Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Investigating memory reactivity with a within-participant manipulation of judgments of learning: support for the cue-strengthening hypothesis

Michelle L. Rivers, Jessica L. Janes & John Dunlosky

To cite this article: Michelle L. Rivers, Jessica L. Janes & John Dunlosky (2021) Investigating memory reactivity with a within-participant manipulation of judgments of learning: support for the cue-strengthening hypothesis, Memory, 29:10, 1342-1353, DOI: 10.1080/09658211.2021.1985143

To link to this article: https://doi.org/10.1080/09658211.2021.1985143

MEMORY
2021, VOL. 29, NO. 10, 1342–1353
https://doi.org/10.1080/09658211.2021.1985143

Investigating memory reactivity with a within-participant manipulation of judgments of learning: support for the cue-strengthening hypothesis

Michelle L. Rivers, Jessica L. Janes and John Dunlosky

Department of Psychological Sciences, Kent State University, Kent, OH, USA

ABSTRACT

When learners make *judgments of learning* (JOLs) for some word pairs but not others, how and why is recall performance affected? Participants studied related and unrelated word pairs and made JOLs for a randomly selected half of the pairs. We evaluated two hypotheses. The *changed-goal hypothesis* states that making JOLs leads learners to notice differences in pair difficulty and to change their learning goal. Because JOLs are manipulated within participants, such a goal change should influence how all (judged or non-judged) pairs are processed on the list, which should lead to no JOL reactivity. The *cue-strengthening hypothesis* predicts greater positive reactivity (i.e., higher recall for judged versus non-judged pairs) for related than unrelated pairs, because making a JOL strengthens the relationship between the two words in a pair, which would be more beneficial for pairs with an a priori relationship. Across experiments, we found positive reactivity for both related and unrelated pairs (albeit to a lesser degree for the latter). We also found no evidence that learners make qualitative changes in their reported strategy use when judging pairs. Making JOLs for some pairs on a list influenced memory performance and the pattern of reactivity provided support for the cue-strengthening hypothesis.

ARTICLE HISTORY
Received 5 March 2021
Accepted 20 September 2021

KEYWORDS
Judgments of learning; measurement reactivity; metacognition; metamemory; monitoring

Hundreds of studies across fifty years of research have used *judgments of learning* (JOLs) to investigate learners' ability to monitor their own learning. This research has provided insight about human metacognition, at times revealing surprising dissociations between how we think we learn and how we actually learn (e.g., Bjork et al., 2013). The majority of prior research has used JOLs to investigate *how* people monitor their learning after study (for a review, see Rhodes, 2016) and only recently has focused on the potential for these immediate judgments to have reactive effects on subsequent memory (e.g., Janes et al., 2018; Mitchum et al., 2016; Myers et al., 2020; Soderstrom et al., 2015; Tauber & Witherby, 2019; Witherby & Tauber, 2017). Accordingly, the current investigation aimed to address the following questions: When learners make JOLs for some pairs on a list but not for others, is recall performance affected? That is, are JOLs reactive when their presence (or absence) is manipulated within each participant? And, assuming reactivity does occur, does asking learners to make immediate JOLs influence the ongoing learning process, such as by changing the strategies used to encode material or by enhancing people's processing of judgment-relevant cues? For the remainder of the Introduction, we first discuss recent research demonstrating that making immediate JOLs (when manipulated *between* participants) can have a reactive effect on memory. We then describe two hypotheses for explaining these reactive effects and introduce our approach to evaluate them using a within-participant manipulation of JOLs.

Making JOLs requires learners to predict the likelihood of future recall during the acquisition of information, typically on an item-by-item basis. Because learners may not monitor their learning using the same inferential processes that JOLs require (e.g., Koriat, 1997), making them may influence how learners process the to-be-learned items and ultimately affect performance on a subsequent memory test (Ericsson & Simon, 1980). The reactive effects of making delayed, cue-only JOLs (i.e., judgments made a short time after initial study) have been extensively explored and debated (e.g., Akdoğan et al., 2016; Jönsson et al., 2012; Nelson & Dunlosky, 1992; Spellman & Bjork, 1992; Tauber et al., 2015; for a review, see Rhodes & Tauber, 2011). However, investigations into the reactive effects of immediate JOLs (i.e., where a judgment is made immediately after studying each item) are more recent, with the majority of this research using related and unrelated cue-target word pairs and measuring memory performance with cued-recall tests.

CONTACT Michelle L. Rivers mlrivers3@gmail.com Department of Psychological Sciences, Kent State University, 800 E. Summit St., Kent, OH 44242, USA

A handful of recent studies have demonstrated the reactive effects of making immediate JOLs on memory (Janes et al., 2018; Mitchum et al., 2016; Myers et al., 2020; Soderstrom et al., 2015) using similar methodology. Participants studied a list of cue-target word pairs, half of which were comprised of two related words (e.g., *feathers – bird*) and half of which were comprised of two unrelated words (e.g., *mask – bread*). After studying each pair, some participants made JOLs immediately after studying each pair (i.e., estimate the likelihood, on a 0–100% scale, of successfully recalling each target when presented with the cue), and other participants did not. Cued-recall performance was then compared between the JOL and no-JOL groups. In most of these studies, related pairs showed *positive reactivity*, with recall for related pairs being higher for the JOL group than the no-JOL group (but see Mitchum et al., 2016, who found no reactivity for related pairs in some experiments). In contrast, unrelated pairs showed no reactivity (statistically equivalent recall for the two groups; Janes et al., 2018; Myers et al., 2020; Soderstrom et al., 2015) or *negative reactivity* (lower recall for the JOL than no-JOL group; Mitchum et al., 2016).

What causes such reactivity? One hypothesis for explaining positive reactivity effects for related pairs incorporates ideas from both Koriat's (1997) cue-utilization framework for JOLs and the transfer-appropriate processing hypothesis for memory (e.g., Morris et al., 1977; see also deWinstanley et al., 1996), which we refer to as the *cue-strengthening hypothesis* (Soderstrom et al., 2015). This hypothesis states that "when a learner is required to make a JOL, the act of doing so can result in the strengthening of the cues or information used as the basis of arriving at such a judgment" (Soderstrom et al., 2015, p. 554). More specifically, because learners base JOLs on associative relatedness (e.g., Dunlosky & Matvey, 2001; Koriat, 1997; Mueller et al., 2013), JOLs strengthen the relationship between the two words, which is beneficial for a later cued-recall test. Thus, the cue-strengthening hypothesis predicts greater positive reactivity for related pairs than unrelated pairs because it is presumably easier to form a meaningful relationship between pairs that already have an inherent association. These outcomes can also be explained by the *changed-goal hypothesis* (Mitchum et al., 2016). According to this hypothesis, the requirement to make JOLs leads participants to consider that some pairs will be remembered and some will not. When studying a list comprised of related and unrelated word pairs, this translates into participants noticing that related pairs will typically be easier to learn and unrelated pairs will be more difficult. In doing so, learners change their goal from attempting to learn all pairs on the list toward learning just the easier, related pairs at the expense of learning the more difficult, unrelated pairs.

