Investigating the Effects of Associative Direction on Judgment of Learning Reactivity

Nicholas P. Maxwell & Mark J. Huff

The University of Southern Mississippi

Author Note

Correspondence concerning this article should be addressed to Nicholas P. Maxwell, 118 College Dr, Hattiesburg, MS, 39406. E-mail: [nicholas.maxwell@usm.edu](mailto:nicholas.maxwell@usm.edu). *R* code used for data screening and analyses as well as all applicable stimuli and data files have been made available on our OSF page (https://osf.io/8yvn3/).

Abstract

Previous research has shown that judgments of learning (JOLs) produce a reactive effect on learning in which correct recall differs between participants who provide JOLs at study versus those who do not. The effects of providing JOLs on memory have been mixed: Some studies have reported a memory benefit (positive reactivity), while others have reported a memory cost (negative reactivity). Importantly, little work has evaluated the interaction between the direction of the associative relationship in cue-target pairs (i.e., credit-card vs. card-credit) and reactivity. Across three experiments, we show that (1) compared to a no-JOL control, providing JOLs produced positive reactivity for paired associates regardless of pair direction but no reactivity for unrelated pairs, (2) recall of paired associates was highest when participants were asked to make JOLs at study relative to deep encoding regardless of pair direction, and (3) frequency judgments produced a reactivity pattern that mimicked JOL reactivity. Our findings suggest that JOLs are largely beneficial to cued-recall and that reactivity effects may be driven by the processes involved when making pair ratings rather than predictions about future performance.

Word count: 180

*Keywords:* Metacognition; Judgments of Learning; Associative Direction; Cued-Recall

Investigating the Effects of Associative Direction on Judgment of Learning Reactivity

An individual’s to ability to accurately monitor the progress of their own learning is a critical component of successful retention. Effective monitoring allows individuals to adjust their study strategies to maximize item retention at test (Nelson & Narens, 1990) and provides insights on how best to allocate resources to optimize performance on future learning tasks (Soderstrom, Clark, Halamish, & Bjork, 2015; see also Bjork, 1999 for a review). Empirically, information about the learning processes can be obtained through metacognitive judgments (i.e., having individuals make judgments about their memorial abilities), which have received significant attention from memory researchers (see Metcalfe, 2000 for a historical overview of metacognitive judgment making within the field of metamemory). Relatively little work, however, has been conducted assessing whether or not the simple act of providing these judgments at study can affect memory performance and, if so, what factors potentially moderate this effect.

One type of metacognitive judgment that is commonly elicited at study is the judgment of learning (JOL). In a standard JOL task, participants are presented with a cue-target study pair (traditionally paired associates) and are asked to respond with the likelihood that they would respond with the correct target at test if given only the cue. While these judgments can be made using a variety of scales (e.g., Likert scales or binary “yes”-“no” responses; Hanczakowski, Zawadzka, Pasek, & Higham, 2013), JOLs are elicited using a continuous 0 to 100 scale that represents the percent likelihood that the cue-target pair would be successfully recalled (e.g., 100% = definitely would remember; 0% = definitely would not remember). The use of a 100 point scale allows for an easy comparison between predicted recall (via JOLs) and the overall proportion of items recalled at test (i.e., predicted recall vs actual recall performance).

Recently, several studies have sought to test whether the act of making JOLs at study is *reactive*. JOL reactivity would occur if the act of making judgments at study inadvertently altered performance at test relative to a condition in which to when no JOLs are elicited. Reactive effects could facilitate learning, a pattern termed positive reactivity, or could produce a memory cost, a pattern termed negative reactivity. The act of providing JOLs may produce either memory benefits or costs relative to studying cue-target pairs in the absence of providing JOLs.

A simple way to assess whether JOLs produce a reactive effect on learning is to compare recall performance for participants who complete a judgment task at study versus those who do not (Soderstrom et al., 2015; Janes, Rivers, & Dunlosky, 2018). Thus, evaluating reactivity simply involves the inclusion of a no-JOL control group, a comparison that is generally absent in JOL studies as researchers have been more interested in condition-specific effects on JOLs themselves (and not overall memory performance) or have made the assumption that the act of providing JOLs at study has no impact later memory. However, given no-JOL control groups are not often included, this assumption cannot be confirmed.

The lack of no-JOL controls across many JOL studies is surprising given that early evidence for the reactive effects of JOL on memory were documented in Arbuckle and Cuddy’s (1969) early study on JOLs. In their second experiment, metacognitive judgments were elicited using a 1-5 Likert scale, and critically, participants provided judgements either during both the study and test phases or only at test. Ratings made at study were framed as a JOL (i.e., subjects indicated their ability to correctly recall pairs at test), while judgments made at retrieval were made as a confidence rating (i.e., confidence that the response provided is correct). This design allowed for a comparison between groups in which metacognitive judgments were provided at both study and test versus a group that only made judgements at test. Importantly, a positive reactivity pattern emerged in which correct recall was greater for participants who were required to make judgements at encoding. Though the authors did not provide an in-depth discussion of these findings, they did note that making predictions did not produce a negative reactivity pattern and therefore did not interfere with recall. Thus, it is clear that providing JOLs at study does affect encoding processes. However, it is important to note that although Arbuckle and Cuddy reported that JOLs can boost recall, both JOL and non-JOL groups provided confidence ratings at test, making it unclear whether confidence ratings were a requisite for positive reactivity.

