p^2 = .1. \) More importantly, relative accuracy for immediate JOLs

---

² One potential criticism of forget-frame JOLs is that people might be confused about how to make these judgements and use the scale “backwards”. These positive relative-accuracy correlations (especially those for delayed JOLs), however, suggest that people can make forget-frame JOLs correctly, and can do with considerable accuracy in some situations.

---

### Table 1. Slopes of JOLs and recall across trials

| Experiment | Remember frame | | | Forget frame | | |
|------------|----------------|----------------|-------------|----------------|----------------|
|            | JOLs | Recall | JOLs | Recall |
|            | M   | SD   | M   | SD   | M   | SD   | M   | SD   |
| Experiment 1 | | | | | | | | |
| Immediate JOLs | +6.1 | 15.6 | +30.9 | 12.1 | +3.5 | 15.2 | +30.1 | 15.2 |
| Delayed JOLs | +16.7 | 13.0 | +27.6 | 13.1 | +14.4 | 16.3 | +28.6 | 13.4 |
| Experiment 2 | | | | | | | | |
| 2-s/item rate | +4.4 | 13.8 | +28.5 | 13.5 | +4.5 | 19.0 | +28.9 | 12.3 |
| 6-s/item rate | +8.5 | 16.3 | +35.6 | 10.0 | +10.3 | 18.0 | +31.5 | 14.9 |
| Experiment 3 | | | | | | | | |
| | +9.5 | 16.7 | +30.9 | 14.2 | +6.1 | 13.9 | +30.6 | 15.4 |
| | +12.3 | 14.5 | +29.1 | 9.3 | +10.4 | 15.1 | +28.6 | 9.2 |

Note: JOL = judgement of learning. Slopes are the mean difference between the Trial 2 value and the Trial 1 value (i.e., Trial 2 JOL – Trial 1 JOL and Trial 2 Recall – Trial 1 Recall).
was higher for remember-framed JOLs than for forget-framed JOLs, $F(1, 90) = 4.4$, $MSE = 0.2$, $p = .04$, $\eta_p^2 = .05$. The interaction between framing and trial was marginally significant, $F(1, 90) = 3.9$, $MSE = 0.2$, $p = .05$, $\eta_p^2 = .04$. Follow-up $t$ tests confirmed that framing did not affect relative accuracy on Trial 1, $t(93) = 0.2$, $p = .9$, $d = 0.03$, but that relative accuracy was higher for remember-framed JOLs than for forget-framed JOLs on Trial 2, $t(91) = 2.6$, $p = .01$, $d = 0.55$. Relative accuracy for all delayed JOLs was quite good (i.e., means around +.7) and did not differ by frame or by trial. That said, the interaction between these two factors was marginally significant for delayed JOLs, $F(1, 87) = 3.6$, $MSE = 0.09$, $p = .06$, $\eta_p^2 = .04$.

It is well established that participants’ use of memory for past test (MPT) information to make JOLs (after an initial study–test trial) plays a major role in improving the relative accuracy of immediate JOLs across trials (e.g., Ariel & Dunsloky, 2011; England & Serra, in press; Finn & Metcalfe, 2007, 2008). As such, we considered whether the forget frame reduced participants’ use of MPT information on Trial 2, which might help explain why relative accuracy was impaired on Trial 2 for forget-framed immediate JOLs. As in Finn and Metcalfe (2007, 2008), we examined participants’ use of the MPT heuristic to make their Trial 2 JOLs by calculating within-participant gamma correlations between Trial 2 JOLs and Trial 1 memory performance for both immediate-JOL groups (Table 2). As expected, the mean MPT correlation was lower for the forget-framed group than for the remember-framed group, $F(1, 91) = 6.9$, $MSE = 0.3$, $p = .01$, $\eta_p^2 = .07$. Although less relevant to the relative accuracy of delayed JOLs, for completeness we also calculated MPT for the delayed groups (Table 2). These correlations did not differ by frame for delayed JOLs, $F(1, 90) = 0.1$, $MSE = 0.1$, $p = .7$, $\eta_p^2 = .001$.

**Discussion**

The purpose of Experiment 1 was to examine the effect of framing on the magnitude and accuracy of immediate and delayed JOLs over two study–test trials. As we expected based on previous research examining the effects of framing on single-trial delayed JOLs (Finn, 2008, Experiment 1B), framing did not affect the magnitude, absolute accuracy, or relative accuracy of
delayed JOLs in our Experiment 1. It seems likely that making delayed JOLs is so likely to elicit participants’ covert retrieval of responses (e.g., Nelson et al., 2004) that the retrieval outcome might be the main cue that people use to make delayed JOLs no matter how we frame the judgements (see also Finn, 2008). Further, using the outcome of such covert retrieval attempts to inform JOLs produces extremely polarized JOLs (i.e., very high and very low JOL values). As such, it seems logical that the magnitude, absolute accuracy, and relative accuracy of the delayed JOLs did not differ by frame or across trials.

In contrast, framing clearly affected the magnitude and accuracy of immediate JOLs in our Experiment 1. First, although JOLs had equivalent relative accuracy across frames on Trial 1 (cf. Finn, 2008), forget-framed JOLs had significantly lower relative accuracy than did remember-framed JOLs on Trial 2. This latter outcome seemed to obtain because participants in the forget-framed group made less use of MPT information to inform their Trial 2 JOLs (a highly predictive cue, see Ariel & Dunlosky, 2011; England & Serra, in press; Finn & Metcalfe, 2007, 2008) than did those in the remember-framed group. These relative-accuracy results are certainly not in line with the forgetting-notion hypothesis, which suggests that a forget frame leads participants to consult more predictive cues to make forget-framed JOLs than to make remember-framed JOLs.