To provide a unique, a priori evaluation of these hypotheses, in the current study we manipulated JOLs within participants and within lists. In particular, participants studied a list comprised of related and unrelated pairs, and most important, they made JOLs for half of the pairs and did not make JOLs for the other half. Regardless of whether JOLs are manipulated within each participant or between participants, the cue-strengthening hypothesis predicts positive reactivity for related pairs. JOL reactivity should be smaller for unrelated pairs, given that it is more difficult to associate words that are unrelated and that making a low JOL for an unrelated pair merely involves noticing that the two words are unrelated. By contrast, in a within-participant design, the changed-goal hypothesis predicts no JOL reactivity for either related or unrelated pairs (but just a main effect for relatedness, with higher performance for related than unrelated pairs). Making JOLs for some pairs is expected to change participants' goal to focusing study on the easier, related items at the expense of learning the more difficult, unrelated pairs. That is, making a JOL draws attention to the differential ease-of-learning for related versus unrelated pairs, so this goal change is expected to be list wide, regardless of whether a JOL is made for a particular pair.

Most important, no prior research has evaluated JOL reactivity using a 2 (related vs. unrelated) × 2 (JOL vs. no JOL) within-participant design (but see Myers et al., 2020, who manipulated JOLs within participant but across two lists). Exploring reactivity using this design will provide unique evidence for testing and developing theory of JOL reactivity effects. Nonetheless, some prior research has found recall advantages for judged items when participants judge some items but not others (Arbuckle & Cuddy, 1969; Yang et al., 2015; Zechmeister & Shaughnessy, 1980), whereas other research has found no differences in recall performance for judged and non-judged to-be-learned material (Benjamin et al., 1998; Kelemen & Weaver, 1997). However, the outcomes from this prior literature are limited for evaluating hypotheses of reactivity, because evaluating judgment reactivity was not the primary goal in these studies, so certain confounds (e.g., time on task was not held constant between JOL and no-JOL conditions) may explain the mixed findings.

Across the three experiments in the current investigation, participants studied a list of related and unrelated word pairs, made immediate JOLs for half of the pairs of each type, and then completed a cued-recall test. To foreshadow, in Experiment 1, we found positive reactivity for related word pairs (and a smaller positive reactive effect for unrelated pairs), and we replicated these findings in Experiment 2. In Experiment 3, we evaluated whether learners make qualitative changes in their strategy use when judging pairs, which provides a proximal explanation for positive reactivity. Given that all experiments included a replication for the novel manipulation of JOLs using a within-participant design, we were also able to provide a high-powered estimate of the size of the reactivity effects across experiments in the General Discussion.

Experiment 1

The main goal of Experiment 1 was to explore JOL reactivity using a within-participant design, but we also included a between-participants manipulation of JOLs in attempts to replicate prior research (as per Janes et al., 2018; Mitchum et al., 2016; Soderstrom et al., 2015). One reason for including the between-participants replication concerns interpretation of the within-participant outcomes. In particular, if the pattern of within-participant outcomes is different than that of prior research, then one possible concern is that our sample is not representative of those used in prior studies. However, if the outcomes from the between-participants manipulation (collected at the same time as the within-participant data using random assignment) do replicate prior ones, then this concern would be minimised.

Another reason for including the between-participants manipulation concerns a potential outcome of the within-participant manipulation of JOLs. Although the cue-strengthening hypothesis predicts positive reactivity for related pairs, another possibility is that the requirement to make JOLs for some pairs on the list and not others will result in participants attending to relatedness of word pairs *in general*, because participants would not know before study which pairs would be judged versus not judged. If so, we would expect reduced or eliminated reactivity when using a within-participant manipulation.

We randomly assigned participants to three groups: One group made JOLs during presentation of each pair, one group did not make any JOLs, and the third group made JOLs for a randomly selected half of the pairs. For the between-participants manipulation, we expected to replicated prior outcomes, with positive JOL reactivity occurring for related pairs and either negative or no reactivity for unrelated pairs. As developed above, the predictions for the within-participant manipulation of JOLs from the hypotheses are as follows. For the cue-strengthening hypothesis, positive reactivity is predicted for related pairs, whereas small-to-no reactivity is expected for unrelated pairs (e.g., Janes et al., 2018; Mitchum et al., 2016; Soderstrom et al., 2015). For the changed-goal hypothesis, no reactivity is expected because making JOLs presumably shifts participants' study to focus on the easier-to-learn related pairs at the expense of the more difficult, unrelated pairs.

Method

All experiments reported here received approval from Kent State University's Institutional Review Board. All participants provided informed consent prior to participation.

Design

Experiment 1 used a 2×3 mixed design in which cue-target association (related, unrelated) was manipulated within participant and judgment group (JOL, no-JOL, mixed) was manipulated between participants. The mixed group made JOLs for a randomly selected half of the studied pairs. Note that even though other variables were manipulated within-participant across these experiments (e.g., cue-target association), when we use the term *mixed*, we are referring to JOLs being elicited for some pairs and not others.

Participants

We aimed for 30 participants in each group. Ninety-three Kent State University undergraduates ($n = 30$ in the JOL group, $n = 30$ in no-JOL group, and $n = 32$ in the mixed group) participated for partial course credit in their Psychology course. One additional participant in the mixed group was removed from analyses because they did not make JOLs when prompted. We conducted sensitivity analyses using G*Power (Faul et al., 2007) for an ANOVA with power set at .80 and $\alpha = .05$. These analyses indicated that the group sizes provided sufficient sensitivity to detect an effect of $\eta_p^2 = .06$ for the within-participant manipulation of JOLs (i.e., the mixed group, $n = 31$) and an effect of $\eta_p^2 = .03$ for the between-participants manipulation of JOLs (i.e., the JOL and no-JOL groups, $n = 60$).

Materials

Materials were 60 cue-target word pairs taken from Soderstrom et al. (2015, Experiment 1b). Of these pairs, half were strongly related (mean forward associative strength = .57; e.g., *feathers – bird*) and half were unrelated (e.g., *mask – bread*) according to the free association norms by Nelson et al. (2004). We used these pairs so as to provide a direct replication of prior research; however, we acknowledge that these materials do not rule out potential item effects because the targets are not the same for related and unrelated pairs. Note that when the same targets were randomly assigned to related or unrelated cue words (as in Myers et al., 2020), a similar pattern of reactivity was found when JOLs were manipulated between participants – positive reactivity for related pairs and no reactivity for unrelated pairs.

