

RIVERS, MICHELLE LAUREN, M.A. DECEMBER 2018 PSYCHOLOGY
INVESTIGATING MEMORY REACTIVITY WITH A WITHIN-PARTICIPANT
MANIPULATION OF JUDGMENTS OF LEARNING (49 pp.)

Thesis Advisor: John Dunlosky

Why does making *judgments of learning* (JOLs) influence subsequent memory, and when learners make JOLs for some items but not others, how is recall performance affected? To answer these questions, participants studied related and unrelated word pairs and made JOLs for half. Pair type was either randomly intermixed within a list (Experiment 1) or blocked (Experiment 2). I evaluated two hypotheses. The *changed-goal hypothesis*, proposed by Mitchum, Kelley, and Fox (*J Exp Psychol Gen*, 2016), states that making JOLs leads learners to notice differences in item difficulty and allocate more resources to learning easier pairs, ultimately leading to higher recall for easier (i.e., related) pairs and impaired recall for more difficult (i.e., unrelated) pairs. In contrast, the *positive-reactivity hypothesis* predicts increased recall performance for both related and unrelated pairs. As predicted by the positive-reactivity hypothesis, recall performance was higher for pairs that were judged versus not judged on both a mixed and blocked list of related and unrelated pairs. In Experiment 3, I evaluated one proximal mechanism for increased performance for judged pairs: The use of more effective encoding strategies during acquisition. Making JOLs did not influence strategy use, which suggests that the benefit of making JOLs on memory performance results from increased attention. These and other findings converge to support the claim that the requirement to monitor learning benefits memory.

INVESTIGATING MEMORY REACTIVITY WITH A WITHIN-PARTICIPANT
MANIPULATION OF JUDGMENTS OF LEARNING

A thesis submitted
to Kent State University in partial
fulfillment of the requirements for the
degree of Master of Arts

By

Michelle Lauren Rivers

December 2018

© Copyright

All rights reserved

Thesis written by

Michelle Lauren Rivers

B.A., University of California at Santa Cruz, 2013

M.A., Kent State University, 2018

Approved by

_____John Dunlosky_____, Advisor

_____Maria S. Zaragoza_____, Chair, Department of Psychological Sciences

_____James L. Blank_____, Dean, College of Arts and Science

TABLE OF CONTENTS.....	iv
LIST OF FIGURES.....	vi
LIST OF TABLES.....	vii
ACKNOWLEDGMENTS.....	viii
CHAPTERS	
I. Introduction.....	1
II. Experiment 1.....	5
2.1 Introduction.....	5
2.2 Method.....	6
2.3 Results.....	7
FIGURE 1.....	9
FIGURE 2.....	10
2.4 Discussion.....	11
III. Experiment 2.....	12
3.1 Introduction.....	12
3.2 Method.....	12
3.3 Results.....	13
FIGURE 3.....	15
FIGURE 4.....	16
FIGURE 5.....	17
3.4 Discussion.....	18
IV. Experiment 3.....	19

4.1	Introduction.....	19
4.2	Method.....	20
4.3	Results.....	21
	FIGURE 6.....	24
	TABLE 1.....	25
	TABLE 2.....	26
	TABLE 3.....	27
4.4	Discussion.....	28
V.	General Discussion.....	29
5.1	Memory Reactivity when JOLs are Manipulated Within Participant.....	29
5.2	Mechanisms of Positive Reactivity.....	31
5.3	Current Status of Theory for JOL Reactivity.....	32
	TABLE 4.....	34
5.4	Summary of the Present Research.....	35
	REFERENCES.....	36
	APPENDICES	
	A. Omnibus Factorial ANOVA for Experiment 1.....	40
	B. Simple Effects Not Reported in Text for Experiments 1 – 3.....	41

LIST OF FIGURES

Figure 1. Recall performance for the JOL and no-JOL groups in Experiment 1.....	9
Figure 2. Recall performance for the within-participant JOL group in Experiment 1.....	10
Figure 3. Recall performance for the mixed group in Experiment 2.....	15
Figure 4. Recall performance for the blocked design of Experiment 2 in which all related pairs were presented prior to unrelated pairs (related-first group).....	16
Figure 5. Recall performance for the blocked design of Experiment 2 in which all unrelated pairs were presented prior to related pairs (unrelated-first group).....	17
Figure 6. Recall performance in Experiment 3.....	24

LIST OF TABLES

Table 1. Proportion of Strategies Reported for Related and Unrelated Pairs that were Judged or Not Judged in Experiment 3.....	25
Table 2. Recall Performance as a Function of Strategies Reported for Related and Unrelated Pairs that were Judged or Not Judged in Experiment 3.....	26
Table 3. Summary of Strategy Descriptions Provided by Participants by Pair Type.....	27
Table 4. Continuously Cumulating Meta-Analysis of Memory Reactivity for Unrelated Word Pairs.....	34

Acknowledgments

I would like to thank John Dunlosky and Jessica Janes for their help in designing and implementing these experiments. I would also like to thank Nola Daley for her continual support as a colleague and friend.