Second, unlike in Finn (2008), in our Experiment 1 our participants’ remember-framed JOLs were consistently lower than were our participants’ forget-framed JOLs across trials (but see Koriat et al., 2004; Kornell & Bjork, 2009). As a result, forget-framed JOLs were highly overconfident on Trial 1 when recall was low, and remember-framed JOLs were highly underconfident on Trial 2 when recall was high. In contrast, remember-framed JOLs were fairly well calibrated on Trial 1 when recall was low, and forget-framed JOLs were fairly well calibrated on Trial 2 when recall was high. Recall performance did not differ by frame (in this or any of our other experiments, so the different patterns of absolute accuracy by frame are not attributable to differential recall.

As such, it does not seem parsimonious (nor is it in line with the forgetting-notion hypothesis) to propose that remember-framed JOLs start out nearly accurate on Trial 1 and then become inaccurate on Trial 2, whereas forget-framed JOLs start out inaccurate on Trial 1 and then become nearly accurate on Trial 2. Rather, a simpler explanation might be that the absolute accuracy of JOLs is largely artefactual in nature and that these apparent differences in absolute accuracy by frame stem mostly from participants’ basing their JOLs around different “anchor points” on the JOL scale (cf. England & Serra, in press; Meeter & Nelson, 2003; Scheck et al., 2004; Scheck & Nelson, 2005) and not adjusting them much across trials (cf. Kornell & Bjork, 2009). It seems possible that our participants based their immediate remember-framed JOLs around a low point on the JOL scale (i.e., around 35%; see also England & Serra, in press; Scheck & Nelson, 2005) and their immediate forget-framed JOLs around the midpoint on the JOL scale (i.e., around 50%). We explore this possibility more directly (and in greater detail) in Experiments 2 and 3.

**EXPERIMENT 2**

In Experiments 2 and 3, we examined the possibility that the differences in immediate-JOL magnitude and absolute accuracy by frame in Experiment 1 stemmed from participants’ using different anchors on the JOL scale for the two different frames. Inspection of Figure 3 (as well as data in Koriat et al., 2004, and Kornell & Bjork, 2009) suggests that remember-framed JOLs might be anchored around the 35% point on the JOL scale whereas forget-framed JOLs might be anchored around the 50% point on the JOL scale. To examine this possibility more directly, in Experiment 2 we manipulated the per-item study time (either 2 s/item or 6 s/item) to create groups with significantly different levels of recall performance. Again, participants made immediate JOLs in terms of either the likelihood of remembering or the likelihood of forgetting. Regardless of the framing of the JOLs, we expected the absolute
accuracy of the JOLs to be dependent on the participants’ level of recall. If anchoring can largely account for the different magnitude and absolute accuracy of JOLs by frame, then manipulating the level of recall performance up or down should not affect JOL magnitude, but should affect absolute accuracy. If, however, judgements of forgetting truly demonstrate better absolute accuracy than do judgements of remembering, then judgements of forgetting should be more sensitive to the manipulated level of memory performance than are judgements of remembering and should therefore demonstrate better absolute accuracy between groups and across study–test trials as performance varies.

Method

Participants, materials, and design
The participants were 112 undergraduates enrolled in General Psychology at Texas Tech University who took part in this experiment for course credit. None had participated in the earlier experiment. We used the same list of paired associates for the study materials as that used in earlier experiments. The design was a 2 (framing: remember vs. forget) × 2 (study duration: 2 s/item vs. 6 s/item) between-participants factorial design.

Procedure

The procedure of Experiment 2 was identical to that of the two immediate-JOL groups in Experiment 1 except that in Experiment 2 we manipulated the per-item study duration between groups to produce different levels of memory performance. Whereas participants in Experiment 1 studied all items for 4 s each on both study–test trials, participants in Experiment 2 studied all items for either 2 s or 6 s each on both study–test trials.

Results

Recall

Figure 5 shows the mean recall for the four groups. Recall increased across the trials, $F(1, 108) = 665.1, \ MSE = 81.6, \ p < .001, \ \eta^2_p = .9$. More importantly, the study-time manipulation was successful at producing a difference in recall: Recall was higher for the groups that studied the items for 6 s than for the groups that studied the items for 2 s, $F(1, 108) = 10.1, \ MSE = 807.0, \ p = .002, \ \eta^2_p = .09$. Trial interacted with study time, as the increase in recall across trials was greater for the groups that studied the items for 6 s than for the groups that studied the items for 2 s,
JOL magnitude
We again reverse-scored forget-framed JOLs (i.e., 100 - JOL) so we could directly compare them to remember-framed JOLs on the same scale. Figure 5 shows the mean JOL for the four groups. JOLs increased across the trials, \( F(1, 108) = 18.8, \ MSE = 142.5, \ p < .001, \ \eta_{p}^{2} = .2 \). Forget-framed JOLs were higher than were remember-framed JOLs, \( F(1, 108) = 15.1, \ MSE = 679.7, \ p < .001, \ \eta_{p}^{2} = .12 \). As expected, the study-time manipulation did not affect JOL magnitude even though it affected the level of recall. No interactions were significant.

Absolute accuracy
Absolute accuracy changed across trials, \( F(1, 108) = 192.4, \ MSE = 170.6, \ p < .001, \ \eta_{p}^{2} = .6 \), and framing affected absolute accuracy, \( F(1, 108) = 4.9, \ MSE = 874.1, \ p = .03, \ \eta_{p}^{2} = .04 \). Remember-framed JOLs were somewhat overconfident on Trial 1 for the 2-s/item group, \( F(1, 27) = 12.4, \ MSE = 225.1, \ p = .002, \ \eta_{p}^{2} = .3 \), but were only marginally overconfident for the 6-s/item group, \( F(1, 27) = 3.6, \ MSE = 202.6, \ p = .07, \ \eta_{p}^{2} = .1 \). Remember-framed JOLs were highly underconfident on Trial 2 for both the 2-s/item group, \( F(1, 27) = 5.1, \ MSE = 272.3, \ p = .03, \ \eta_{p}^{2} = .2 \), and the 6-s/item group, \( F(1, 27) = 32.8, \ MSE = 169.2, \ p < .001, \ \eta_{p}^{2} = .5 \). In contrast, forget-framed JOLs were highly overconfident on Trial 1 for both the 2-s/item group, \( F(1, 27) = 20.4, \ MSE = 306.7, \ p < .001, \ \eta_{p}^{2} = .4 \), and the 6-s/item group, \( F(1, 27) = 10.2, \ MSE = 301.9, \ p = .004, \ \eta_{p}^{2} = .3 \), but neither group was significantly underconfident on Trial 2 (Figure 5). Although the effect of study time on absolute accuracy was only marginally significant, \( F(1, 108) = 2.8, \ MSE = 874.1, \ p = .1, \ \eta_{p}^{2} = .03 \), the differences were nevertheless in the predicted direction (i.e., greater overconfidence when performance was lower and greater underconfidence when performance was higher) in that the study-time manipulation affected recall but not the magnitude of the JOLs. No interactions were significant.