Procedure

Participants were run in small groups of up to six. Each participant was run in an individual cubicle with a computer programmed with LiveCode. The experimental procedure consisted of three phases: Study (with or without JOLs), distractor, and test. Before study, participants were told to expect a cued-recall test on each pair ("Later in the experiment, you will be asked to recall the second word when provided with the first.") During study, each participant was exposed to the 60 word pairs in a random order, presented individually for 8 s each. Participants in the JOL group were prompted to make a JOL (i.e., "On a scale of 0–100%, please estimate the likelihood that you will be able to successfully recall this pair on a later test.") halfway through the exposure duration (i.e., after 4 s), whereas participants in the no-JOL group made no JOLs during the 8-s exposure duration. If at any point participants did not make a JOL for a given pair, they were briefly reminded (for 4 s) by the computer programme to type in their judgment within the allotted time. Participants in the mixed group were prompted to make JOLs for a randomly selected half of the pairs presented, with the restriction that half of the JOLs were made for related pairs, and half for unrelated pairs. Following study, participants completed a 3-min distractor task (paper-and-pencil arithmetic problems). Finally, participants were given a self-paced cued-recall test, with the order of cue words randomised for each participant. The full procedure took approximately 20 min.

Results

In Experiment 1, our primary goal was to investigate JOL reactivity using a within- participant manipulation of JOLs. First, we present analysis relevant to our replication of prior research that used a between-participants manipulation of JOLs, which mitigated any concerns about having a representative sample compared to prior research. Next, we report our novel findings of within-participant JOL reactivity.

Responses were marked as correct if the first three letters matched the target.[1] Effect sizes were calculated using formulas from Lakens (2013). All *t*-tests are two-tailed.

Between-participants manipulation of JOLs

A 2 (cue-target association: related vs. unrelated) × 2 (judgment group: judgment vs. no judgment) mixed ANOVA was conducted on memory performance (top panel of Figure 1), calculated as the mean proportion correct on the cued-recall test. The ANOVA revealed a significant main effect of cue-target association with recall being higher for related than unrelated pairs ($M = .77$ vs. $M = .30$), $F(1, 58) = 540.83$, $p < .001$, $\eta_p^2 = .90$, but no significant main effect of judgment condition, $F(1, 58) = .01$, $p = .95$, $\eta_p^2 = .00$. The interaction between cue-target association and judgment condition was significant, $F(1, 58) = 19.80$, $p < .001$, $\eta_p^2 = .25$. Follow-up tests revealed that recall was higher for related pairs that were judged versus not judged, $t(58) = 2.48$, $p = .016$, 95% CI [.02, .15], Hedges' $g_s = 0.63$, and was lower for unrelated pairs that were judged versus not judged; $t(58) = 1.68$, $p = .098$, 95% CI [−.20, .02], $g_s = 0.43$. The latter negative reactivity for unrelated pairs using a between-participants manipulation is central for constraining theories of JOL reactivity; however, given that our focus across all experiments is on the within-participant manipulation of JOLs (vs. no JOLs), we do not consider this effect again until the General Discussion.

Within-participant manipulation of JOLs (mixed group)

A 2 (cue-target association) × 2 (judgment condition) repeated-measures ANOVA was conducted on memory performance (bottom panel of Figure 1), which revealed that recall was higher for related than unrelated pairs ($M = .75$ vs. $M = .23$), $F(1, 31) = 322.51$, $p < .001$, $\eta_p^2 = .91$. Recall was also higher for pairs that were judged versus not judged ($M = .53$ vs. $M = .46$), $F(1, 31) = 17.99$, $p < .001$, $\eta_p^2 = .37$. The interaction between cue-target association and judgment condition was not significant, $F(1, 31) = 2.56$, $p = .12$, $\eta_p^2 = .08$. Despite the non-significant interaction, given that the predictions from the hypotheses pertained to the reactivity effects separately for related pairs and unrelated pairs, we also conducted follow-up comparisons for the two kinds of pairs. Follow-up *t*-tests revealed that recall was higher for related pairs that were judged versus not judged, $t(31) = 4.50$, $p < .001$, 95% CI [.06, .15], Hedges' $g_{av} = 0.61$. Recall did not significantly differ for unrelated pairs that were judged versus not judged; $t(31) = 1.56$, $p = .13$, 95% CI [−.01, .10], $g_{av} = 0.23$.

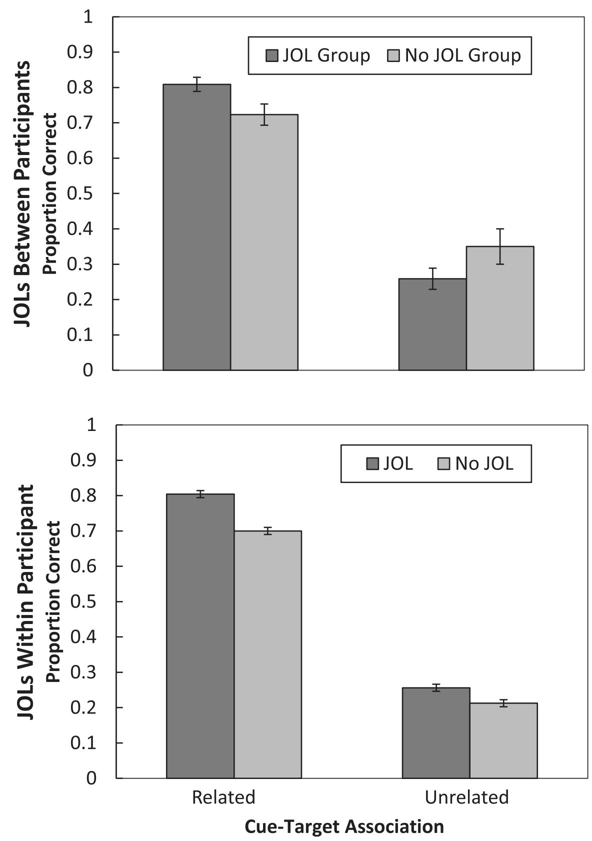

Figure 1. Recall performance for the JOL, No-JOL, and mixed groups in Experiment 1.

Note: JOL = judgment of learning. For the between participants manipulation of JOLs, error bars reflect standard error of the mean. For the within participant manipulation of JOLs, error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions (for details, see Loftus & Masson, 1994).

Between-participants versus within-participant reactivity effects

Because we manipulated judgments both between and within participants in the same experiment, we were able to compare their effects in a 2 (judgment manipulation: between participants or within participant) × 2 (judgment condition: judgment vs. no judgment) × 2 (cue-target association: related vs. unrelated) mixed ANOVA. However, we could not treat the judgment manipulation as a repeated measure. Instead, we treated the judgment manipulation as a between-participants variable, which may have reduced our power (by ignoring the relatedness of scores in the within-participant group) and resulted in a negatively biased *F*-statistic (Erlebacher, 1977). Despite this conservative estimate, we still obtained a marginally significant 3-way interaction, $F(1, 120) = 3.37$, $p = .069$. Follow-up tests revealed a significant interaction between judgment condition and cue-target association for participants with a between-participants manipulation of JOLs, but not for participants who received a within-participant manipulation of JOLs (see above).