More recently, Soderstrom et al. (2015) had participants study a list of cue-target pairs in which contained a mix of related and unrelated word pairs. After studying each pair, one group of participants were instructed to provide a JOL, while a no-JOL control group simply studied each pair in isolation. Participants were then tested on their recall of the target word when presented with the cue without additional metacognitive judgments (cf. Arbuckle & Cuddy, 1969). Overall, target recall was found to be greater for participants who provided JOLs initially versus those who did not, however, this positive reactivity pattern was restricted to related pairs. For unrelated pairs, target recall did not differ between the JOL and non-JOL groups. A similar pattern was reported by Janes et al. (2018), who also showed that initial JOLs produced positive reactivity for targets from related but not unrelated pairs. Finally, Witherby and Tauber (2017) showed that positive reactivity for paired associates extends to long-term retention.

In contrast to the positive reactivity for JOLs associated with related pairs reported by Soderstrom et al. (2015) and Janes et al. (2018), Mitchum, Kelly, and Fox (2016) reported a divergent pattern of reactivity. In their study, participants who provided JOLs at study showed no difference in later recall relative to a no-JOL group on related pairs and produced a negative reactivity pattern relative to the no-JOL group for unrelated pairs. Mitchum et al. (2016) interpreted this discrepancy as arising from methodological differences between their study and the one conducted by Soderstrom et al. (2015), though they noted that results of Soderstrom et al. (2015) were most similar to findings obtained in their fifth experiment, in which the study/JOL block was experimenter paced. Taken together, these studies demonstrate that providing JOLs when studying cue target pairs can induce reactivity on target learning, but the direction of reactivity has been mixed with positive or no reactivity reported when pairs are related and negative or no reactivity reported with unrelated pairs.

**Mechanisms Driving JOL Reactivity**

To account for JOL reactivity processes, Mitchum et al. (2016) proposed three accounts. First, the *positive reactivity hypothesis* states that given monitoring is a key aspect for determining the effectiveness of the learning process (e.g., Nelson & Narens, 1990), retention will benefit from any additional monitoring that may occur as a byproduct of providing JOLs at study. Thus, recall should be more accurate when JOLs are provided versus a non-JOL control regardless of the relatedness of the study pairs. Separately, the *dual-task* *hypothesis* suggests that generating JOLs will produce negative reactivity for all study materials versus a no-JOL control as providing JOLs is resource demanding which may interfere with the learning of word pairs (Hertzog, Dunlosky, Powell-Moman & Kidder, 2002). Finally, the *changed-goal hypothesis* proposes that JOL reactivity occurs due to online changes in participant study goals that arise at encoding. According to this hypothesis, participants set an initial goal of memory mastery when employing a study task, and strategically allocate more time and/or effort towards studying items perceived as being difficult to remember than those perceived as easy, to maximize retention. However, certain conditions have been shown to induce a change of study goal. Metcalfe & Kornell (2003) have shown that when study time is limited, participants prioritize learning of items that are perceived as “easy” with the goal of mastering as many total items as possible during the time restraint. Furthermore, when providing JOLs (specifically those utilizing a 0 to 100 rating scale), it becomes more obvious to participants that not all items will be equally recalled. Thus, participants may use their perceptions of difficulty to shift their study goals towards mastering easier items (Ariel, Dunlosky, & Bailey, 2009).

Within the context of JOL reactivity effects found with word pairs, the changed-goal hypothesis assumes that study lists will provide participants with at least two distinct pair types and that participants will be able to discriminate between. This hypothesis predicts that providing JOLs will induce positive reactivity for pairs that are perceived as easy to remember, but negative reactivity for pairs that are perceived as difficult to remember. This is because when individuals detect differences in difficulty between pair types, they will prioritize encoding of the easier to remember related pairs which will come at a cost to the encoding of more difficult unrelated pairs. Thus, the changed-goal hypothesis predicts a divergent memory pattern when comparing JOL to non-JOL groups as a function of individual’s perceptions of pair difficulty.

Although reactivity patterns as a function of pair association have been mixed (e.g., Janes et al., 2018; Mitchum et al., 2016; Soderstrom et al., 2015), a meta-analysis conducted by Double, Birney, and Walker (2018) which included 17 published and non-published experiments comparing JOL to non-JOL groups has provided partial support for the changed-goal hypothesis. Specifically, providing JOLs yielded a positivity effect for recall of related targets overall, but showed no effect of negative reactivity on recall of unrelated targets relative to no-JOL controls. In terms of the changed-goals hypothesis, it therefore appears that individuals prioritize encoding of related pairs when making JOL ratings, but this priority is not accompanied by a concomitant cost to the encoding of unrelated pairs.

**Associative Direction and JOL Accuracy**

The associative direction between word pairs has been shown to directly influence both how well individuals recall items at test and the accuracy of metacognitive judgments made at study. Koriat and Bjork (2005; see too Koriat & Bjork, 2006) demonstrated across three experiments that JOLs for pairs associated in the forward direction (e.g., credit-card) were accurate at predicting later recall of the target item. When forward association strength between pairs was weak (e.g., article-newspaper), JOLs were less predictive of later recall relative to when the forward association between pairs was strong (e.g., lost-found). For weak forward pairs, JOLs ratings were similar to those given to strong associates, but recall was reduced as weakly related cue words were less effective in aiding retrieval of the target. Thus, the *calibration* between JOLs and recall was moderated by the strength of the forward cue-target association.