We also considered participants’ shift-to-underconfidence scores as another measure of absolute accuracy as in Experiment 1. These scores were also not affected by either framing or the study time manipulation, further indicating that there were no actual differences in absolute accuracy by frame.

Slopes
Although all values were positive (Table 1), slopes were greater for changes in recall than for changes in JOLs, \( F(1, 108) = 192.4, \ MSE = 170.6, \ p < .001, \ \eta_{p}^{2} = .6 \). Slopes were also greater for the groups that studied the items for 6 s than for the groups that studied the items for 2 s, \( F(1, 108) = 4.9, \ MSE = 277.5, \ p = .03, \ \eta_{p}^{2} = .04 \). No other main effects or interactions were significant.

Relative accuracy
Overall, the relative accuracy of the JOLs (Table 2) improved across trials, \( F(1, 100) = 14.4, \ MSE = 0.1, \ p < .001, \ \eta_{p}^{2} = .1 \). More importantly, relative accuracy was higher for remember-framed JOLs than for forget-framed JOLs, \( F(1, 100) = 6.5, \ MSE = 0.2, \ p = .01, \ \eta_{p}^{2} = .06 \). The study time manipulation did not affect relative accuracy. Follow-up t tests confirmed that relative accuracy was higher for remember-framed JOLs than for forget-framed JOLs on Trial 2, \( t(106) = 2.5, \ p = .01, \ d = 0.5 \), but did not differ on Trial 1. No interactions were significant.

The mean MPT correlation (Table 2) was lower for the forget-framed groups than for the remember-framed groups, \( F(1, 100) = 4.5, \ MSE = 0.2, \ p = .04, \ \eta_{p}^{2} = .04 \). The study-time manipulation did not affect MPT utilization, nor did study time interact with frame.

Discussion
Our study-time manipulation was successful at producing differences in recall performance but—as would be expected based on the anchoring explanation for JOLs—differential performance did not affect the magnitude of the JOLs, which was consistently higher for the forget-framed groups than for the remember-framed groups.
the main effect of this study time manipulation on absolute accuracy was only marginally significant, the mean scores nevertheless reflect exactly what we predicted: JOLs were more overconfident when recall was low than when recall was high and were more underconfident when recall was high than when recall was low. These outcomes provide further evidence that JOL magnitudes are largely dependent upon psychological anchor points on the JOL scale (England & Serra, in press; Scheck et al., 2004; Scheck & Nelson, 2005) and that remember- and forget-framed JOLs are anchored around different points on the JOL scale. We further examined these two issues in Experiment 3.

**EXPERIMENT 3**

The purpose of Experiment 3 was (a) to once again replicate the immediate-JOL findings from the earlier experiments and (b) to provide further evidence that JOLs framed in terms of either remembering or forgetting utilize different anchor points on the JOL scale. To do so, we again used a variation of the methodology of the immediate-JOL groups in Experiment 1. As in Experiment 1, we had participants in Experiment 3 study 64 paired associates over two study–test trials while making immediate JOLs in terms of either the likelihood of remembering or the likelihood of forgetting. In Experiment 3, however, participants also made second-order judgements—SOJs—to indicate their confidence in the accuracy of their JOLs. We have previously used this methodology to demonstrate that participants are more confident in the accuracy of delayed than immediate JOLs (Dunlosky, Serra, Matvey, & Rawson, 2005), that participants are more confident in JOLs for sets of items with higher than with lower relative accuracy (also Dunlosky et al., 2005), and that younger adults are more confident in the accuracy of their immediate JOLs than are older adults (Serra et al., 2008). More recently, Miller and Geraci (2011) used a similar SOJ procedure to demonstrate that low-performing students in a classroom setting are less confident in their predictions of their grades on upcoming exams than are high-performing students.

By considering the SOJ–JOL relationship for JOLs framed in terms of either remembering or forgetting, we can better identify potential anchors for the two types of JOL. As previously described, as a participant acquires more information that they will or will not remember a given item, they presumably adjust their JOL away from the anchor as they feel is appropriate (England & Serra, in press; Scheck et al., 2004; Scheck & Nelson, 2005). Therefore, participants should be more confident—and assign higher SOJs—to items for which they acquired more information (see Dunlosky et al., 2005) and should be less confident—and assign lower SOJs—to items for which they acquired little information (i.e., for items at or near the anchor). As such, plotting the SOJ–JOL relationship for JOLs framed in terms of either remembering or forgetting should reveal the potential anchor point(s) of these judgements (i.e., SOJ magnitude should be the lowest near the anchor point). Based on the results of the preceding experiments, we expect this point to be around 35% for remember-framed JOLs and around 50% for forget-framed JOLs.

**Method**

**Participants, materials, and design**

The participants were 56 undergraduates enrolled in General Psychology at Texas Tech University who took part in this experiment for course credit. None had participated in the earlier experiments. Although such “metacognition about metacognition” has not received much empirical study in the cognitive-psychology literature, it has been considered in greater depth in the social-psychology literature (albeit in a less direct way, e.g., Briñol, Petty, & Tormala, 2006; McCaslin, Petty, & Wegener, 2010). Furthermore, prominent cognitive conceptualizations of metacognition (i.e., Nelson & Narens, 1994) concede the likelihood of “meta-metacognition” and stress that such thinking is in fact no different from any other metacognition, as it is simply another form of “thought about thought”.