Discussion

We replicated results from prior research using a between-participants manipulation of JOLs; namely, positive reactivity occurred for related pairs. Most important, with a within-participant manipulation of JOLs, we found positive reactivity for related pairs, suggesting that JOLs have a proximal effect on recall performance. That is, the requirement to make JOLs for some pairs on the list and not others does not necessarily lead to participants attending to relatedness of word pairs *in general*, rather, the effect is limited to the pairs being judged. These outcomes provide more competitive support for the cue-strengthening hypothesis than for the changed-goal hypothesis.

Experiment 2

Given the novel finding of positive reactivity for the within-participant manipulation of JOLs, we attempted a high-powered replication of the mixed group in Experiment 2. We also further evaluated the cue-strengthening hypothesis by presenting related and unrelated word pairs in a blocked order for two additional groups. For the latter, we used the same materials as Experiment 1, but participants either studied all thirty related pairs first followed by the thirty unrelated pairs, or vice-versa.

For the cue-strengthening hypothesis, JOLs presumably have a direct effect on processing at the level of each pair. Accordingly, the prediction is straightforward: Regardless of whether pair type (related vs. unrelated) is mixed within a list or blocked, positive reactivity is expected for related pairs. The prediction from the changed-goal hypothesis is not as straightforward, although according to this hypothesis, mixing related and unrelated pairs within a list is critical for obtaining JOL reactivity (e.g., Mitchum et al., 2016) because making JOLs presumably focuses participants on this difference between related and unrelated items. Thus, one prediction from this hypothesis that reactivity will be minimal using the blocked list design. To evaluate these possibilities in Experiment 2, some participants studied either a list wherein related and unrelated pairs were intermixed (to replicate the novel outcomes from Experiment 1), whereas other participants either studied all thirty related pairs first followed by the thirty unrelated pairs (i.e., blocked list design) or vice-versa.

In research by Janes et al. (2018; Experiment 2), participants learned either all related pairs in a list or all unrelated pairs in a list. Half of the participants made JOLs for each pair, and half did not. In support of the changed-goal hypothesis, no JOL reactivity was observed for either related or unrelated pairs. However, Witherby and Tauber (2017) reported JOL reactivity for a list of pairs that were relatively homogeneous with respect to learning difficulty (i.e., all pairs were moderately related), which cannot be explained by the changed-goal hypothesis. The pattern of reactivity for a blocked list when learners make JOLs for some pairs and not others remains an open issue.

Method

Participants and procedure

Participants were randomly assigned to one of three groups. Eighty-three participants followed an identical procedure to the within-participant JOL group of Experiment 1 (*mixed* group). An additional 163 participants were assigned to the *blocked* groups – 82 participants studied all 30 related pairs first followed by the 30 unrelated pairs (*related-first* group), and 81 participants studied all the unrelated pairs first followed by the related pairs (*unrelated-first* group). One participant was excluded from the unrelated-first group due to a computer malfunction. We conducted a sensitivity analysis using G*Power for an ANOVA with power set at .80 and $\alpha = .05$, which indicated that these group sizes ($n \approx 80$) provided sufficient sensitivity to detect an effect of $\eta_p^2 = .02$.

All participants made JOLs for a randomly selected half of the pairs. The distractor task and cued-recall test procedures were identical to the mixed group of Experiment 1.

Results

Mixed group

Recall performance for the mixed group is presented in the top panel of Figure 2. A 2 (cue-target association) × 2 (judgment condition) repeated-measures ANOVA revealed a significant main effect of cue-target association with recall being higher for related than unrelated pairs ($M = .80$ vs. $M = .26$), $F(1, 82) = 1439.55$, $p < .001$, $\eta_p^2 = .95$. The main effect of judgment condition was also significant, with recall being higher for pairs that were judged versus not judged ($M = .55$ vs. $M = .51$), $F(1,82) = 15.29$, $p < .001$, $\eta_p^2 = .16$. The interaction between cue-target association and judgment condition was not significant, $F(1, 82) = 0.08$, $p = .77$, $\eta_p^2 = 001$. As in Experiment 1, however, we

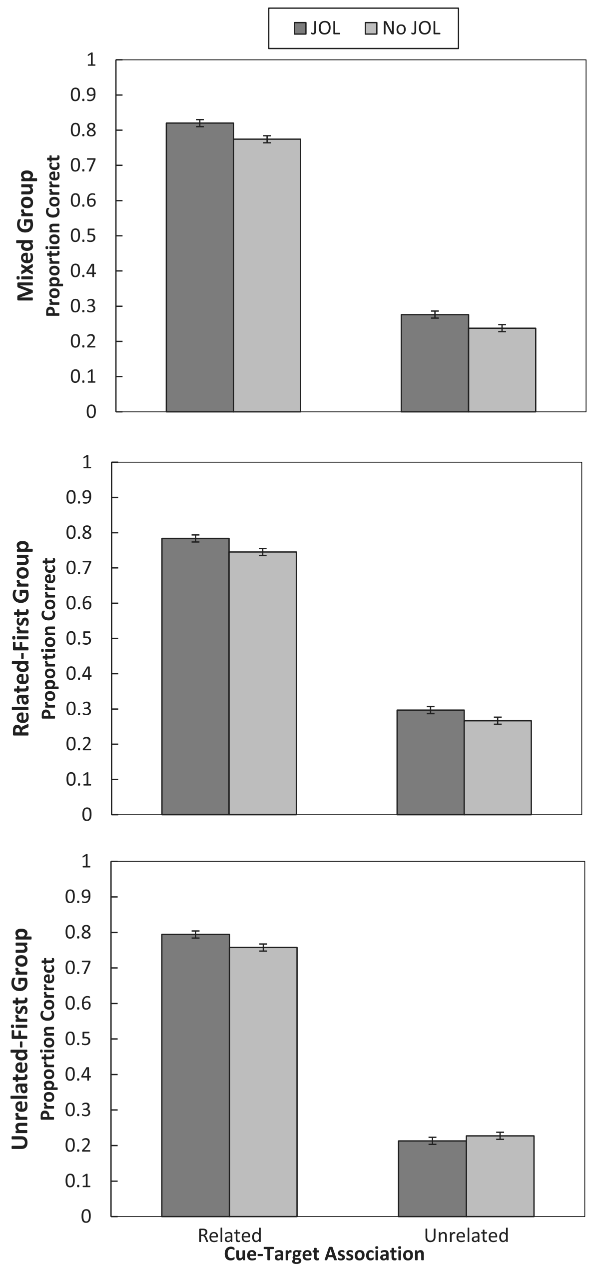

Figure 2. Recall performance for the mixed, related-first, and unrelated-first groups in Experiment 2.

Note: Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions. JOL = judgment of learning.

conducted separate comparisons for related and unrelated pairs. Follow-up *t*-tests revealed that recall was higher for related pairs that were judged versus not judged, $t(82) = 3.07$, $p = .003$, 95% CI [.02, .07], $g_s = 0.35$, and - to a lesser extent - for unrelated pairs that were judged versus not judged; $t(82) = 2.13$, $p = .036$, 95% CI [.002, .07], $g_s = 0.23$.