In addition to forward associates, Koriat and Bjork (2005) also evaluated the correspondence between JOLs and target recall for pairs associated in the backward direction (e.g., card-credit). Like weak associates, backward associates received high JOL ratings, but recall for the target word was considerably lower relative to forward pairs. Dubbed the *illusion of competence,* this overestimationpattern has been extended to other pair types. Castel et al. (2007) showed that the illusion of competence extended to identical pairs in which the cue is perfectly predictive of the target (e.g., lost-lost). More recently, Maxwell and Huff (in press) further investigated the influence of associative direction on JOLs and found that the illusion of competence holds for backward associates after controlling for lexical and semantic properties of the cue and target (e.g., word length, concreteness, etc.) and extends to symmetrical associates (e.g., off-on) that were matched on associative strength with forward and backward associates.

The illusion of competence serves as an example of how the directional correspondence between associative pairs can affect the predictive capacity of JOLs on later recall. Although most studies investigating the reactive effects of JOLs (as well as the larger body of JOL literature as a whole) have relied almost exclusively on comparing cued-recall performance between forward associates and unrelated pairs, a notable exception was conducted by Mitchum et al. (2016, Experiment 1), who presented participants with forward associates, backward associates, and unrelated pairs within the same study list. As reported above, no reactivity effect was found for related pairs, and this null effect was found on both forward and backward associates. Despite this null pattern, the authors concluded that the changed-goal hypothesis was partially supported as participants allocated their study time differently depending on the type of task they received. While participants spent the most time studying unrelated pairs across JOL and no-JOL conditions, JOL participants spent significantly less time studying unrelated item pairs relative to participants who did not make JOL ratings.

Although Mitchum et al. (2016) showed reactivity results that were inconsistent with other JOL reactivity studies (e.g., Janes et al., 2018; Soderstrom et al., 2015), it is also worth pointing out another inconsistency in their data: No illusion of competence pattern emerged for backward pairs (cf. Castel et al., 2007; Koriat et al., 2005; Maxwell & Huff, in press). Though Mitchum et al. reported lower recall for backward than forward pairs across JOL and non-JOL groups, these differences were much smaller than those typically reported. This discrepancy may have resulted from how association was measured across these studies. Koriat and Bjork for instance used Hebrew word pairs derived from a set of Hebrew free association norms, while Mitchum et al. used English word pairs derived from the University of South Florida Free Association Norms (USF norms; Nelson, McEvoy, & Schreiber, 2004) as well as a relatedness score calculated with Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). However, we Maxwell and Huff (in press) similarly utilized the USF norms as in Mitchum et al. (2016) and found a more robust illusion of competence pattern.

A second possibility for this discrepancy is that although association between pair types was assessed and manipulated, neither Koriat and Bjork (2005) nor Mitchum et al. (2016) controlled for lexical and semantic item characteristics that may have covaried across pair types. Characteristics such as word length, frequency, and concreteness have each been shown to affect later recall (Madan, Glaholt, & Caplan, 2010; Criss, Aue, & Smith, 2011) and are therefore confounded with associative direction in these studies. Thus, given discrepancies in recall that occur as a result of pair direction (i.e., the illusion of competence), it remains unclear whether pair direction could moderate JOL reactivity (i.e., reactivity greater for forward vs. backward pairs).

**Experiment 1: Judgments of Learning vs Study**

In Experiment 1, we sought to provide a further test of the changed-goal hypothesis of JOL reactivity. Participants studied three types of related pairs (forward, backward, and symmetrical) and a set of unrelated items. Additionally, we controlled for potential lexical and semantic item effects that were not equated for across pair types in previous studies (e.g., Soderstrom et al., 2015; Janes et al., 2018). All items were matched on word frequency, concreteness, and length. Related pairs were further matched on associative strength. In addition to controlling for item effects, we also included a comparison to symmetrical associates (i.e., pairs with equivalent forward and backward relatedness)—a novel comparison. For the changed-goal hypothesis to be upheld, related pairs should show a boost to correct recall when JOLs are made versus the No-JOL group, regardless of pair direction. Furthermore, based on Mitchum et al.’s (2016) description of the changed-goal hypothesis, unrelated pairs should show a memory cost when JOLs are made. However, it is unclear whether this latter pattern will occur given studies have shown little evidence of negative reactivity within this context. Thus, Experiment 1 was provided an additional test for the presence of negativity reactivity.

**Methods**

**Participants**

55 participants were recruited online using Prolific Academic and were compensated at a rate of $9.50/hour. A sensitivity analysis conducted with *G\*Power* (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that our sample size provided adequate power (0.80) to detect a small effect size (Cohen’s *d* = 0.27) or larger. All participants were native English speakers with normal or corrected-to-normal vision.

**Materials**

Stimuli pairs were derived from Maxwell and Huff (in press). These pairs consisted of 180 word pairs generated from the University of Florida Free Association Norms (Nelson et al., 2004). Pairs were equally split into four types. As such, the stimuli consisted of 40 forward pairs (e.g., credit-card), 40 backward pairs (e.g., card-credit), 40 symmetrical pairs in which forward and backward strength were equivalent (e.g., ball-bounce), and 40 unrelated pairs (e.g., artery-bronze). Additionally, 20 non-tested buffer pairs were generated to control for primacy and recency effects. Item pairs were equally distributed across two study lists. Thus, each list contained 20 of each pair type and ten buffer items. All item pairs have been made available at [OSF LINK].