---

3 Although such “metacognition about metacognition” has not received much empirical study in the cognitive-psychology literature, it has been considered in greater depth in the social-psychology literature (albeit in a less direct way, e.g., Briñol, Petty, & Tormala, 2006; McCaslin, Petty, & Wegener, 2010). Furthermore, prominent cognitive conceptualizations of metacognition (i.e., Nelson & Narens, 1994) concede the likelihood of “meta-metacognition” and stress that such thinking is in fact no different from any other metacognition, as it is simply another form of “thought about thought”.
experiments. We used the same list of paired associates for the study materials that we used in earlier experiments. The type of framing (remember vs. forget) was the only between-participants independent variable.

Procedure
The procedure of Experiment 3 was identical to that of the two immediate-JOL groups in Experiment 1 except that in Experiment 3 participants also made an SOJ for each of their JOLs on both study–test trials. After participants made a JOL for an item, the computer program asked them, “You judged that you feel you have a [value]% chance of [remembering/forgetting] the current item on the test. How accurate do you think the judgement you just made is? (0 = ‘definitely NOT accurate’; 100 = ‘definitely accurate’).” The [value] was replaced with the JOL the participant assigned to the just-judged item. The [remember/forget] option matched the framing condition of the JOLs. Participants did not see any aspect of the actual item (i.e., the entire pair or the stimulus word alone) referred to at this point, to focus their judgement on the JOL, not on the item. As noted, the prompt instructed participants to make their SOJs on a 0 (the JOL is definitely not accurate) to 100 (the JOL definitely is accurate) scale and hence did not need to be reverse-scored for the forget group.

Results
Recall
Figure 6 shows the mean recall for the two groups. Recall increased across the trials, $F(1, 54) = 242.5$, $MSE = 109.2$, $p < .001$, $\eta^2_p = .8$. Framing did not affect recall, nor did frame interact with trial.

JOL magnitude
We again reverse-scored forget-framed JOLs (i.e., 100 – JOL) so we could directly compare them to remember-framed JOLs on the same scale. Figure 6 shows the mean JOL for the two groups. JOLs increased across the trials, $F(1, 54) = 14.6$, $MSE = 117.6$, $p < .001$, $\eta^2_p = .2$, and forget-framed JOLs were higher than were remember-framed JOLs, $F(1, 54) = 5.1$, $MSE = 445.3$, $p = .03$, $\eta^2_p = .09$. Framing did not interact with trial.

Absolute accuracy
Absolute accuracy of the JOLs changed across trials, $F(1, 54) = 65.2$, $MSE = 225.5$, $p < .001$, $\eta^2_p = .6$, and framing affected absolute accuracy, $F(1, 54) = 8.9$, $MSE = 782.1$, $p = .004$, $\eta^2_p = .1$
Figure 6. Remember-framed JOLs were overconfident on Trial 1, \( F(1, 27) = 5.2, \; MSE = 267.3, \; p = .03, \; \eta^2_p = .2 \), and underconfident on Trial 2, \( F(1, 27) = 6.1, \; MSE = 299.3, \; p = .02, \; \eta^2_p = .2 \), whereas forget-framed JOLs were highly overconfident on Trial 1, \( F(1, 27) = 84.3, \; MSE = 123.7, \; p < .001, \; \eta^2_p = .8 \), but accurate on Trial 2. Framing did not interact with trial. We also considered participants’ shift-to-underconfidence scores as in Experiments 1 and 2. These scores were not affected by framing.

**Slopes**

Although all values were positive (Table 1), slopes were greater for changes in recall than for changes in JOLs, \( F(1, 54) = 65.2, \; MSE = 225.5, \; p < .001, \; \eta^2_p = .6 \). No other main effects or interactions were significant.

**SOJ magnitude and distribution**

Figure 6 shows the mean SOJ for the two groups. SOJs did not differ by frame or by trial, and these two factors did not interact. More importantly, we also considered the relationship between JOL and SOJ magnitudes on the two trials (Figure 7). Although the computer program (in all experiments) permitted participants to use any value from 0 to 100 for their JOLs, participants most often responded with values that were multiples of 10 (i.e., 30; 70). They also used multiples of 5 (i.e., 25; 65) frequently, but rarely used other values (e.g., 22; 73). As such, we grouped JOLs into 11 categories to more easily consider the SOJ–JOL relationship (i.e., 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100) by rounding all JOLs to the nearest multiple of 10 (e.g., a JOL of 38 became a 40). On both trials, the SOJ–JOL relationship for the remember-framed group demonstrated a “checkmark” pattern as in earlier studies (Dunlosky et al., 2005; Serra et al., 2008). This suggests that remember-framed JOLs are psychologically anchored even lower than expected, at around the 25% point on the JOL scale. Presumably, participants adjusted their JOL value for each item away from that point as they accumulated evidence that they would or would not remember each item on the test (cf. Dunlosky et al., 2005; Scheck & Nelson, 2005). In contrast, on both trials the SOJ–JOL relationship for the forget-framed group demonstrated a more symmetric “u-shape” pattern. This suggests that forget-framed JOLs are psychologically anchored around the 65% point on the JOL scale, and participants adjusted their JOL value for each item.
away from that point as they accumulated evidence that they would or would not forget each item on the test.4

Relative accuracy
Overall, the relative accuracy of the JOLs (Table 2) improved across trials, $F(1, 52) = 23.9, MSE = 0.0, p < .001, \eta_p^2 = .3$. Relative accuracy was higher for remember-framed JOLs than for forget-framed JOLs, but the difference was not significant in this experiment (likely because the comparison was greatly underpowered: observed power was .1). Although replicating the significant effect of framing on relative accuracy was not the focus of this experiment, the pattern for the mean relative accuracy correlations for Experiment 3 is nevertheless completely in line with those in the other experiments (Table 2). Framing did not interact with trial.