I. Introduction

Hundreds of studies across almost 50 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, Dunlosky, & Kornell, 2013). The majority of prior research has used JOLs to investigate *how* people monitor their learning immediately after study (for a review, see Rhodes, 2016) and only recently has focused on the potential for these judgments to have reactive effects on subsequent memory (e.g., Janes, Rivers, & Dunlosky, 2018; Mitchum, Kelley, & Fox, 2016; Soderstrom, Clark, Halamish, & Bjork, 2015; Witherby & Tauber, 2017). The current investigation aimed to address the following questions: Does asking learners to provide immediate JOLs influence the ongoing learning process, such as by changing the strategies used to encode pairs? In particular, when learners make JOLs for some items but not others, how is recall performance affected? For the remainder of the introduction, I first discuss recent research demonstrating that making immediate JOLs can have a reactive effect on memory. I then describe two hypotheses for explaining these reactive effects, and introduce my approach to competitively evaluate them by 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 JOLs may reactively influence how learners process the to-be-learned items, ultimately influencing performance on a subsequent memory test (Ericsson & Simon, 1980). Three studies demonstrated the reactive effects of making JOLs on subsequent memory performance (Janes et

al., 2018; Mitchum et al., 2016; 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 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. Final recall performance was then compared between the JOL and no-JOL groups. Each study revealed a similar interaction: Whereas related pairs benefited from the requirement to make JOLs, unrelated pairs did not. Specifically, related pairs showed *positive reactivity*, with recall for related pairs being higher for the JOL group than the no-JOL group. In contrast, recall for the unrelated pairs showed either *negative reactivity* (lower recall for the JOL than no-JOL group) or no reactivity (statistically equivalent recall across the two groups).

What causes such reactivity? According to the *changed-goal hypothesis* (Mitchum et al., 2016), the requirement to make JOLs leads participants to consider that some pairs will be remembered and some will not. When studying a mixed list 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. Given the change in goals toward learning easier versus difficult pairs, this hypothesis predicts both positive reactivity for related pairs and negative reactivity for unrelated pairs.

To evaluate this hypothesis in the current study, 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. When JOLs are manipulated using a within-participant design, how will recall performance be affected? As with a between-participants manipulation of JOLs, the changed-goal hypothesis predicts that when learners make judgments, they shift their learning goal. If the mechanisms of reactivity are similar for both a within and between-participants manipulation of JOLs, then the changed-goal hypothesis predicts positive reactivity for related pairs and negative reactivity for unrelated pairs. Another possibility is that the changed learning goal invoked by making JOLs will carry over to non-judged pairs. That is, participants will focus on related pairs at the expense of unrelated pairs regardless of judgment condition, in which case I would expect no recall differences between judged and non-judged pairs.

Another explanation for reactivity is defined by the *positive-reactivity hypothesis* (Mitchum et al., 2016), which is that participants' memory generally benefits from the requirement to monitor learning. For instance, when participants are required to judge some pairs but not others, they may focus their attention on judged pairs, leading to higher recall performance. In contrast to the changed-goal hypothesis, the positive-reactivity hypothesis predicts positive reactivity for both related and unrelated pairs.

To competitively evaluate these hypotheses, I used a within-participant manipulation of JOLs¹, as noted above. In the within-participant group, 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, I found a main effect in which recall was greater

¹ No prior research has evaluated JOL reactivity with a 2 (related vs. unrelated) x 2 (JOL vs. no JOL) within-participant design, which was critical for competitively evaluating predictions made by the two hypotheses. 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, Bjork, & Schwartz, 1998; Kelemen & Weaver, 1997).

for judged than non-judged pairs – results that are better explained by the positive-reactivity hypothesis. In Experiment 2, I attempted to replicate this key finding and further evaluated the two hypotheses in a novel context – with related and unrelated pairs presented using a blocked list design (i.e., with all related pairs presented first followed by unrelated pairs for one group and vice-versa for the other). Finally, in Experiment 3, I evaluated whether learners make qualitative changes in their strategy use when judging pairs, which provides a proximal explanation for positive reactivity.

II. Experiment 1

2.1 Introduction

The main goal of Experiment 1 was to explore JOL reactivity using a within-participant design, but I 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). An added benefit of 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 (e.g., positive reactivity occurs for both pair types), then one possible concern is that the sample is not representative of those used in prior studies. However, if the between-participants outcomes (collected at the same time as the within-participant data using random assignment) do replicate prior ones (i.e., positive reactivity for related pairs and negative or no reactivity for unrelated pairs), then it would minimize these concerns.

I 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 critical group made JOLs for half of the pairs. For the JOL and no-JOL groups, I expected to replicate prior research showing positive reactivity for related pairs and either negative or no reactivity for unrelated pairs (e.g., Janes et al., 2018; Mitchum et al., 2016; Soderstrom et al., 2015). The recall pattern for participants who make JOLs for some pairs but not others, however, remains to be seen. Most important, as explained above, the two hypotheses make competitive predictions: Whereas the changed-goal hypothesis predicts positive reactivity for related pairs and negative reactivity for unrelated pairs (or no reactivity for either pair type), the positive-reactivity hypothesis predicts positive reactivity for both related and unrelated pairs.

2.2 Method

Design. Experiment 1 used a 2 x 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 random half of the studied pairs.

Participants. I aimed for 30 participants in each group as per Soderstrom et al. (2015; Experiment 1B). 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 or attempt recall on the final test.

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 = 0.57; e.g., *feathers – bird*) and half were unrelated (e.g., *mask – bread*) according to Nelson, McEvoy, and Schreiber (1998) free association norms.

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. During study, each participant was exposed to the 60 word pairs in a random order. Pairs were individually presented 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 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 (4 s) by the

computer program to type in their judgment within the allotted time for future pairs. Participants in the mixed group were prompted to make JOLs for a random 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 (i.e., recall the target word when presented with the cue word), with the order of cue words randomized for each participant. The full procedure took approximately 20 min.