Each item list was presented as an individual study-test block, and list order was counter-balanced across participants. Lists were created such that the 80 tested pairs (20 of each item type) were always proceeded and followed by five buffer pairs, leading to 90 item pairs per study list. Additionally, lists were constructed such that pair types were equated on both associative strength (e.g., FAS and BAS values derived from the Nelson et al. (2004) free association norms), frequency (SUBTLEX; Brysbaert & New, 2009), word length, and concreteness (taken from the English Lexicon Project; Balota et al., 2007; See the Tables A1-A2 in the Appendix for associative strength and the lexical properties for each pair type). Finally, counterbalanced versions of each study list were created such that flipped the order of words with each of the four pair types (i.e., king-queen becomes queen-king). While the order was switched across all pair types, this was especially impactful for the two asymmetrical pair types (e.g., forward and backward pairs) and provided us with greater control of item differences (i.e., forward pairs became backward pairs in and vice versa across counter-balances).

The cued-recall test was generated from all 80 cue items (buffers were not tested) by replacing the target item with a question mark (i.e., credit - ?). Test items were presented in a random order for each participant.

**Procedure**

Data collection was conducted online using *Collector*, an open source program for presenting web-based psychological experiments (Garcia & Kornell, 2015). In both the JOL and no-JOL conditions, participants were instructed that they would view a series of cue-target word pairs and that their memory for the target item would be tested. Participants in the JOL condition received further instructions to rate the likelihood that they would be able to remember the target word if shown only the cue at test. Judgments were elicited using a scale of 0-100, in which 0 indicated that they would be completely unable to recall the item at test, while a rating of 100 represented full certainty in their ability to correctly recall the target. After receiving instructions, participants then began the first study list. Study was self-paced, with participants in both conditions pressing the Enter key to advance to the next pair. Additionally, participants in the JOL condition were asked to type a JOL rating before proceeding to the next study pair. JOL ratings were provided concurrently with study such that ratings were typed while the pair was displayed on the computer screen.

Following presentation of the first study list, participants completed a two-minute filler task in which they were asked to list the 50 U.S. states in alphabetical order. This was immediately followed by a cued-recall test that presented participants with the cue word from each of the previously studied items. Participants were asked to type the correct target item. If participants could not retrieve the correct item, the Enter key could be pressed to move forward to the next pair. Following completion of the first cued-recall test, participants began the second block, which followed the same format as the first block. Participants were fully debriefed following the completion of the second cued-recall test. Each session lasted approximately 30 minutes.

**Results**

A *p* < 0.05 significance level was used for all analysis. Partial eta-squared (*ηp*2)and Cohen’s *d* effect sizes were reported alongside all significant analyses of variance (ANOVAs) and *t* tests. Figure 1 (top panel) plots mean JOL ratings and cued-recall rates for each

pair type for participants in the JOL study condition, while the bottom panel compares recall rates for participants who made JOLs at study vs those who silently read pairs at encoding. For completeness, all individual comparisons between of JOL ratings and correct recall proportions are displayed in Table A3, and all comparisons between correct recall proportions for each encoding strategy are reported in Table A4.

First, we conducted a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 2 (Measure: JOL vs. Recall) repeated measures ANOVA to assess whether the illusion of competence first reported by Koriat and Bjork (2005) replicated for participants in the JOL study condition. First, a main effect of Pair Type, *F*(3, 78) = 318.46, *MSE* = 96.85, *ηp*2 = .68, was detected in which JOLs/recall rates for were highest for forward (66.74) pairs, followed by symmetrical pairs (63.26), backward pairs (45.88), and unrelated pairs (14.46). Pairwise *t*-tests showed that JOLs/recall rates significantly differed across all comparisons, *t*s ≥ 3.07, *d*s ≥ xx. Next, a significant effect of measure was observed, *F*(1, 26) = 5.14, *MSE* = 602.92, *ηp*2 = .07, in which overall JOL ratings (51.37) exceeded later recall rates (43.80). Importantly, a significant interaction between Pair Type and Measure, *F*(3, 78) = 44.85, *MSE* = 45.82, *ηp*2 = .12, confirmed that the illusion of competence replicated with this dataset. Follow up *t* tests indicated a robust illusion of competence for backward pairs wherein JOLs exceeded later recall accuracy (57.06 vs. 34.69), *t*(26) = 5.32, *SEM* = 3.61, *d* = XX. Additionally, the illusion of competence extended unrelated pairs (18.91 vs 10.01), *t*(26) = 2.65, *SEM* = 3.51, *d* = XX, though to a lesser magnitude relative to backward pairs. For forward pairs, the pattern reversed, with JOL ratings significantly lower than recall (62.86 vs. 70.62), *t*(26) = 2.25, *SEM* = xx, *d* = XX, indicating that participants generally underestimated their performance for this pair type rather than overestimating it. Finally, for symmetrical pairs, JOLs exceeded recall (66.66 vs 59.87), though this difference was marginal, *t*(26) = 1.81, *SEM* = 3.61, *p* = .08, *d* = XX.