The mean MPT correlation was lower for the forget-framed group than for the remember-framed group (Table 2), but the difference was not significant in this experiment (again likely because the comparison was greatly underpowered: observed power was .2). Again, replicating the significant effect of framing on MPT utilization was not the focus of this experiment.

Discussion
Framing did not affect participants’ overall confidence in their JOLs (i.e., mean SOJs). Although this suggests that participants might use the same processes to make judgements of both remembering and forgetting, further consideration of the data tells a more nuanced story. Differences in the relationship between JOL and SOJ magnitudes on the two trials (Figure 7) suggest that the anchor point for forget-framed JOLs is at the midpoint of the JOL scale (around 50–70%) whereas the anchor point for remember-framed JOLs is at the lower end of the JOL scale, around 20–40% (see also Dunlosky et al., 2005; England & Serra, in press; Scheck & Nelson, 2005).

Although participants seem to make both types of judgement using anchoring-and-adjustment processes, they probably start out at two different points on the scale and might factor evidence into their JOLs somewhat differently to adjust from there. The SOJ–JOL relationship for forget-framed JOLs seems to indicate that participants weigh information relevant to either remembering or forgetting quite evenly (i.e., the symmetrical u-shaped shape), whereas the SOJ–JOL relationship for remember-framed JOLs seems to indicate that participants mostly consider information relevant to remembering (i.e., the asymmetrical checkmark shape). The latter relationship is similar to theories of how people make delayed JOLs that suggest that participants first make a quick yes/no recall decision and only then adjust their JOLs higher as they accumulate evidence they will remember the item later (e.g., Son & Metcalfe, 2005). Such processes would probably result in the checkmark pattern seen in Figure 7 for immediate JOLs if participants first quickly made a yes/no decision: There would be little reason to consider the “no” items further, and hence they might confidently assign them low JOLs such as a 0% or 10% (with relatively high SOJs). For items that participants feel they will recall, however, more deliberate processes would then take over to better consider the likelihood of remembering each item, with higher JOLs receiving higher SOJs. This would result in a positive linear relationship between JOLs and SOJs as seen for remember-framed JOLs above 20%.

Framing aside, the present results further highlight the utility of SOJs for identifying aspects of JOL data not immediately apparent from the JOLs alone (cf. Dunlosky et al., 2005). For example, if we had assumed that higher judgements are naturally associated with higher confidence then are lower judgements, we would have been forced to conclude that participants in the forget-framed group were more confident in the accuracy of their JOLs than were participants in the remember-framed group (at least based on the mean JOLs in Figure 6). The mean SOJs, however,

---

4 Note that although the reverse scoring of the forget-framed JOLs might have produced a “flipped” SOJ–JOL pattern for these JOLs, the two halves of Figure 7 do not seem to be mirror images of each other.
demonstrate that the two groups were equally confident in the accuracy of their judgements despite the obvious differences in the magnitudes of their JOLs across the two frames. More important for the present purposes, if we had estimated the anchor of the remember-framed JOLs based solely on inspection of the data from Experiments 1 and 2, we would probably have continued to overestimate the anchor at around 35% (cf. Scheck & Nelson, 2005). Instead, the SOJ data suggest that the anchor might be even lower for these judgements (and that the anchor for forget-framed JOLs might be higher than 50%), a fact that is probably normally masked by simply using the mean JOL to estimate the anchor.

EXPERIMENT 4

So far, three data sets have demonstrated no differences in JOL magnitude or absolute accuracy by frame (although immediate JOLs in Finn’s, 2008, Experiment 3 differed in magnitude by frame, absolute accuracy did not differ in that case): the delayed JOLs in Finn (2008, Experiment 1B), the delayed JOLs in the present Experiment 1, and the immediate JOLs for large and small font sizes in Rhodes and Castel (2008, Experiments 1 and 5). So why didn’t these JOLs show magnitude or absolute-accuracy differences across the two frames? Both Finn and Rhodes and Castel suggested that differences by frame did not obtain in their respective experiments because participants had salient cues to use to make their JOLs regardless of the framing. Finn (2008) suggested that “when diagnostic information about whether the item can be retrieved from memory is available, as is the case when one is making a delayed JOL, other cues do not play as large a role in making the judgement” (p. 819). It is widely accepted that people make delayed JOLs by attempting to retrieve the target word when presented with the cue word alone during a delayed JOL (e.g., Nelson et al., 2004). This cue is presumably quite salient to participants regardless of the framing of the JOL, which helps to explain why delayed JOLs do not demonstrate differences in magnitude or absolute accuracy by frame (see also the delayed JOLs in the present Experiment 1 as well). Similarly, Rhodes and Castel’s (2008) font-size manipulation was presumably so salient that participants used it as a cue for their JOLs regardless of the framing of the JOLs and regardless of the fact that font size was not predictive of later recall.

Given these past results and suppositions, in Experiment 4 we attempted to extend the lack of a difference by frame in terms of the magnitude and accuracy of immediate JOLs when a memory cue is salient (cf. Rhodes & Castel, 2008) to a different cue: image generation. JOLs are known to be negatively correlated with the latency to generate an interactive image during the encoding of memory items (e.g., paired associates as in the present experiments; see Hertzog, Dunlosky, Robinson, & Kidder, 2003), so we predicted that participants would naturally utilize this cue regardless of frame without us explicitly prompting them to do so.

To examine this possibility, in Experiment 4 we adapted the basic procedure of the immediate-JOL groups of Experiment 1 and added image-generation instructions. Specifically, we instructed participants to try to generate an interactive image for each paired associate they studied. They clicked icons on the computer screen to indicate when they generated an image (or if they did not). They then made an immediate JOL framed in terms of either remembering or forgetting. Although we did not instruct participants to use the outcome or latency of their image generation as a cue for their JOLs, we assumed this procedure would make these cues salient and that participants would factor the information into their JOLs as in previous studies (e.g., Hertzog et al., 2003). As such, we predicted that the magnitude, absolute accuracy, and relative accuracy of the JOLs would be equal for the two groups because they would be using the same cue(s) for their JOLs regardless of the frame.