Blocked groups

Recall performance for the related-first group is presented in the middle panel of Figure 2, and for the unrelated-first group in the bottom panel of Figure 2. We conducted a 2 (group: related-first, unrelated first) × 2 (cue-target association) × 2 (judgment condition) mixed ANOVA on cued-recall performance. The 3-way interaction was not significant, $F(1, 160) = 1.74$, $p = .19$, $\eta_p^2 = .01$, nor was the 2 (group) × 2 (judgment condition) interaction, $F(1, 160) = 1.90$, $p = .17$, $\eta_p^2 = .01$. The 2 (group) × 2 (cue-target association) interaction was significant, $F(1, 160) = 8.00$, $p = .005$, $\eta_p^2 = .05$. In the related-first group, participants recalled more related pairs ($M = .76$, $SD = .13$) than unrelated pairs ($M = .28$, $SD = .15$), $t(81) = 31.36$, $p = <.001$, 95% CI [.45, .51], $g_{av} = 3.41$. In the unrelated-first group, participants also recalled more related pairs ($M = .78$, $SD = .15$) than unrelated pairs ($M = .22$, $SD = .16$), $t(79) = 26.94$, $p = <.001$, 95% CI [.51, .60], $g_{av} = 3.55$.

The 2 (cue-target association) × 2 (judgment condition) interaction was significant, $F(1, 160) = 3.34$, $p = .07$, $\eta_p^2 = .02$. For related pairs, recall was higher for pairs that were judged versus not judged, $t(161) = 3.30$, $p = .001$, 95% CI [.02, .06], $g_{av} = 0.24$. For unrelated pairs, the recall difference for the JOL and no-JOL conditions was not significant, $t(161) = 0.69$, $p = .49$, 95% CI [−.02, .03], $g_{av} = 0.05$.

Discussion

Using the within-participant design wherein JOLs (vs. no JOLs) were mixed across the list, we replicated results of Experiment 1 and again found positive reactivity. For both the related-first and unrelated-first groups, we found significant positive reactivity for related word pairs, but no significant reactivity for unrelated pairs. As with the outcomes presented in Experiment 1, the present outcomes cannot be easily explained by the changed-goal hypothesis and provide more competitive support for the cue-strengthening hypothesis.

Experiment 3

Results from the first two experiments provide support for the cue-strengthening hypothesis, which leads to the question: What is the *proximal* mechanism that results in enhanced performance for judged pairs? The main goal of Experiment 3 was to investigate one potential mechanism – that the effect of making a JOL increases the likelihood of generating an effective mediator for judged pairs. Specifically, making JOLs may induce learners to use more effective strategies during study. The positive reactivity observed when participants make JOLs may be due to switching from relatively ineffective strategies, such as rote repetition, to more effective strategies, such as imagery or sentence generation.

Some preliminary evidence that the requirement to judge learning leads to the use of more effective strategies was reported by Pressley et al. (1984) and Sahakyan et al. (2004), who found that participants who made judgments about memory were more likely to use effective strategies on a future learning task compared to participants who did not make judgments. However, both studies had participants make *global* judgments (i.e., predict the percentage of studied pairs they would be able to recall on a final test), which may operate differently from the pair-by-pair JOLs used in the present experiments.

More relevant, Mitchum et al. (2016, Experiment 1) examined whether participants use different strategies when they make item-level JOLs versus when they do not. Participants completed a post-experiment questionnaire, in which they were to rate (on a 0–10 scale) the extent to which they used various encoding strategies (e.g., rote rehearsal, interactive imagery, sentence generation) during study. No differences occurred in participants' self-reported strategy use between the JOL and no-JOL groups. Note, however, such general reports do not allow analyses at the pair level, which are relevant for potentially discovering the nature of reactivity. For instance, although participants reported using the same strategies whether they made JOLs or not (Mitchum et al., 2016), making JOLs may have shifted the pattern of strategy use across related and unrelated pairs.

To further evaluate the contribution of using effective strategies for producing JOL reactivity, we collected item-by-item reports of strategy use, which allowed us to examine whether participants use qualitatively different strategies making JOLs, and whether strategy use varied by cue-target association. We also used the mixed group (i.e., participants made JOLs for a randomly selected half of the pairs) from Experiments 1 and 2, so as to replicate key outcomes and further evaluate the cue-strengthening hypothesis.

Method

Participants and procedure

Our target sample size was 60 participants. Sixty-four undergraduates participated.

A sensitivity analysis conducted using G*Power (Faul et al., 2007) with power set at .80 and two-tailed $\alpha = .05$ indicated that this sample size provided sufficient sensitivity to detect an effect of Cohen's $d = 0.36$, which was relevant to our primary analysis comparing strategy reports for pairs that were judged versus not judged.

We replicated the design of the mixed group from Experiments 1 and 2 – all participants made JOLs for half of the word pairs. The procedure was identical up until the cued-recall test. Following the distractor task, participants made strategy reports. We chose to use delayed reports because having participants make them concurrently (i.e., immediately after studying each pair) may in itself be reactive (Dunlosky & Hertzog, 2001). To obtain reports, participants first received information about three common strategies used for paired-associate learning: sentence generation, rote repetition, and interactive imagery (Dunlosky & Hertzog, 2001; Richardson, 1998). For each strategy, participants were provided with a one-sentence description (e.g., "when you use *sentence generation*, you try to link the two words together by completing a sentence that includes both words") and an example (e.g., "If shown the pair *clown* – *paper*, you may have generated a sentence to help yourself remember the pair, such as 'The *clown* always wanted to buy a *paper* hat'"). Participants were also told that many other strategies may have been used to learn the pairs, and that they may have used multiple strategies during learning. Participants were also given the option to report *other strategy*, *no strategy* or *don't remember*. After attempting cued recall for each pair, participants reported the strategy they had used during learning and were given the option to provide a brief description (e.g., number of times they rehearsed a given pair, specific sentence or image they generated, or a short description of another strategy as per Morehead et al., 2018). Specifically, participants were told:

> After attempting to recall the response to a given pair, you will then be asked to report the strategy that you used to learn the pair. If you don't remember the strategy that you had used, that's fine, just click on "don't remember." If you think you used "imagery" (or some other strategy) but don't remember the specific image, that's fine too: just click on "imagery" and then type "don't remember the image" in the response box.

Even if participants did not recall a particular target, they were still required to make a strategy report. Neither the cue nor the target of the pair was available during these reports. Strategy reports and descriptions were self-paced.

Results

Recall performance

Recall performance is presented in Figure 3. A 2 (cue-target association) × 2 (judgment condition) repeated-measures ANOVA revealed recall was higher for related than unrelated pairs ($M = .81$ vs. $M = .22$), $F(1, 63) = 993.51$, $p < .001$, $\eta_p^2 = .94$. The main effect of judgment condition was significant, with higher recall for judged versus non-judged pairs ($M = .53$ vs. $M = .50$), $F(1, 63) = 5.67$, $p = .02$, $\eta_p^2 = .08$.