Next, a 4 (Pair Type: Forward vs Backward vs Symmetrical vs Unrelated) × 2 (Study Condition: JOL vs No-JOL) mixed measures ANOVA was used to test for differences in recall rates between within the four pair types and tested whether correct recall differed as a function of pairs receiving JOLs. First, this analysis yielded a main effect of Pair Type, *F*(3, 156) = 371.04, *MSE* = 73.07, *ηp*2 = .64, in which across study conditions, correct recall was greatest for forward pairs (60.97), followed by symmetrical pairs (48.29), backward pairs (25.69), and unrelated pairs (10.98). Post-hoc *t*-tests indicated that all comparisons significantly differed, *t*s ≥ 6.83, *d*s ≥ xx. Next, a significant main effect of Study Condition was observed, *F*(1, 52) = 17.71, *MSE* = 656.88, *ηp*2 = .20, in which correct recall in the JOL condition (43.82) exceeded that of the read only study condition (29.14), indicating that making JOLs at study had a reactive effect that was largely beneficial to cued-recall. Finally, a significant interaction between Pair Type and Study Condition, was observed, *F*(3, 156) = 23.55, *MSE* = 73.07, *ηp*2 = .10. Post-hoc testing indicated that positive JOL reactivity occurred, but only when participants studied related pairs. Correct recall in the JOL condition exceeded that of the read only group for forward pairs (70.56 vs 51.39), backward pairs (34.81 vs 16.57), and symmetrical pairs (59.91 vs 36.67), *t*s ≥ 4.20, *d*s ≥ xx. When participants studied unrelated pairs (10.00 vs 11.94), the reactive effects failed to occur, as correct recall did not differ as a function of encoding strategy, *t*(49) = 0.71, *SEM* = 2.81, *p* = .48. Thus, JOLs only appear to benefit study when item pairs are related.

**Discussion**

In Experiment 1, we showed that JOLs have a reactive effect on learning relative to a read only control condition and that this reactivity generally results in greater recall for pairs that receive JOLs at encoding. This finding lends partial support to the Changed-Goal hypothesis, as both the Positive JOL and Dual-Task hypotheses predict that reactivity will occur for all pairs that are judged at study, regardless of pair relatedness. The Changed-Goal hypothesis, however, is sensitive to the associations between item pairs and predicts that recall rates will increase for forward, backward, and symmetrical pairs while simultaneously decreasing for unrelated pairs. As shown in Experiment 1, JOLs indeed had a positive reactive effect on recall, but reactivity (regardless of direction) was only observed when study pairs were related. The expected negative reactivity did not occur for unrelated pairs. One possible explanation is that recall rates for unrelated pairs are generally lower relative to related pairs, even when participants are not making JOLs at study. Thus, correct recall of unrelated pairs may be at floor and the inclusion of JOLs at encoding has little impact on further lowering JOLs. To further investigate this, Experiment 2 introduced two additional encoding manipulations (e.g., a deep encoding task and a shallow encoding task) and tested whether these manipulations showed the same reactivity patterns as JOLs for paired associates and whether these tasks affected recall of unrelated pairs.

**Experiment 2: JOLs vs Deep and Shallow Encoding Tasks**

In Experiment 2, we tested whether the memorial benefits of JOLs are equivalent to those that are generally observed when participants are asked to engage in deeper encoding tasks at study relative to using more shallow encoding tasks (e.g., silent reading). Based on the levels-of-processing framework (Craik & Lockhart, 1972), JOLs would be expected to operate at a deeper level of encoding relative to simply reading the study pairs; thus, the memorial benefits of JOLs in Experiment 1 would be expected to occur due to the depth of the task rather than any processes that are unique to making metacognitive judgments. Thus, to provide a further test of JOL reactivity, Experiment 2 included two additional encoding conditions: A deep encoding task in which participants were asked to use relational processing at study (i.e., having participants think about how paired items are similar) and a shallow encoding task in which participants counted the number of vowels within each pair at study.

The inclusion of these additional encoding tasks allowed us to test whether JOL reactivity occurs as a byproduct of judgment making or if it is simply due to the additional processing that occurs at encoding. Specifically, the relational encoding task was chosen because it mimics some of the processes that occur when participants are asked to provide JOLs at study. For example, as illustrated by the illusion of competence, participants generally take pair relatedness into consideration when providing JOLs (i.e., pairs that are perceived as being more closely related generally receive higher JOL ratings). Based on the levels-of-processing approach, we expected that correct recall rates would increase as a function of processing depth. Thus, we expected correct recall to be highest when participants studied pairs using the deeper JOL and relational encoding tasks at study and lowest for participants completing the shallower vowel-counting task. If correct recall is elevated when participants make JOLs relative to using relational processing, then this would suggest that JOLs operate at a deeper level of processing relative to the relational task and that some of the memorial benefits of JOLs may result from processes unique to making metacognitive judgements at study (i.e., having to predict future performance and assign a value).

**Methods**

**Participants and Stimuli**

A total of XX participants were recruited to take part in Experiment 2. Participants in Experiment 2 were recruited from two sources. First, we recruited XX participants from Prolific Academic who were compensated at a rate of $9.50/hour. The remainder were XX undergraduate students recruited from University of Southern Mississippi who completed the study online for partial course credit. All participants were native English speakers with normal or corrected-to-normal vision. The same materials from Experiment 1 were used as stimuli.

**Procedure**

Experiment 2 followed the same general procedure as Experiment 1 but included the two additional encoding tasks described above. Participants completing the relational encoding task were instructed to think about how the two concepts were related to one another. The pair *cat-turtle­* was provided as an example, and participants in this condition were instructed to consider overlapping features shared between the two concepts while studying the pairs (i.e., both are animals, have four legs, and can be kept as pets). This condition differed from the read condition in which no specific instructions were given other than to read each pair silently. In the vowel counting condition, participants were instructed to count the number of vowels in the both the cue and target items. Finally, the JOL condition was identical to the one used in Experiment 1. The exact instructions used for each condition can be viewed at [OSF LINK HERE]. After viewing each pair and studying it using their respective encoding strategy, participants pressed the enter key to move to the next pair.