Method

Participants, materials, and design
The participants were 74 undergraduates enrolled in General Psychology at Texas Tech University who took part in this experiment for course credit.
None had participated in the earlier experiments. We used the same list of paired associates for the study materials as that used in earlier experiments. The type of framing (remember vs. forget) was the only between-participants independent variable.

**Procedure**

The procedure of Experiment 4 was identical to that of the two immediate-JOL groups in Experiment 1 except that in Experiment 4 we included image-generation instructions. Specifically, the computer program instructed participants to “try to think of an image to relate the words of each pair”. If they could come up with an image for a pair, participants were to click an “Image Formed” icon on the screen as soon as they did; if they could not, participants were to click an “Image Not Formed” icon on the screen. The computer program recorded the latencies of either selection. As such, study time in Experiment 4 was essentially participant paced. Participants engaged in image generation during the study phase of both trials in Experiment 4.

**Results**

**Recall**

Figure 8 shows the mean recall for the two groups. Recall increased across the trials, $F(1, 72) = 721.5$, $MSE = 42.6$, $p < .001$, $\eta_p^2 = .9$. Framing did not affect recall, nor did frame interact with trial.

**JOL magnitude**

We again reverse scored forget-framed JOLs (i.e., $100 - \text{JOL}$) so we could directly compare them to remember-framed JOLs on the same scale. Figure 8 shows the mean JOL for the two groups. JOLs increased across the trials, $F(1, 72) = 43.5$, $MSE = 109.5$, $p < .001$, $\eta_p^2 = .4$. As expected, the magnitude of forget-framed JOLs was not different from that of remember-framed JOLs in this experiment. Framing did not interact with trial.

**Absolute accuracy**

Absolute accuracy changed across trials, $F(1, 72) = 73.2$, $MSE = 154.5$, $p < .001$, $\eta_p^2 = .5$, but framing did not affect absolute accuracy. Both groups’ judgements were accurate on Trial 1, but both the remember-framed, $F(1, 36) = 28.8$, $MSE = 174.1$, $p < .001$, $\eta_p^2 = .4$, and the forget-framed, $F(1, 36) = 18.7$, $MSE = 192.1$, $p < .001$, $\eta_p^2 = .3$, groups’ JOLs were highly underconfident on Trial 2 (Figure 8). Framing did not interact with trial. We also considered participants’ shift-to-underconfidence scores as in the previous experiments. These scores were not affected by framing.

Figure 8. Mean JOL (judgement of learning) and recall performance across trials in Experiment 4. Forget-framed JOLs were reverse scored. Error bars are the standard error of the mean.
**Slopes**

Although all values were positive (Table 1), slopes were greater for changes in recall than for changes in JOLs, \( F(1, 72) = 73.2, \text{ MSE} = 154.5, p < .001, \eta^2_p = .5 \). No other main effects or interactions were significant.

**Image-generation data**

We examined whether the ability to generate an image for a pair of words or not was related to JOLs by calculating gamma correlations between image-generation outcome and JOL magnitude (Table 3); these did not differ significantly by frame. For items for which participants were able to successfully generate an image, we calculated gamma correlations between the latency to generate an image and JOL magnitude (Table 3). Again, there were no significant differences by frame.

**Relative accuracy**

Overall, the relative accuracy of the JOLs (Table 2) improved across trials, \( F(1, 70) = 23.7, \text{ MSE} = 0.1, p < .001, \eta^2_p = .3 \). Surprisingly—and despite the fact that participants’ utilization of the image-generation information did not differ by frame—relative accuracy was higher for remember-framed JOLs than for forget-framed JOLs on Trial 2, \( t(70) = 2.5, p = .01, d = 0.6 \), but did not differ on Trial 1. Framing did not interact with trial. The mean MPT correlation (Table 2) was lower for the forget-framed group than for the remember-framed group, \( F(1, 71) = 4.2, \text{ MSE} = 0.2, p = .05, \eta^2_p = .06 \).

**Discussion**

As expected, including a salient cue for immediate JOLs in Experiment 4 eliminated differences in the magnitude and absolute accuracy of the JOLs by frame (see also Finn, 2008; Rhodes & Castel, 2008). It seems that the magnitude and absolute accuracy of the immediate JOLs were equivalent across frames in this experiment because the two groups incorporated information about image generation (success or latency) into their JOLs in the same way regardless of the framing of the JOLs (Table 3). These outcomes mimic earlier studies utilizing salient memory cues such as font size (Rhodes & Castel, 2008) and delayed JOLs (Finn, 2008).

Surprisingly, though, relative accuracy and MPT utilization still differed by frame in Experiment 4 (Table 2) even though other cue usage did not differ by frame (see Table 3; but note that most of these other correlations were

---

**Table 3. Image-generation data in Experiment 4**

<table>
<thead>
<tr>
<th>Experiment Measure</th>
<th><strong>Remember frame</strong></th>
<th><strong>Forget frame</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Trial 1</strong></td>
<td><strong>Trial 2</strong></td>
</tr>
<tr>
<td>Percentage of items for which images were successfully generated</td>
<td>77.9</td>
<td>17.6</td>
</tr>
<tr>
<td>Correlation of JOLs to whether images were successfully generated</td>
<td>+.80</td>
<td>.36</td>
</tr>
<tr>
<td>Mean JOL for items with successful image generation</td>
<td>49.7</td>
<td>18.2</td>
</tr>
<tr>
<td>Mean JOL for items without successful image generation</td>
<td>24.1</td>
<td>16.5</td>
</tr>
<tr>
<td>Image-generation latency (in seconds)</td>
<td>5.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Correlation of JOLs to image-generation latency</td>
<td>-.05</td>
<td>.21</td>
</tr>
</tbody>
</table>

**Note:** JOL = judgement of learning.
numerically lower for the forget-framed group than for the remember-framed group as well). As in the previous experiments, these results suggest that a forget frame can impair the relative accuracy of JOLs and that a forget frame not only does not help participants to attune to predictive cues about their memory, but might actually cause them to discount or even neglect such cues altogether (as well as some nonpredictive cues). These results provide further evidence against the forgetting-notion hypothesis for the effects of a forget frame on the accuracy of JOLs.