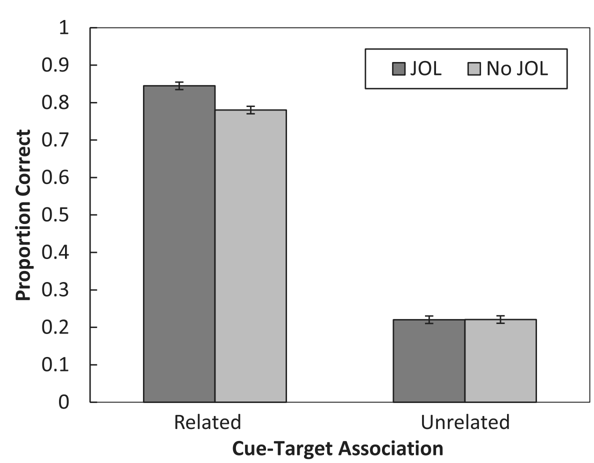

Figure 3. Recall performance in Experiment 3.

Note: JOL = judgment of learning. Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions.

The cue-target association by judgment condition interaction was significant, $F(1, 63) = 5.82$, $p = .019$, $\eta_p^2 = .09$. For related pairs, recall was higher for pairs that were judged versus not judged, $t(63) = 3.48$, $p = .001$, 95% CI [.03, .10], $g_{av} = 0.45$. For unrelated pairs, the recall difference for the JOL and no-JOL conditions was not significant, $t(63) = 0.03$, $p = .98$, 95% CI [−.04, .04], $g_{av} < 0.01$.

Strategy reports

To answer our primary question regarding strategy use, pairs in which participants reported using imagery or sentence generation were grouped into a single category, which we refer to as normatively effective strategies (as per Dunlosky & Hertzog, 2001). Reports of using rote repetition, some other strategy, or no strategy were classified as normatively ineffective. As shown in Table 1, the pattern of reported strategies did not differ for judged pairs versus non-judged pairs. Overall, participants were no more likely to report using effective strategies for pairs that were judged versus not judged ($M = .24$, $SD = .16$ vs. $M = .24$, $SD = .18$), $t(63) = .63$, $p = .53$, 95% CI [−.02, .03], $g_{av} = 0.05$. This was also true for related pairs that were judged versus not judged, and unrelated pairs that were judged versus not judged (see Table 1).

We conducted conditional analyses on recall performance by reported strategy and pair type (Table 2) and found no significant differences between the JOL versus no-JOL conditions (all $ps > .05$). For judged related pairs, recall performance was near the ceiling and hence did not differ for pairs studied with effective versus ineffective strategies, $t(55) = 1.28$, $p = .21$, 95% CI [−.02, .09], $g_{av} = 0.25$. For judged unrelated pairs, recall was higher for pairs that were reportedly studied with effective strategies relative to those studied with ineffective strategies, $t(46) = 3.80$, $p < .001$, 95% CI [.13, .43], $g_{av} = 0.72$.

Following their strategy report, participants were also asked to provide a brief description of the strategy they used (e.g., the specific sentence or image they generated, or a description of another strategy). These fine-grained analyses may reveal qualitative differences in strategy use that are relevant to recall performance, such as if making JOLs increase the likelihood participants would generate interactive versus separate images (Begg, 1978).

Table 1. Mean proportion of strategies reported for related and unrelated pairs that were judged or not judged in Experiment 3.

	Effective strategies *M* (*SD*)	Ineffective strategies *M* (*SD*)	Don't remember *M* (*SD*)
Related Pairs			
JOL	.34 (.23)	.49 (.25)	.18 (.16)
No JOL	.32 (.25)	.43 (.27)	.24 (.21)
Unrelated Pairs			
JOL	.15 (.14)	.30 (.27)	.55 (.28)
No JOL	.15 (.16)	.29 (.29)	.56 (.29)

Note: JOL = judgment of learning. Effective strategies = reports of imagery or sentence generation. Ineffective strategies = reports of rote repetition, other strategy, or no strategy.

Table 2. Mean recall performance as a function of strategies reported for related and unrelated pairs that were judged or not judged in Experiment 3.

	Effective strategies *M* (*SD*)	Ineffective strategies *M* (*SD*)	Don't remember *M* (*SD*)
Related Pairs			
JOL	.96 (.14)	.91 (.15)	.42 (.39)
No JOL	.93 (.17)	.87 (.17)	.42 (.33)
Unrelated Pairs			
JOL	.69 (.38)	.39 (.37)	.03 (.08)
No JOL	.71 (.40)	.45 (.37)	.06 (.16)

Note: JOL = judgment of learning.

Descriptions were coded for total number of words, total number of content words (operationalised as nouns, verbs, adjectives, and adverbs), total number of rehearsals for pairs studied with a rote rehearsal strategy, and whether or not the cue and/or target were present in their description (as per Dunlosky et al., 2005). For cases where both the cue and target were present in the participants' description, two coders independently rated whether or not the cue and target were interacting (e.g., a *bird* has *feathers* on its body). The two coders agreed on 93.6% of the interactive descriptions, and disagreements were resolved by the first author. Table 3 contains a summary of these data by pair type. None of these qualitative analyses revealed any consistent pattern between pairs that were judged versus not judged.

Discussion

We again found a main effect of judgment condition in which recall was higher for pairs that were judged versus not judged. Can this recall benefit be explained by improved strategy use for judged pairs? The method used to collect strategy reports allowed us to conduct pair-level analyses of strategy reports, and our evidence suggests the answer to this question is "no". When participants reported their strategy use following recall, they were no more likely to report using effective strategies for pairs that were judged versus not judged.

Nevertheless, the method we used to obtain retrospective strategy reports – after cued recall was attempted – does have limitations that should be considered when interpreting these outcomes. First, for targets that are not recalled, participants may be less likely to remember the strategy that had originally been used to study the pair. Note, however, such forgetting would result in a greater number of effective strategies being reported when recall performance is higher, such as for pairs in which JOLs had been made. That is, this limitation would be expected to lead to more reports of effective strategies when JOLs are made, which did not occur. Second, participants may report using a particular strategy to learn a word pair (e.g., imagery) when they actually used another (e.g., repetition). When directly evaluating this possibility by collecting both concurrent strategy reports (made immediately after studying each word pairs) and retrospective strategy reports from each participant (as in Dunlosky & Hertzog, 2001), the reported strategies did align on the majority of the trials, but not all of them. Thus, finding no differences in self-reported strategy use between the JOL and no-JOL conditions may be at least partially driven by forgetting the strategy used to study some of the pairs when later making the retrospective reports. Considering these limitations, further research is needed to more fully explore other ways strategy use may contribute to JOL reactivity.