2.3 Results

In Experiment 1, my primary goal was to investigate JOL reactivity using a within-participant manipulation of JOLs. First, I present analysis relevant to the replication of prior research that used a between-participants manipulation of JOLs (Figure 1), which mitigated any concerns about having a representative sample compared to prior research. Next, I report the novel findings of within-participant JOL reactivity (Figure 2). Although I present analyses of these groups separately below, this design also allowed me to analyze results with between vs. within effects as an independent variable. Results from this omnibus factorial analysis of variance (ANOVA) are presented in Appendix A. Cohen's d effect sizes were calculated using the formula from Cortina and Nouri (2000).

Between-participants manipulation of JOLs. A 2 (cue-target association: related vs. unrelated) x 2 (judgment group: judgment vs. no judgment) mixed ANOVA was conducted on memory performance (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. $M = .30$), $F(1, 58) = 540.83$, $p < .001$, partial $\eta^2 = .90$, but no significant main effect of judgment condition, $F(1, 58) = .01$, $p = .95$, partial $\eta^2 =$

.00. The interaction between cue-target association and judgment condition was significant, $F(1, 58) = 19.80, p < .001$, partial $\eta^2 = .25$. Planned comparisons (one-tailed test, for replications of prior research) confirmed that recall was higher for related pairs that were judged ($M = .81, SD = .11$) versus not judged ($M = .72, SD = .15$), $t(58) = 2.48, p = .01, d = .64$, and was lower for unrelated pairs that were judged ($M = .26, SD = .16$) versus not judged ($M = .35, SD = .25$; $t(58) = 1.68, p = .05, d = .43$). The latter negative reactivity for unrelated pairs using a between-participants manipulation is central for constraining theories of JOL reactivity; however, given that the focus across all experiments is on the within-participant manipulation of JOLs (vs. no JOLs), I do not consider this effect again until the General Discussion.

Within-participant manipulation of JOLs. A 2 (cue-target association) x 2 (judgment condition) repeated-measures ANOVA was conducted on memory performance (Figure 2), which revealed recall was higher for related than unrelated pairs ($M = .75$ vs. $M = .23$), $F(1, 31) = 322.51, p < .001$, partial $\eta^2 = .91$. Recall was also higher for pairs that were judged versus not judged ($M = .53$ vs. $M = .46$), $F(1, 31) = 17.99, p < .001$, partial $\eta^2 = .37$. The interaction between cue-target association and judgment condition was not significant, $F(1, 31) = 2.56, p = .12$, partial $\eta^2 = .08$. Despite the non-significant interaction, I still conducted paired-samples t -tests to estimate the effect size of JOL reactivity for the related and unrelated pairs (values presented in Appendix B).

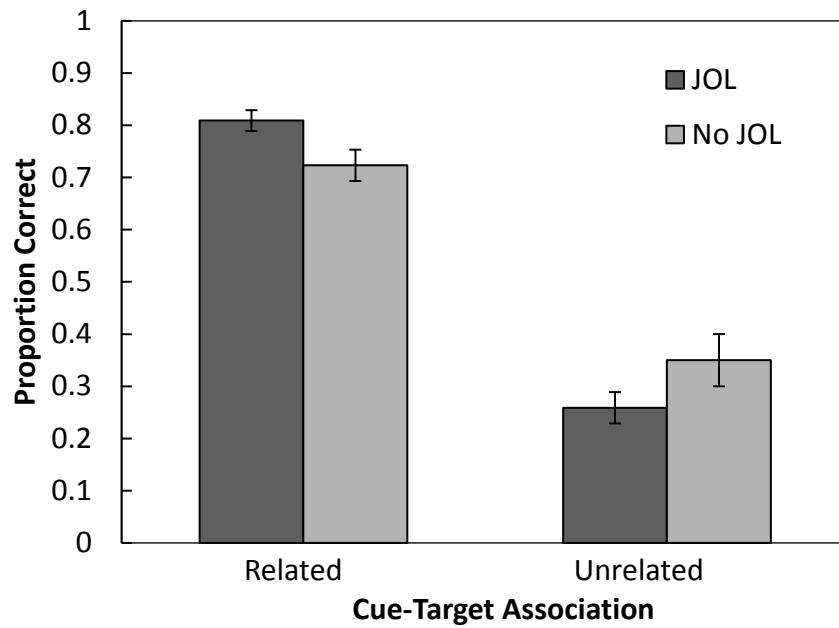


Figure 1. Recall performance for the JOL and no-JOL groups in Experiment 1. Error bars reflect standard error of the mean. JOL = judgment of learning.

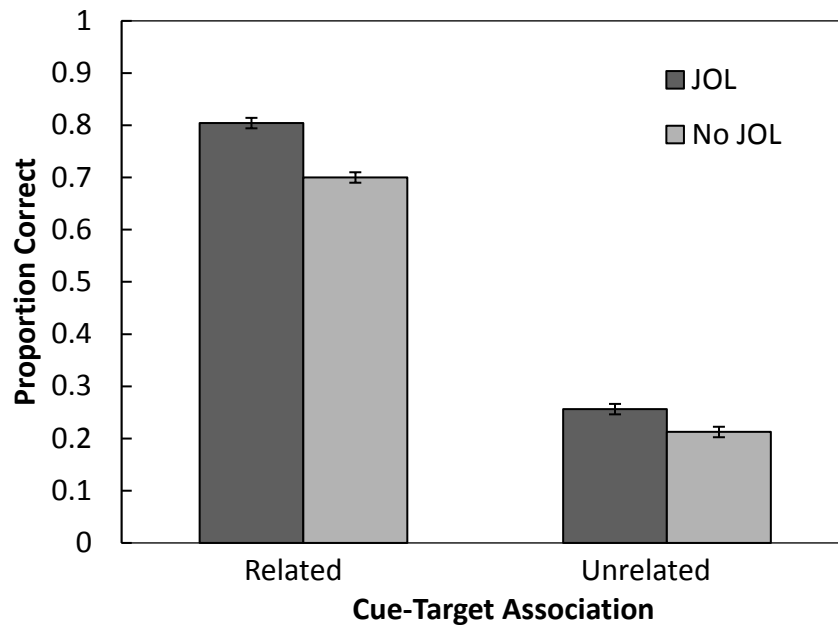


Figure 2. Recall performance for the within-participant JOL group in Experiment 1. Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions (for details, see Loftus & Masson, 1994).