GENERAL DISCUSSION

The purpose of the present experiments was to further examine the effects of framing JOLs in terms of either remembering or forgetting on the magnitude, absolute accuracy, relative accuracy, and cue utilization of people’s JOLs. We sought to determine whether a forget frame consistently produces greater or lesser confidence in memory than does a remember frame (cf. Finn, 2008; Korniat et al., 2004; Kornell & Bjork, 2009) and to compare the forgetting-notion hypothesis for the magnitude and absolute accuracy of JOLs (Korniat et al., 2004; see also Finn, 2008) with our own hypothesis that the effect of framing on the magnitude and absolute accuracy of JOLs is largely the result of anchoring differences by frame (cf. Scheck & Nelson, 2005). We also examined whether the presence of salient memory cues (i.e., image generation or covert retrieval for delayed JOLs) eliminates differences in the magnitude and accuracy of JOLs by frame (cf. Finn, 2008; Rhodes & Castel, 2008) and examined whether framing affects the relative accuracy and cue utilization of JOLs over multiple study–test trials.

First, the present Experiments 1 through 3 suggest that a forget frame consistently produces greater confidence in memory than does a remember frame. Although this outcome was in contrast to the results of Finn (2008), the same pattern of results was obtained by Kornell and Bjork (2009, Trial 1 predictions) and Korniat et al. (2004, 10-minute retention predictions). Kornell and Bjork (2009) themselves recognized the discrepancy in JOL magnitudes between their studies and that of Finn (2008), noting that:

[framing the predictions in terms of forgetting, in Experiment 7, appears to have globally increased participants’ predictions. . . . It is unclear why making predictions in terms of forgetting rather than learning increased [predictions of learning]; based on previous research, one might expect a question framed in terms of forgetting to make people more conservative, not more confident. (Kornell & Bjork, 2009, p. 458).]

One possible interpretation of the discrepancy across studies is that the effect of framing on the magnitude of JOLs is simply inconsistent and unpredictable. Or, fine differences in methodology between the past and present studies that we have not yet identified might have produced the differences noted. Another possibility is that personality differences existed between the samples used by Finn and those used in the studies exhibiting the opposite pattern (cf. Stankov & Lee, 2008). Regardless, the present data provide further documentation that a forget frame most often produces greater confidence in memory than does a remember frame (Korniat et al., 2004; Kornell & Bjork, 2009) and suggests that the pattern obtained by Finn (2008) might be the exception rather than the rule. Why exactly this pattern obtains is beyond the scope of the present research, but some other recent data from our lab might help to inform this issue. England and Serra (in press) recently demonstrated that remember-framed JOLs seem to be anchored around the 35% point on the JOL scale because participants expect the memory task to be difficult; when told the task will be easy, they anchor these JOLs around 50%. As such, one possibility is that the forget frame leads participants to believe that the memory task will be easier than does the remember frame. Given that this possibility seems counterintuitive (cf. Kornell & Bjork, 2009), it certainly deserves further empirical investigation.

Second—and most important for the present purposes—we considered the main tenet of the forgetting-notion hypothesis (Finn, 2008; Korniat et al., 2004), which is that forget-framed JOLs demonstrate better absolute accuracy than do
remember-framed JOLs because the former utilize more predictive cues for memory than do the latter. Instead, we proposed that any difference in absolute accuracy by frame is actually an artefactual aspect of differential anchoring by frame (cf. Scheck & Nelson, 2005), the stability bias in memory (Kornell & Bjork, 2009), and limits of the absolute-accuracy measure itself (cf. Scheck & Nelson, 2005). In support of our hypothesis, we demonstrated that people’s judgements of remembering and forgetting were both equally insensitive to natural improvements in memory performance across study–test trials (Experiments 1 through 3) and when we purposefully manipulated memory performance between groups (Experiment 2). Increases in participants’ JOLs across trials did not keep pace with increases in their memory performance in either frame (Table 1), and forget-framed JOLs were biased by a nonpredictive cue as equally as were remember-framed JOLs (Experiment 4). As such, we currently cannot endorse the forgetting-notion hypothesis that a forget frame yields JOLs that demonstrate better absolute accuracy than does a remember frame.

We suggest that future research that claims to support the improved-absolute-accuracy aspect of the forgetting-notion hypothesis must demonstrate that forget-framed JOLs are more sensitive to factors that affect memory than are remember-framed JOLs across different levels of memory performance. Based on the present results, it seems that simply calculating absolute accuracy for one level of memory performance (e.g., on a single study–test trial) is not a reliable way to demonstrate that one frame results in better absolute accuracy than does another frame. Although the anchoring explanation for JOL magnitude turns absolute accuracy into a largely artefactual measure, absolute accuracy nevertheless seems to factor into participants’ study-time allocation and restudy decisions (e.g., Finn, 2008; Metcalfe & Finn, 2008) and hence should not be completely abandoned as a topic of study. Towards this end, further examining the effects of framing on study choices might be a more fruitful avenue for research than is examining the effects of framing on the absolute accuracy of JOLs. For example, if restudy choices are largely determined by absolute accuracy, then given the pattern of absolute accuracy in the present experiments, had we asked our participants to make restudy choices, we probably would have obtained the opposite outcome to Finn (2008), with participants selecting more items for restudy in the remember-framed groups than in the forget-framed groups. If, however, restudy choices are largely determined by a gain/loss orientation that can be affected by the frame of the JOLs (cf. Kahneman & Tversky, 1979), then we might have obtained the same outcomes as Finn (2008). Again, although this question is beyond the scope of the present experiments, we do feel that it is a worthwhile question for future research.