**Results**

Figure 2 plots mean cued-recall rates as a function of each of the four encoding strategies split by each of the four study pair types. Using a (Pair Type: Forward vs Backward vs Symmetrical vs Unrelated) × 4 (Study Condition: JOL vs Read vs Relational Encoding vs Vowel Counting) between subjects ANOVA, a main effect of Pair Type was detected in which [DESCRIBE PATTERN] [STATS]. Next, a main effect of Study Condition was detected, indicating that cued-recall rates were [PATTERN AND STATS]. [INTERACTIONS WILL GO HERE]

**Discussion**

[DISCUSS LOP PATTERN AND HOPEFULLY THAT JOLS GIVE A GREATER RECALL BOOST RELATIVE TO RELATIONAL PROCESSING] [SOMETHING HERE ABOUT HOW JOLS MAY ENGAGE RELATIONAL PROCESSING]

**Experiment 3: JOLs vs Frequency Judgments**

Given that our findings in Experiment 2 suggested that JOL reactivity was driven in part by processes unique to the judgment task, our third experiment tested whether the reactive effects of JOLs were specifically exclusive to metacognitive judgments or if other judgment paradigms could show similar memory benefits. As such, Experiment 3 compared JOLs to a novel judgment type: Word frequency judgments in which participants are tasked with rating how likely to words would appear together in natural language. The frequency judgment task was due to its similarity to the JOL task. Both tasks require participants to think about how the pairs are related (either conceptually or their use together) and assign some sort of value. Furthermore, both judgment types can be elicited using the same scale, allowing for easy comparison. A key difference between the two tasks, however, is that JOLs require participants to also think about both the relationship between the pairs and their future test performance, while frequency judgments only require participants to think about the pairs being judged.

As such, the goal of Experiment 3 was to further investigate the mechanisms driving JOL reactivity by testing whether the memorial benefit of JOLs is driven by processes that occur at judgment or if it is specifically driven by metacognitive processes that are activated when participants make specific predictions about their future performance. If frequency judgments produced a similar pattern reactivity pattern as JOLs (i.e., increased cued-recall performance for paired associates, no difference for unrelated items), then this would suggest that JOL reactivity is primarily driven by the judgment aspect of the task. Alternatively, increased correct recall in the JOL condition would indicate that JOL reactivity is driven more so by participants making predictions about their future performance.

**Methods**

**Participants**

A total of 121 participants took part in Experiment 3. All participants were recruited from the University of Southern Mississippi’s undergraduate research pool and completed the study for partial course credit. Two outliers were detected (*z*s > 3.00), leading to 119 participants included in the final dataset.

**Materials and Procedure**

Experiment 3 used the same set of materials as the previous two experiments and followed the same general procedure described in Experiment 1 with one notable exception. In addition to the JOL and No-JOL study conditions, Experiment 3 included a frequency judgment condition in which participants were asked to rate the likelihood in which the cue and target items would appear together in everyday language. The frequency judgment task utilized the same 0-100 rating scale employed by the JOL task. As with the JOL task, frequency judgments were made concurrently with study such that participants typed their ratings while the pairs were displayed on the screen. Thus, the only difference between the two judgments tasks was the focus of the judgment.

**Results**

Figure 3 reports mean recall rates as function of encoding strategy and pair direction. To test for differences in correct recall as function of encoding strategy and pair direction, Experiment 3 used a 4 (Pair Type: Forward vs Backward vs Symmetrical vs Unrelated) × 3 (Study Condition: JOL vs Frequency vs Read) ANOVA. Consistent with the two previous experiments, a main effect of Pair Type was detected, *F*(3, 348) = 590.71, *MSE* = 99.13, *ηp*2 = 0.50. Across study conditions, mean correct recall was highest for forward pairs (63.18), followed by symmetrical pairs (56.44), backward pairs (30.37), and lowest for unrelated pairs (15.87). These differences were significant across all comparisons, *t*s ≥ xx, *d*s ≥ xx. Next, an effect of Study Condition was observed, *F*(2, 116) = 6.00, *MSE* = 1205.07, *p* = .003, *ηp*2 = 0.08, such that correct recall was highest when participants made JOLs (47.11) and frequency judgments (43.32) relative to the read only study group (34.05; *t*s ≥ 2.16, *d*s ≥ XX.) We note, however, that the difference in recall rates between the JOL and frequency conditions was non-significant, *t*(79) = .79, *p* =.43. These patterns were qualified by a significant interaction between Pair Type and Study Condition, *F*(6, 348) = 12.34, *MSE* = 1205.07, *ηp*2 = 0.04, in which correct recall in both the JOL and frequency judgment conditions exceeded that of the read condition for both forward associates (JOLs = 72.57, Frequency = 66.58, Read = 50.50), *t*s ≥ 3.48, *d*s ≥ XX, and symmetrical associates (JOLs = 62.91, Frequency = 62.05, Read = 44.56), *t*s ≥ 3.56, *d*s ≥ XX. Across each of these comparisons, correct recall did not significantly differ between the JOL and frequency conditions *t*s ≤ 1.41, *p*s ≥ .16. For backward associates, correct recall did not differ between JOLs (35.44) and frequency judgments (31.28), *t*(78) = 0.66, *SEM* = 4.72, *p* = .51, or between frequency judgments and silent reading (24.44), *t*(78) = 1.36, *SEM* = 4.79, *p* = .18. However, the difference between correct recall in the JOL and the read conditions was marginally significant, *t*(79) = 1.99, *SEM* = 4.84, *p* = .051, *d* = XX. Finally, when participants studied unrelated items, recall rates were equivalent across the JOL (17.53), frequency judgment (13.34), and read (16.69) study conditions, *t*s ≤ 1.01, *p*s ≥ .32. Thus, the JOL reactivity pattern reported in the previous two experiments replicated and this reactivity pattern extended to frequency judgments.