2.4 Discussion

I replicated results from prior research using a between-participants manipulation of JOLs; namely, positive reactivity occurred for related pairs and a trend occurred for negative reactivity for unrelated pairs (consistent with results reported by Janes et al., 2018). Most important, with a within-participant manipulation of JOLs, I found an overall trend for positive reactivity, although reactivity tended to be larger for related than unrelated pairs (see Appendix B). These results are difficult to explain with the changed-goal hypothesis, which predicts negative reactivity for unrelated pairs, and provide more competitive support for the positive-reactivity hypothesis.

III. Experiment 2

3.1 Introduction

In Experiment 2, I attempted to replicate the main findings in Experiment 1 relevant to the within-participant manipulation of JOLs, and I further evaluated these hypotheses by presenting related and unrelated word pairs in a blocked order. For the latter, I 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.

The changed-goal hypothesis suggests that because participants will no longer be considering pair difficulty when making their JOLs on the first half of the blocked study list, no differences in recall should occur for pairs that are judged versus not judged (for either related or unrelated word pairs). However, on the second half of the blocked list, participants will notice the drastic differences in pair difficulty. In this case, the changed-goal hypothesis predicts that participants who study related pairs first will invest less effort on the second half of the list (comprised of the more difficult pairs) and thus show negative reactivity for the unrelated pairs on the second half of the list. For participants who study unrelated pairs first, the hypothesis predicts participants will invest more effort in learning the related pairs on the second half of the list, leading to positive reactivity for related pairs. In contrast, the positive-reactivity hypothesis predicts positive reactivity for both related and unrelated pairs.

3.2 Method

Participants and procedure. I used the software program G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to conduct a power analysis for a *t*-test comparing recall for judged versus non-judged pairs for unrelated pairs in the mixed group of Experiment 1. The goal was to obtain .80 power to detect a small effect size ($d = .28$) at the standard alpha error probability (.05), which yielded a sample size of 243. The final sample size was 246 participants, but one

participant was excluded from the unrelated-first group due to a computer malfunction.

Materials were the same as in Experiment 1. Participants were randomly assigned to one of three conditions: Eighty-three participants studied the related and unrelated pairs in a random order (identical to the mixed group of Experiment 1), 82 participants studied all 30 related pairs first followed by the 30 unrelated pairs (*related-first* group), and 80 participants studied all the unrelated pairs first followed by the related pairs (*unrelated-first* group). All participants made JOLs for a random half of the pairs. The distractor task and cued-recall test procedures were identical to Experiment 1.

3.3 Results

Mixed group. Recall performance is presented in Figure 3. A 2 (cue-target association) x 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. $M = .26$), $F(1, 82) = 1439.55$, $p < .001$, partial $\eta^2 = .95$. The main effect of judgment condition was also significant, with recall being higher for pairs that were judged versus not judged ($M = .55$ vs. $M = .51$), $F(1, 82) = 15.29$, $p < .001$, partial $\eta^2 = .16$. The interaction between cue-target association and judgment condition was not significant, $F(1, 82) = 0.08$, $p = .77$, partial $\eta^2 = .00$.

Related-first group. Recall performance is presented in Figure 4. A 2 (cue-target association) x 2 (judgment condition) repeated-measures ANOVA revealed a significant main effect of cue-target association with recall being higher for related than unrelated pairs ($M = .77$ vs. $M = .28$), $F(1, 81) = 983.28$, $p < .001$, partial $\eta^2 = .93$. The main effect of judgment condition was significant, with recall being higher for pairs that were judged versus not judged ($M = .54$ vs. $M = .51$), $F(1, 81) = 7.26$, $p < .01$, partial $\eta^2 = .08$. The interaction between cue-target association and judgment condition was not significant, $F(1, 81) = 0.12$, $p = .73$, partial $\eta^2 = .00$.

Unrelated-first group. Recall performance is presented in Figure 5. A 2 (cue-target association) x 2 (judgment condition) repeated-measures ANOVA revealed recall was higher for related than unrelated pairs ($M = .78$ vs. $M = .22$), $F(1, 79) = 725.71$, $p < .001$, partial $\eta^2 = .90$. The main effect of judgment condition was not significant, ($M = .50$ vs. $M = .49$), $F(1, 79) = 1.10$, $p = .30$, partial $\eta^2 = .14$. The cue-target association by judgment condition interaction was significant, $F(1, 79) = 5.38$, $p < .05$, partial $\eta^2 = .06$. For related pairs, the recall difference for the JOL ($M = .79$, $SD = .16$) and no-JOL ($M = .76$, $SD = .17$) conditions was significant, $t(79) = 2.41$, $p < .05$, $d = .22$. For unrelated pairs, the recall difference for the JOL ($M = .21$, $SD = .18$) and no-JOL ($M = .23$, $SD = .17$) conditions was not significant, $t(79) = .92$, $p = .36$, $d = .08$.