Also related to the issues of JOL magnitude and absolute accuracy, we demonstrated that people anchor immediate JOLs at different points on the JOL scale depending on the frame of the judgements. The results of Experiments 1 through 3 suggest that people tend to anchor judgements of forgetting around the midpoint of the scale (i.e., around 50%, but perhaps as high as 70%), whereas they tend to anchor judgements of remembering around a lower point on the scale (i.e., around 30%, with a range between 20% to 40%; see also England & Serra, in press; Scheck et al., 2004; Scheck & Nelson, 2005). Furthermore, the SOJ results of Experiment 3 suggest that people might make judgements of remembering and forgetting somewhat differently. When people make judgements of forgetting, they seem to start at a psychological anchor point in the middle of the scale and adjust their JOLs up or down as they accumulate evidence that they will remember or forget the current item. This results in a U-shaped SOJ–JOL relationship. In contrast, when people make judgements of remembering, they seem to start at a psychological anchor point lower on the scale and first decide whether they will or will not remember the item. A decision that they will not remember the item results in a low JOL (i.e., 0% or 10%) with moderate confidence; a decision that they will remember the item leads them to further adjust their JOLs higher—and with greater confidence—as they accumulate evidence that they will remember the
item. This results in a checkmark-shaped SOJ–JOL relationship (see also Dunlosky et al., 2005; Serra et al., 2008).

In contrast to immediate JOLs, which were the main focus of the present experiments, it is worth noting that delayed JOLs do not seem to demonstrate framing effects (the present Experiment 1; Finn, 2008) and do not seem to be influenced by anchoring. Most likely, both of these outcomes obtain because participants use the outcome of such covert retrieval attempts to inform their JOLs, and this produces extremely polarized JOLs (i.e., very high and very low JOL values). As such, the magnitude and accuracy of these JOLs are not affected by framing, and accuracy—both absolute and relative—is quite good (cf. Nelson et al., 2004).

Third, we examined whether the presence of salient memory cues eliminated differences in the magnitude and accuracy of JOLs by frame (cf. Finn, 2008; Rhodes & Castel, 2008; see also Tauber & Rhodes, in press-a). In the present experiments, both delayed JOLs (Experiment 1) and immediate JOLs made after participants generated an interactive image for each paired associate (Experiment 4) demonstrated equal magnitude and absolute accuracy regardless of how we framed the JOLs. Although delayed JOLs exhibited equivalent relative accuracy by frame, immediate remember-framed JOLs continued to demonstrate superior relative accuracy to forget-framed JOLs even when other apparently salient cues were available that were either predictive of memory performance (i.e., MPT information) or not (i.e., cues related to image generation). As such, although the present results confirm the idea that some cues are so salient to participants that they will utilize them regardless of how the JOLs are framed, these data further disconfirm that the forget frame attunes participants to focus on and utilize more predictive cues than does the remember frame.

Finally, relative accuracy was consistently better for immediate JOLs framed in terms of remembering than for those framed in terms of forgetting. Although this difference was never significant within any single experiment on the first study–test trial (Table 2; cf. Finn, 2008), combining the Trial 1 relative-accuracy correlations for immediate JOLs across the four experiments yielded a small—and only marginally significant—advantage for remember-framed JOLs over forget-framed JOLs, $F(1, 320) = 3.0$, $MSE = 0.1$, $p = .09$, $\eta_p^2 = .009$ (Cohen’s $d = 0.16$). More notably, relative accuracy on Trial 2 was consistently better for remember-framed JOLs than for forget-framed JOLs, both within each experiment (Table 2), and when data from all four experiments were combined, $F(1, 320) = 20.0$, $MSE = 0.2$, $p < .001$, $\eta_p^2 = .06$ (Cohen’s $d = 0.49$). The Trial 2 advantage of remember-framed JOLs over forget-framed JOLs seemed to stem from participants’ greater use of past test information (MPT; see Table 2) to inform the former type of judgement (cf. Ariel & Dunlosky, 2011; England & Serra, in press; Finn & Metcalfe, 2007, 2008; King et al., 1980; Tauber & Rhodes, in press-b). MPT use was greater for the remember-framed groups than for the forget-framed groups both within each experiment (Table 2) and when data from all four experiments were combined, $F(1, 320) = 16.9$, $MSE = 0.2$, $p < .001$, $\eta_p^2 = .05$ (Cohen’s $d = 0.46$). Most important for the present purposes, this outcome contradicts the main tenet of the forgetting-notion hypothesis in that not only does a forget frame not seem to direct participants to use more predictive cues for memory than does the remember frame, but it might actually cause them to completely discount this highly predictive cue when making JOLs. Of course, MPT information is not the only cue that people consult to make JOLs across trials (Ariel & Dunlosky, 2011; England & Serra, in press; Finn & Metcalfe, 2008; Tauber & Rhodes, in press-b) so future research should examine whether people’s use of other information such as new learning or new forgetting (cf. Ariel & Dunlosky, 2011) changes based on frame as well.

These findings not only suggest that forget-framed JOLs might be a poor choice to use in applied learning situations where maximizing the relative accuracy of students’ JOLs is of tantamount importance (cf. Serra & Metcalfe, 2009), but are also important for helping us to better understand how people make JOLs in general (e.g., Koriat, 1997).
For example, current theories of how people make JOLs (e.g., Koriat, 2008) assume a bottom-up selection of cues for memory based on the utility of those cues (e.g., how predictive those cues have been for memory in the past). If people select and utilize cues for memory via bottom-up processes, however, then framing should never affect JOLs, because the cues (and their previously experienced utility) would always be the same. Instead, the present results suggest that cue selection is a more top-down process such that the framing of the prompt for the judgement can change which cues people select and how people make their judgements, regardless of the utility of the cues.

In summary, the present results add to the small—but growing—literature examining the effect of framing on metacognitive judgements. This is a promising research area that can greatly inform theories of how people make metacognitive judgements, even if the effects of different frames on such judgements do not actually yield conclusions or outcomes (i.e., changes in monitoring accuracy) that are directly applicable in the classroom. That said, future research should continue to examine the effects of framing on restudy decisions (cf. Finn, 2008), which might have important implications for applied situations.

Original manuscript received 16 June 2011
Accepted revision received 27 March 2012
First published online 30 May 2012

REFERENCES