**Discussion**

Findings from Experiment 3 were consistent with the previous two experiments. In both the JOL and study conditions, participants displayed the same patterns of correct recall as in Experiment 1. Additionally, the same reactivity pattern emerged in which JOLs had a positive reactive effect on recall but only when item pairs were related. When participants studied unrelated word pairs, JOLs had no influence on correct recall relative to the read only study condition. These findings lent partial support to the changed-goal hypothesis, though again, the lack of negative reactivity for unrelated pairs prevented this hypothesis from being fully supported. Furthermore, Experiment 3 showed that the reactive benefits of JOLs are not inherently unique to the JOL task and that they can extend to other judgment paradigms. When participants made frequency judgments at study, correct recall of paired associates increased relative to the read only study condition and mirrored the recall boosts observed in the JOL condition. Critically, correct recall of paired associates did not differ between the two judgment conditions, suggesting that both judgment types are equally beneficial to encoding. Thus, JOL reactivity appears to be driven primarily by processes that occur when making judgments, rather than those specifically related to thinking about future performance. For example, judgment making may attune participants to characteristics of item pairs that help with recall (i.e., in both a JOL and frequency judgment task, participants may become aware that some items are more likely to be remembered relative to others). Thus, reactivity may occur when participants make any type of judgment at encoding that highlights pair relatedness.

**General Discussion**

Overall, the primary goal of this study was to… [EXPAND AND SUMMARIZE EACH EXPERIMENT]

[REVISED CHANGED-GOAL HYPOTHESIS HERE]

References

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

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81* (1), 126–131.

Balota, D. A., Yap, M. J., Hutchsison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B, & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39* (3), 445-459.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Bjork, R.A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp.435–459). Cambridge, MA: MIT Press.

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14* (1), 107–111.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11* (6), 671-684.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64* (2), 119-132.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgments of learning. *Memory, 26* (6), 741-750.

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.

Garcia, M. & Kornell, N. (2015). Collector [Computer software]. Retrieved April 3rd, 2020 from https://github.com/gikeymarica/Collector.

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language, 69*, 429–444.

Hertzog, C., Dunlosky, J., Powell-Moman, A., & Kidder, D. P. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired*? Psychology and Aging, 17*, 209–225.

Janes, J. L., Rivers, M. L, & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review, 25* (6), 2356-2364.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31* (2), 187–194.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners’ sensitivity to retrieval conditions at test. *Memory & Cognition, 34* (5), 959–972.

Landauer, T. K, & Dumais, S. T. (1997). A Solution to Plato’s Problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104* (2), 211-240.

Madan, C R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63* (1), 46-63.

Maxwell, N. P., & Huff, M. J. (in press). The deceptive nature of associative word pairs: Effects of associative direction on judgments of learning. *Psychological Research*, 1-19.

Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 197-211). New York, NY, US: Oxford University Press.

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*, 530–542.

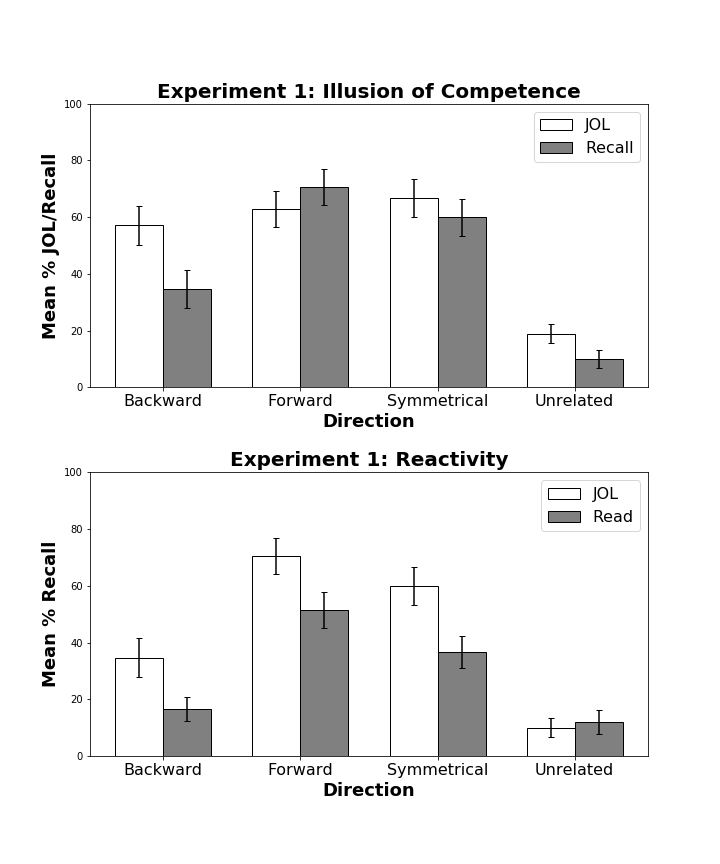
Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145* (2), 200-219.