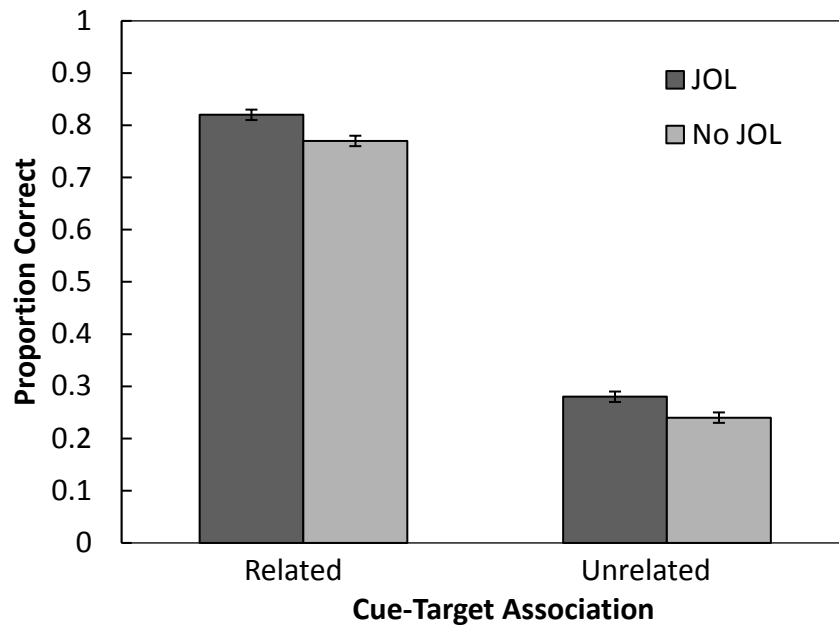


Figure 3. Recall performance for the mixed group in Experiment 2. JOL = judgment of learning. Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions.

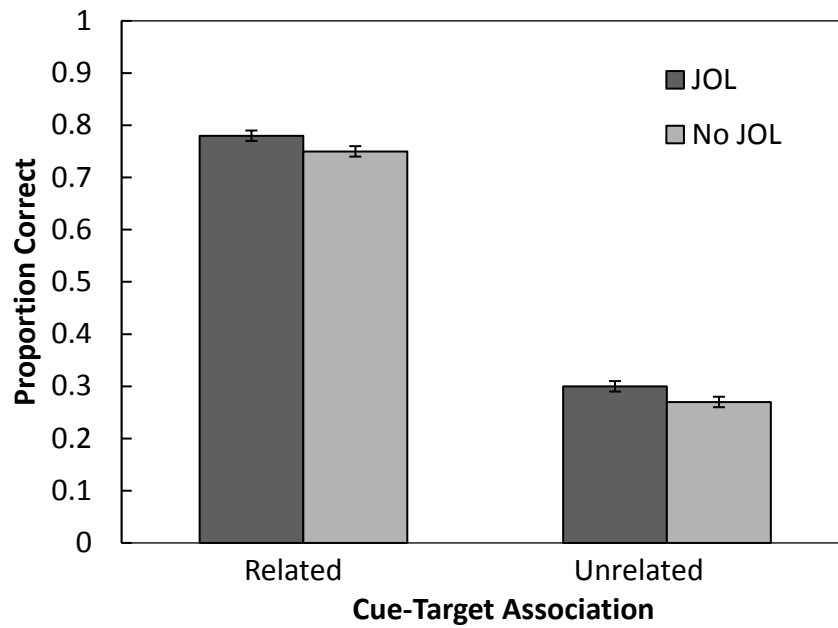


Figure 4. Recall performance for the blocked design of Experiment 2 in which all related pairs were presented prior to unrelated pairs (related-first group). JOL = judgment of learning. Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions.

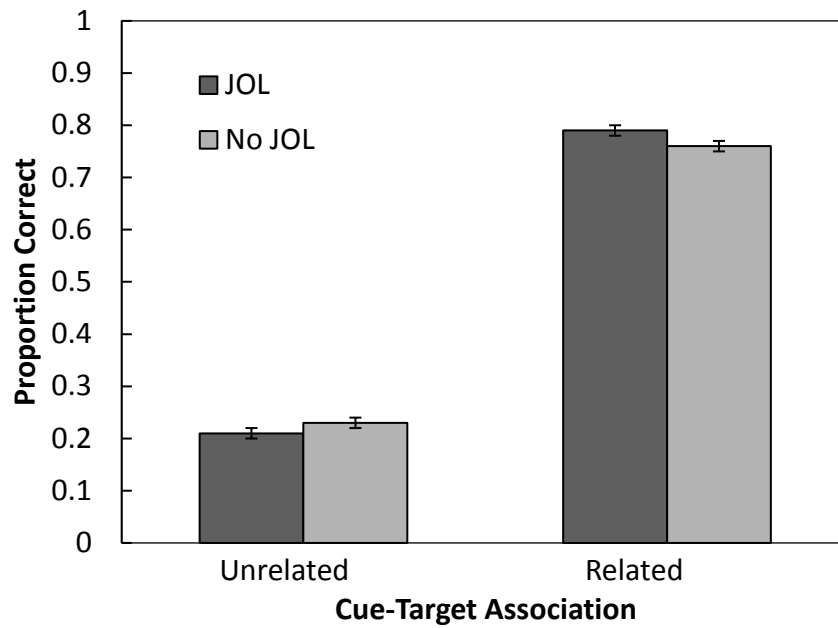


Figure 5. Recall performance for the blocked design of Experiment 2 in which all unrelated pairs were presented prior to related pairs (unrelated-first group). JOL = judgment of learning. Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions.

3.4 Discussion

Using the within-participant design wherein JOLs (vs. no JOLs) were mixed across the list, I replicated results of Experiment 1 and again found positive reactivity. Competitive support for the positive-reactivity hypothesis was also found with the related-first and unrelated-first groups. For the related-first group, I found positive reactivity for both pair types – results that are difficult to explain with the changed-goal hypothesis. For the unrelated-first group, I found significant positive reactivity for related word pairs, but no significant reactivity for unrelated pairs. I further consider these outcomes (i.e., the differential reactivity for related and unrelated pairs) with respect to current hypotheses of reactivity in the General Discussion.