General discussion

Memory reactivity when JOLs are manipulated within participant

Our primary goal in the present research was to investigate JOL reactivity using a within-participant manipulation of JOLs. To obtain a high-powered evaluation of possible reactivity effects, we estimated effect sizes using a 2 (cue-target association) × 2 (judgment condition) × 3 (experiment) mixed ANOVA with the 341 participants who made JOLs for a randomly selected half of the studied pairs (including the related-first and unrelated-first groups of Experiment 2). The interaction between cue-target association and judgment condition was significant, $F(1, 340) = 8.34$, $p = .004$, $\eta_p^2 = .02$. For related pairs, recall was higher for pairs that were judged ($M = .81$, $SD = .15$) versus not judged ($M = .76$, $SD = .15$), $t(340) = 6.58$, $p < .001$, 95% CI [.04, .07], $g_{av} = 0.34$. For unrelated pairs, recall was also higher for pairs that were judged ($M = .25$, $SD = .17$) versus not judged ($M = .24$, $SD = .18$), $t(340) = 2.04$, $p = .042$, 95% CI [.0006, .03], $g_{av} = 0.10$, albeit a much smaller effect. These outcomes provide more competitive support for the cue-strengthening hypothesis than for the changed-goal hypothesis.

Mechanisms of reactivity when JOLs are manipulated within participant

What are the mechanisms underlying positive reactivity for related pairs? According to the cue-strengthening hypothesis, JOLs have a direct positive impact on recall by strengthening the association between a cue and target, but how does such strengthening occur? It could arguably arise from either quantitative or qualitative changes that are triggered by making a JOL for a specific pair. Concerning qualitative changes, one possibility is that making JOLs may encourage learners to use more effective strategies for judged pairs compared to non-judged pairs. In Experiment 3, we investigated this possibility by having participants report their strategy use following recall of each pair, which allowed us to conduct pair-level analyses of strategy reports. Converging with conclusions from Mitchum et al. (2016), we did not find any evidence that making JOLs induced learners to use more effective strategies for learning. However, one limitation is that JOLs may induce learners to change strategies in a way that was not captured by the strategy reports. For instance, perhaps participants who reported using imagery for a given pair generated a different or more detailed mental image when JOLs were elicited (such as in Pyc & Rawson, 2012), which would also be consistent with the cue-strengthening hypothesis. Although the current evidence suggests a minimal contribution of strategy shifts (Tables 1–3 and Mitchum et al., 2016), a lack of such strategy shifts does not necessarily negate the cue-strengthening hypothesis.

Another non-mutually exclusive possibility is that JOLs may lead to quantitative changes in learning. In particular, eliciting JOLs may reduce mind-wandering or increase attention to the on-going task. One possibility is that reduced mind-wandering acts at a global level: Simply expecting to make JOLs for some pairs might reduce mind-wandering for all studied pairs. If this were the case, we would not expect any reactive effects of JOLs using a within-participant design. Another possibility is that making JOLs do not globally increase attention but instead increase attention for only the pairs that are judged. In this case, the JOL prompt presented halfway through the presentation of a pair would serve to re-orient participants to the pair, leading to learning gains during the last few seconds of presentation. Our results suggest that the specific mechanism that leads to positive reactivity does not entirely act at a global level, given that a benefit occurred for judged pairs, and no carry-over effects occurred for non-judged pairs, such as predicted by the changed-goal hypothesis. Instead, at least in the present experiments, JOLs appear to act locally, by helping learners remember the judged pairs. In summary, the current research provides a unique outcome for reactivity research in being the first to demonstrate the reactive effect of JOLs using a within-participant design.

Finally, although our research provides more support for the cue-strengthening hypothesis than the changed-goal hypothesis, the changed-goal hypothesis could be revised to explain positive reactivity for related pairs even when JOLs are manipulated within participant. Perhaps JOLs do not lead participants to make a *global* shift in their overall learning goals for an entire list of word pairs (i.e., focus on the easier pairs at the expense of more difficult pairs). Instead, changes in learning goals may occur at the pair level. For example, perhaps when participants make a high JOL for a related pair, they exert extra effort on learning the pair, whereas when they make a low JOL for an unrelated pair, they stop attempting to learn it. In contrast, they do not change their learning efforts for pairs that do not receive JOLs. This modified version of the changed-goal hypothesis predicts positive reactivity for related pairs and negative reactivity for unrelated pairs. Given we did not observe negative reactivity for unrelated pairs, our results do not

Table 3. Summary of strategy descriptions provided by participants by pair type in Experiment 3.

	# Words *M* (*SD*)	# Content words *M* (*SD*)	# Rehearsals *M* (*SD*)	Cue present *M* (*SD*)	Target present *M* (*SD*)	Cue/target interacting *M* (*SD*)
Related Pairs						
JOL	5.39 (2.07)	3.60 (1.22)	3.63 (1.89)	.73 (.31)	.75 (.29)	.40 (.25)
No JOL	4.92 (1.87)	3.37 (1.13)	3.30 (1.74)	.71 (.33)	.73 (.30)	.36 (.28)
Unrelated Pairs						
JOL	6.86 (3.40)	4.66 (2.28)	3.71 (2.43)	.84 (.36)	.60 (.38)	.33 (.32)
No JOL	7.53 (3.80)	4.83 (2.12)	4.07 (2.08)	.86 (.23)	.57 (.39)	.30 (.33)

Note: JOL = judgment of learning. # Words: Average number of words in participants' description. # Content Words: Average number of nouns, verbs, adjectives, and adverbs in participants' description. # Rehearsals: Average number of times participants reported rehearsing pairs when using the rote repetition strategy. Cue Present: Proportion of times cue word was present in participants' description. Target Present: Proportion of times target word was present in participants' description. Cue/Target Interacting: Proportion of times cue and target word were coded as interacting in participants' descriptions.

support this version of the changed-goal hypothesis. However, the two hypotheses we evaluated in the current research are not mutually exclusive, and further research will be needed to reveal the specific, proximal mechanism that leads to reactive effects.

Mechanisms of reactivity when JOLs are manipulated between participants

Although the outcomes from this set of experiments support the cue-strengthening hypothesis, we would be remiss to neglect outcomes reported in prior research. Our experiments are the first to investigate reactivity for related and unrelated word pairs using a within-participant manipulation of JOLs within a study list (but see Myers et al., 2020, who manipulated JOLs within-participant using a blocked design). Multiple published studies have investigated reactivity with a between-participants design using related and unrelated pairs. The typical pattern of results found with a between-participants manipulation of JOLs is positive reactivity for related pairs, and negative or no reactivity for unrelated pairs (Janes et al., 2018; Mitchum et al., 2016; Myers et al., 2020). In our first experiment using a between-participants manipulated of JOLs, we observed positive reactivity for related pairs, and found a non-significant trend toward negative reactivity for unrelated pairs. Any amount of negative reactivity is difficult to explain solely with the cue-strengthening hypothesis.

What can we conclude about the different outcomes observed when a within versus between-participants manipulation of JOLs is used? We do not have a definitive answer to this question, but we offer two speculations to guide future inquiry. One is that a different mechanism is responsible for explaining reactive effects when JOLs are manipulated between versus within-participants. Perhaps results found with a within-participant manipulation of JOLs are best explained by the cue-strengthening hypothesis, whereas results found with a between-participants manipulation of JOLs are best explained with another hypothesis.

For example, the dual-task hypothesis (Mitchum et al., 2016) can explain why negative reactivity occurs for unrelated pairs in a between-participants manipulation of JOLs but positive reactivity occurs in a within-participant manipulation. For a between-participants design, this hypothesis states that the requirement to monitor learning could interfere with the primary task of memorising word pairs, particularly for learning difficult pairs. Thus, the non-judged unrelated pairs are processed more fully when JOLs are not made. By contrast, for a within-participant manipulation of JOLs, participants may begin monitoring their learning of all pairs in anticipation of making a JOL, so regardless of whether unrelated pairs are actually judged or not, they would suffer from the dual-task costs. The unrelated pairs that are then actually judged demonstrate (small) positive reactivity presumably because some of these pairs benefit from cue strengthening. Both of these explanations (and their corresponding assumptions) are speculative but can be empirically evaluated in future research.

Future research should strive to develop a coherent explanation of memory reactivity after learners make JOLs across multiple contexts (Double et al., 2018). In particular, research would benefit from continuing to explore memory reactivity with other materials (e.g., categorised lists, as per Senkova & Otani, 2021; longer text materials, as per Ariel et al., 2021), other types of judgments (e.g., confidence judgments, Double & Birney, 2017), other types of tests (e.g., free recall or recognition; Myers et al., 2020), and with other populations (e.g., older adults, Tauber & Witherby, 2019).

Summary of the present research

The present research establishes that even when immediate JOLs are manipulated within each participant, they have a reactive effect on memory performance. Such positive reactivity cannot be explained by changes in the strategies that learners used to study the judged pairs or focusing more attention on judged pairs than non-judged pairs. Instead, the reactive effects more likely arise from enhanced processing of the cue-target relationship (Soderstrom et al., 2015), as predicted by the cue-strengthening hypothesis. Along with recent studies (e.g., Janes et al., 2018; Mitchum et al., 2016; Myers et al., 2020; Soderstrom et al., 2015; Witherby & Tauber, 2017), these outcomes indicate that making JOLs can change the underlying learning process, both when JOLs are made for all studied material or just some studied material.

Note

1. Two targets shared the first three letters ("food" and "foot"), and only a single participant (in Experiment 3) recalled the word "food" when the correct target was "foot", which had no impact on any outcomes or conclusions.

Acknowledgements

The authors thank Rachel Hall, Chris Gifford, Bryton Hunt, Alex Knopps, Alesia Lambert, Mason McLeod, David Meekins, Benjamin Mitchell, Emily Moore, Jacob O'Connor, Aubrey Polaski, Mary Grace Swain, and Olivia Yee for their assistance with data collection, and Alexa Barrett and Alex Knopps for their assistance with strategy coding.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Michelle L. Rivers http://orcid.org/0000-0002-4931-2895

References

Akdoğan, E., Izaute, M., Danion, J. M., Vidailhet, P., & Bacon, E. (2016). Is retrieval the key? Metamemory judgment and testing as learning strategies. *Memory*, *24*(10), 1390–1395. https://doi.org/10.1080/09658211.2015.1112812

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126–131. https://doi.org/10.1037/h0027455

Ariel, R., Karpicke, J. D., Witherby, A., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, *33*(2), 693–712. https://doi.org/10.1007/s10648-020-09556-8

Begg, I. (1978). Imagery and organization in memory: Instructional effects. *Memory & Cognition*, *6*(2), 174–183. https://doi.org/10.3758/BF03197443

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68. https://doi.org/10.1037/0096-3445.127.1.55

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory (Hove, England)*, *4*(1), 31–48. https://doi.org/10.1080/741940667

Double, K. S., & Birney, D. P. (2017). Are you sure about that? Eliciting confidence ratings may influence performance on Raven's progressive matrices. *Thinking & Reasoning*, *23*(2), 190–206. https://doi.org/10.1080/13546783.2017.1289121

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, *26*(6), 741–750. https://doi.org/10.1080/09658211.2017.1404111

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*(2), 247–253. https://doi.org/10.3758/BF03194918

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*(2), 389–400. https://doi.org/10.1037/0012-1649.41.2.389

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1180–1191. https://doi.org/10.1037/0278-7393.27.5.1180

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Erlebacher, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin*, *84*(2), 212–219. https://doi.org/10.1037/0033-2909.84.2.212

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Janes, J., Rivers, M., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*, *25*(6), 2356–2364. https://doi.org/10.3758/s13423-018-1463-4

Jönsson, F. U., Hedner, M., & Olsson, M. J. (2012). The testing effect as a function of explicit testing instructions and judgments of learning. *Experimental Psychology*, *59*(5), 251–257. https://doi.org/10.1027/1618-3169/a000150

Kelemen, W. L., & Weaver, C. A. (1997). Enhanced memory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1394–1409. https://doi.org/10.1037/0278-7393.23.6.1394

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. https://doi.org/10.3758/BF03210951

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*(2), 200–219. https://doi.org/10.1037/a0039923

Morehead, K., Dunlosky, J., Rawson, K. A., Bishop, M., & Pyc, M. A. (2018). Does mediator use contribute to the spacing effect for cued recall? Critical tests of the mediator hypothesis. *Memory*, *26*(4), 535–546. https://doi.org/10.1080/09658211.2017.1381266

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*(2), 378–384. https://doi.org/10.3758/s13423-012-0343-6

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*(5), 745–758. https://doi.org/10.3758/s13421-020-01025-5

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Nelson, T. O., & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, *3*(5), 317–319. https://doi.org/10.1111/j.1467-9280.1992.tb00681.x

Pressley, Michael, Levin, Joel R, & Ghatala, Elizabeth S.. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 270–288. https://doi.org/10.1016/S0022-5371(84)90181-6

Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 737–746. https://doi.org/10.1037/a0026166

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, *137*(1), 131–148. https://doi.org/10.1037/a0021705

Richardson, J. T. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*(4), 597–614. https://doi.org/10.3758/BF03208837

Sahakyan, Lili, Delaney, Peter F, & Kelley, Colleen M.. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review*, *11*(*1*), 131–136. https://doi.org/10.3758/BF03206472

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition*, *49*(5), 955–967. https://doi.org/10.3758/s13421-020-01133-2

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 553–558. https://doi.org/10.1037/a0038388

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*(5), 315–317. https://doi.org/10.1111/j.1467-9280.1992.tb00680.x

Tauber, S. K., Dunlosky, J., & Rawson, K. A. (2015). The influence of retrieval practice versus delayed judgments of learning on memory: Resolving a memory-metamemory paradox. *Experimental Psychology*, *62*(4), 254–263. https://doi.org/10.1027/1618-3169/a000296

Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, *34*(6), 836–847. https://doi.org/10.1037/pag0000376

Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, *6*(4), 496–503. https://doi.org/10.1016/j.jarmac.2017.08.004

Yang, H., Cai, Y., Liu, Q., Zhao, X., Wang, Q., Chen, C., & Xue, G. (2015). Differential neural correlates underlie judgment of learning and subsequent memory performance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01699

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*(1), 41–44. https://doi.org/10.3758/BF03329756