IV. Experiment 3

4.1 Introduction

Results from the first two experiments provide more competitive support for the positive-reactivity hypothesis, which leads to the question: How might monitoring one's learning improve recall performance? That is, 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 – the generation of effective mediators. Specifically, making JOLs may induce learners to use more effective strategies during study. That is, the positive reactivity observed when participants make JOLS may 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 monitor learning leads to the use of more effective strategies was reported by Pressley, Levin, and Ghatala (1984) and Sahakyan, Delaney, and Kelley (2004), who found that participants who made judgments about memory performance 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 items they would be able to recall on a final test), which may operate differently from the item-by-item JOLs used in the present experiments.

Mitchum et al. (2016, Experiment 1) examined whether participants use different strategies when they make JOLs versus when they do not, as they were interested in whether a changed learning goal led learners to use better strategies for pairs that were judged versus not judged. Participants completed a post-experiment questionnaire, in which they were to rate (on a 0 to 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 item 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. Perhaps making JOLs led participants to use effective strategies more often for related pairs and less often for unrelated pairs, as per the changed-goal hypothesis. To evaluate this and other possibilities, I collected item-by-item reports of strategy use, which allowed me to examine whether participants use qualitatively different strategies making JOLs, and whether strategy use varied by pair relatedness.

4.2 Method

Participants and procedure. The target sample size was 60 participants. Sixty-four Kent State undergraduates participated. I 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 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, 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., 2017). 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 rating. Strategy reports and descriptions were self-paced.

4.3 Results

Recall performance is presented in Figure 6. A 2 (cue-target association) x 2 (judgment condition) repeated-measures ANOVA revealed that recall was higher for related than unrelated pairs ($M = .81$ vs. $M = .22$), $F(1, 63) = 993.51$, $p < .001$, partial $\eta^2 = .94$. Recall was also higher for pairs that were judged versus not judged ($M = .53$ vs. $M = .50$), $F(1, 63) = 5.67$, $p < .05$, partial $\eta^2 = .08$. Unlike prior experiments, the interaction between cue-target association and judgment condition was significant, $F(1, 63) = 5.82$, $p < .05$, partial $\eta^2 = .09$. For related pairs, the recall difference between the JOL ($M = .84$, $SD = .12$) and no-JOL ($M = .78$, $SD = .16$) conditions was significant, $t(63) = 3.48$, $p < .01$, $d = .45$. For unrelated pairs, the recall difference between the JOL ($M = .22$, $SD = .16$) and no-JOL ($M = .22$, $SD = .18$) conditions was not significant, $t(63) = .03$, $p = .98$.

To answer my primary question regarding strategy use, pairs in which participants reported using imagery or sentence generation were grouped into a single category, referred 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. Participants were no more likely to report using effective strategies for judged pairs compared to non-judged pairs ($M = .24$, $SD = .16$ vs. $M = .24$, $SD = .18$), $t(63) = .63$, $p = .53$.

I conducted conditional analyses on 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 ($M = .95$, $SD = .15$ vs. $M = .92$, $SD = .14$), $t(55) = 1.28$, $p = .21$, $d = .13$. For judged unrelated pairs, recall performance was higher for pairs that were reportedly studied with effective strategies relative to those studied with ineffective strategies ($M = .67$, $SD = .38$ vs. $M = .39$, $SD = .37$), $t(46) = 3.80$, $p < .001$, $d = .59$.

Recall that following their strategy report, participants were 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 subsequent memory performance, such as if making JOLs increase the likelihood participants would generate interactive versus separate images (Begg, 1978). Descriptions were coded for total number of words, total number of content words (operationalized 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, Hertzog, Powell-Moman, 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 I resolved any disagreements. 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.

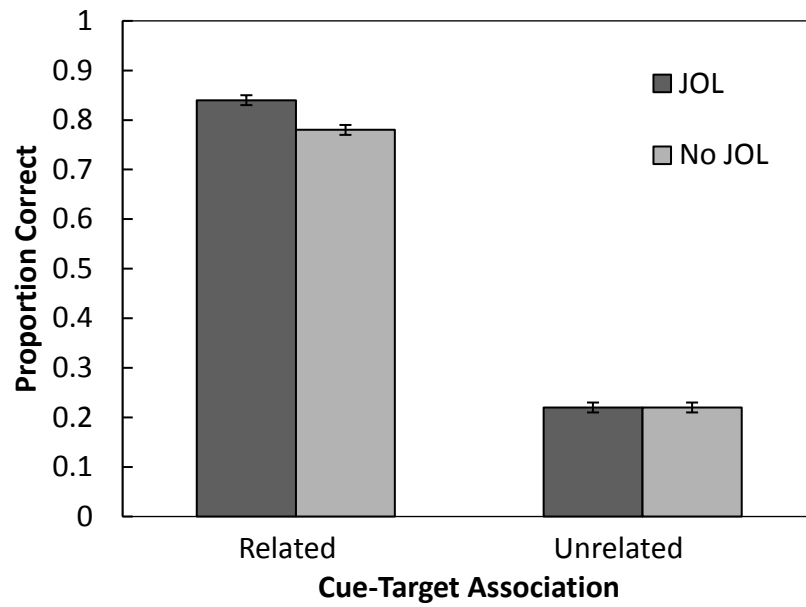


Figure 6. Recall performance in Experiment 3. JOL = judgment of learning. Error bars are standard errors computed (separately for related and unrelated pairs) for the within-participant contrast between the JOL versus no-JOL conditions.

Table 1

Proportion of Strategies Reported for Related and Unrelated Pairs that were Judged or Not Judged in Experiment 3

	Effective Strategies	Ineffective Strategies	Don't Remember
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

Recall Performance as a Function of Strategies Reported for Related and Unrelated Pairs that were Judged or Not Judged in Experiment 3

	Effective	Ineffective	Don't
	Strategies	Strategies	Remember
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
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.

Table 3

Summary of Strategy Descriptions Provided by Participants by Pair Type

	# Words	# Content Words	# Rehearsals	Cue Present	Target Present	Cue/Target Interacting
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.

4.4 Discussion

I again found a main effect of judgment condition in which recall was higher for pairs that were judged versus not judged, although the interaction revealed this effect was limited to related pairs. Can this recall benefit be explained by improved strategy use for judged pairs?

The method used to collect strategy reports allowed me to conduct item-level analyses of strategy reports, and the evidence suggests the answer to this question is “no” – I did not find any evidence that making JOLs induces learners to use more effective strategies for learning. 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.

V. General Discussion

5.1 Memory Reactivity when JOLs are Manipulated Within Participant

The primary goal in the present research was to investigate JOL reactivity using a within-participant manipulation of JOLs, which allowed me to evaluate two hypotheses. The changed-goal hypothesis (Mitchum et al., 2016) states that making JOLs leads learners to notice differences in pair difficulty and allocate more resources to learning easier pairs, ultimately leading to higher recall performance for easier (related) pairs and impaired recall for more difficult (unrelated) pairs. In contrast, the positive-reactivity hypothesis predicts increased recall performance for both related and unrelated pairs. For a high-powered test of these hypotheses, because there were three replications across Experiments 1 – 3, I estimated effect sizes using a 2 (cue-target association) x 2 (judgment condition) x 3 (Experiment: 1, 2, or 3) mixed ANOVA with the 179 participants that had a within-participant manipulation of JOLs (not including the related-first and unrelated-first conditions of Experiment 2). The main effect of judgment condition was significant, with recall being higher for pairs that were judged versus not judged ($M = .54$ vs. $M = .49$), $F(1, 176) = 35.99$, $p < .001$, partial $\eta^2 = .17$. Most important, the interaction between cue-target association and judgment condition was significant, $F(1, 176) = 6.08$, $p < .05$, partial $\eta^2 = .03$. For related pairs, the recall difference between the JOL ($M = .83$, $SD = .13$) and no-JOL ($M = .76$, $SD = .15$) conditions was significant, $t(178) = 6.0$, $p < .001$, $d = .44$. For unrelated pairs, the recall difference between the JOL ($M = .25$, $SD = .17$) and no-JOL ($M = .23$, $SD = .17$) conditions was also significant, $t(178) = 2.12$, $p < .05$, $d = .15$, albeit a much smaller effect. These outcomes provide more competitive support for the positive-reactivity than the changed-goal hypothesis. Further support for the former hypothesis was found in

Experiment 2, where positive reactivity occurred when related and unrelated pairs were presented in a blocked format.

I did not find negative reactivity for unrelated pairs in these within-participant JOL experiments, but as demonstrated above, I did find larger positive reactivity for related pairs compared to unrelated pairs. Why might positive reactivity be greater for related than unrelated pairs? One answer is provided by the *dual-task hypothesis* (Mitchum et al., 2016). This hypothesis states that the requirement to monitor learning could interfere with the primary task of memorizing word pairs, particularly for learners with low working memory capacity or for particularly demanding learning tasks. Because the task is arguably more demanding for learning unrelated than related word pairs, the former may suffer more from the dual-task costs and hence the benefits of making JOLs for unrelated pairs would be reduced.

An alternative hypothesis useful for understanding the larger positive reactivity effects for related (versus unrelated) pairs incorporates ideas from Koriat's (1997) cue-utilization framework for JOLs and transfer-appropriate processing (e.g., Morris, Bransford, & Franks, 1977; see also de Winstanley, Bjork, & Bjork, 1996), which I refer to as the *cue-relevant reactivity 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). Because learners base JOLs on associative relatedness (e.g., Dunlosky & Matvey, 2001; Koriat, 1997; Mueller, Tauber, & Dunlosky, 2013), JOLs strengthen the relationship between the two words, which is beneficial for a later cued-recall test. This hypothesis predicts greater positive reactivity for related than unrelated pairs (Soderstrom et al., 2015), because it is presumably easier to form a meaningful relationship between pairs that already have an inherent association.

Given that the dual-task and the cue-relevant reactivity hypotheses are not mutually exclusive, estimating their joint contribution to JOL reactivity poses a challenge for future research.

5.2 Mechanisms of Positive Reactivity

What are the mechanisms underlying positive reactivity? The positive-reactivity hypothesis (Mitchum et al., 2016) is mute about the proximal mechanism, but such benefits could arguably arise from either quantitative or qualitative changes during study. 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, I investigated this qualitative learning mechanism by having participants report their strategy use following recall of each pair, which allowed me to conduct item-level analyses of strategy reports. Converging with conclusions from Mitchum et al. (2016), I did not find any evidence that making JOLs induced learners to use more effective strategies for learning (Tables 1 & 3). Instead, making JOLs may lead to purely quantitative changes in learning.

Concerning possible quantitative changes, consider reduced mind-wandering (or increased 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, I 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 reorient participants to the pair, leading to learning gains during the last few seconds of presentation. As noted above, these 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 there were no carry-over

effects for non-judged pairs). Instead, JOLs appear to act locally, by helping learners remember the judged pairs.

5.3 Current Status of Theory for JOL Reactivity

Although the outcomes from this set of experiments provide more competitive support for the positive-reactivity hypothesis than the changed-goal hypothesis, I would be remiss to neglect outcomes reported in prior research. These experiments are the first to investigate reactivity for related and unrelated word pairs using a within-participant manipulation of JOLs, but three prior studies have investigated reactivity with a between-participants design, reporting results that are difficult to explain solely with the positive-reactivity hypothesis. In particular, the typical pattern of results found with a between-participants manipulation of JOLs is positive reactivity for related pairs and a negative trend for unrelated pairs (e.g., Janes et al., 2018; Mitchum et al., 2016).

I replicated this pattern in the first experiment, using a between-participants design. Specifically, I observed positive reactivity for related pairs, and found a non-significant trend toward negative reactivity for unrelated pairs. Given recent emphasis on basing conclusions on multiple estimates of effect sizes, I conducted a *continuously cumulating meta-analysis* (CCMA; Braver Thoemmes, & Rosenthal, 2014) to estimate the size of the negative reactivity for unrelated pairs. This analysis included five studies that have followed the exact procedures I used in the between-participants groups of Experiment 1 (Table 4). Although this analysis revealed only a small negativity effect ($p = .01$, $d = .23$), any amount of negative reactivity cannot be explained with the positive-reactivity hypothesis.

What can be concluded about the different outcomes observed when a within versus between-participants manipulation of JOLs is used? One plausible answer is that a different

mechanism is responsible for explaining reactive effects when JOLs are manipulated between versus within-participants. For instance, perhaps results found with a between-participants manipulation of JOLs are best explained by the changed-goal hypothesis, whereas the results from a within-participant manipulation of JOLs are better explained by the positive-reactivity hypothesis. I leave evaluation of these possibilities for future research.

Table 4

Continuously Cumulating Meta-Analysis of Memory Reactivity for Unrelated Word Pairs

	Mean diff	S_{pooled}	t	p	Cohen's d	Z
Soderstrom et al., 2015; Experiment 1b	.01	.17	.24	.82	.07	.23
Janes et al., 2018; Experiment 1	-.09	.23	1.67	.10	.40	1.65
Janes et al., 2018; Experiment 2	-.06	.27	.87	.39	.23	.86
Janes et al., 2018; Experiment 3	-.05	.26	1.56	.12	.18	1.56
Current Study; Experiment 1	-.09	.21	1.67	.10	.43	1.64
CCMA Results				.01	.23	2.66

Note. *Mean diff*: Mean difference between JOL and no-JOL groups in proportion correct on final recall test. Cohen's d reported as absolute values. Effect size homogeneity test was not significant, $Q(4) = 2.15, p = .71$.

5.4 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 subsequent memory performance. Such positive reactivity cannot be explained by changes in the strategies learners used to study the judged pairs. Instead, the reactive effects more likely arise from a quantitative change in study, perhaps by reorienting learners to engage in study when making JOLs. Along with recent studies (Janes et al., 2018; Mitchum et al., 2016; 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.

References

- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, 81, 126.
- Begg, I. (1978). Imagery and organization in memory: Instructional effects. *Memory & Cognition*, 6, 171-183.
- 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.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417-444.
- Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9(3), 333-342.
- Cortina, J. M., & Nouri, H. (2000). Effect size for ANOVA designs. Thousand Oaks, CA: Sage.
- de Winstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory*, 4, 31-48.
- Double, K. S., Birney, D. P., & Walker, S. A. (2017). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, 1-10.
- 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.
- 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.

- 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.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 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.
- 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.
- Janes, J., Rivers, M., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*.
<https://doi.org/10.3758/s13423-018-1463-4>
- Kelemen, W. L., & Weaver III, 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.
- 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.
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476-490.
- 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.

- Morehead, K., Dunlosky, J., Rawson, K. A., Bishop, M., & Pyc, M. A. (2017). Does mediator use contribute to the spacing effect for cued recall? Critical tests of the mediator hypothesis. *Memory*, 1-12.
- 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.
- 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.
- 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.
- Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (65 – 80). New York: Oxford University Press.
- 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.
- 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.
- Wetherby, 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.

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.

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.

Appendix A

Omnibus Factorial ANOVA for Experiment 1

Because I manipulated judgments both between and within participants, I could not treat judgment manipulation as a repeated measure for the 2 (judgment manipulation: between participants or within participant) x 2 (judgment condition: judgment vs. no judgment) x 2 (cue-target association: related vs. unrelated) mixed ANOVA. Instead, I treated judgment manipulation as a between-participants variable, which may have reduced the 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, I still obtained a marginally significant 3-way interaction, $F(1, 120) = 3.37, p = .07$. 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 (refer to the results section of Experiment 1).

Appendix B

Simple Effects Not Reported in Text for Experiments 1 – 3

	JOL	No JOL	<i>t</i>	<i>d</i>
	<i>M (SD)</i>	<i>M (SD)</i>		
Experiment 1				
Mixed Group				
Related Pairs	.80 (.15)	.70 (.18)	4.50	.71
Unrelated Pairs	.26 (.18)	.21 (.19)	1.56	.24
Experiment 2				
Mixed Group				
Related Pairs	.82 (.13)	.77 (.13)	3.07	.35
Unrelated Pairs	.28 (.17)	.24 (.17)	2.13	.23
Experiment 2				
Related-First Group				
Related Pairs	.78 (.15)	.75 (.15)	2.27	.25
Unrelated Pairs	.30 (.18)	.27 (.16)	1.69	.18

Note. JOL = judgment of learning.