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.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In: *The psychology of learning and motivation*, ed. G. Bower. American Psychologist.

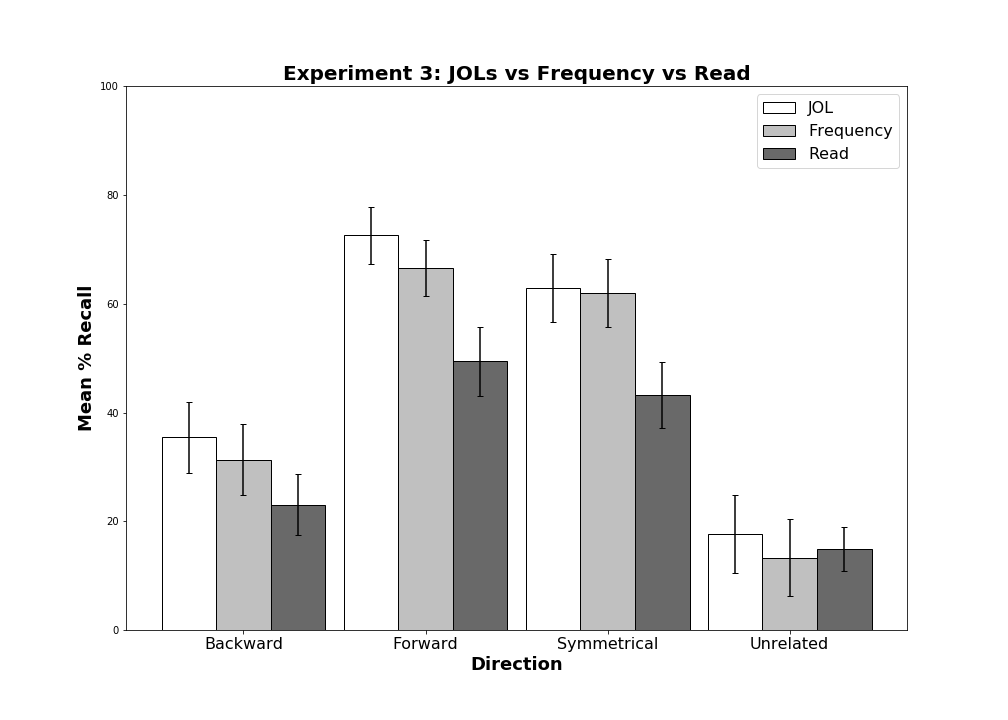
Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 553–558.

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.



*Figure 1.* Comparison of mean JOL ratings and recall rates in the JOL encoding condition (top panel) and recall rates in the JOL and Read conditions (bottom panel). Error bars represent 95% confidence intervals.

[FIGURE 2]



*Figure 3.* Comparison mean recall rates between the JOL, Frequency, and Read encoding conditions for each pair type. Error bars represent 95% confidence intervals.

Table A1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Condition | Variable | *M* | *SD* | *Min.* | *Max.* |
| Forward | FAS | .37 | .21 | .05 | .81 |
|  | BAS | .00 | .00 | .00 | .00 |
| Backward | FAS | .00 | .00 | .00 | .00 |
|  | BAS | .37 | .21 | .05 | .81 |
| Symmetrical | FAS | .19 | .13 | .01 | .46 |
|  | BAS | .19 | .13 | .02 | .52 |

*Summary Statistics for associative overlap variables.*

*Notes.* Values are grouped by associative direction. FAS and BAS values for unrelated pairs are not included as by deﬁnition these pairs have not been normed. Mean FAS and BAS values were computed by taking the average association strength for each pair.

Table A2

*Summary statistics for cue and target item properties.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Position | Variable | *M* | *SD* |
| Forward | Cue | Concreteness | 4.97 | 1.22 |
|  |  | Length | 6.20 | 1.86 |
|  |  | Frequency | 3.74 | 0.67 |
|  | Target | Concreteness | 4.96 | 1.14 |
|  |  | Length | 4.46 | 1.27 |
|  |  | Frequency | 2.49 | 0.63 |
| Backward | Cue | Concreteness | 4.96 | 1.14 |
|  |  | Length | 4.46 | 1.27 |
|  |  | Frequency | 2.49 | 0.63 |
|  | Target | Concreteness | 4.97 | 1.22 |
|  |  | Length | 6.20 | 1.86 |
|  |  | Frequency | 3.74 | 0.67 |
| Symmetrical | Cue/Target | Concreteness | 4.70 | 1.38 |
|  |  | Length | 5.21 | 1.94 |
|  |  | Frequency | 3.23 | 0.67 |
| Unrelated | Cue/Target | Concreteness | 4.63 | 128 |
|  |  | Length | 5.21 | 1.52 |
|  |  | Frequency | 2.49 | 0.85 |

*Notes.* Values are grouped by associative direction. Forward and backward pairs are grouped by position within cue-target pair. Symmetrical and unrelated pairs are averaged across cues and targets, as they did not differ by position within the pairs. Frequency is measured using SUBTLEX word frequency measure (Brysbaert & New, 2009). Concreteness and length were taken from the English Lexicon Project (Balota et al., 2007).

Table A3

*Comparison of Mean JOL Ratings and Correct Recall Percentages across all Associative Direction Pair Types for Participants in Experiment 1 who Made JOLs at Encoding.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Task | Pair Type | *M* | *95% CI* | F | B | S |
| JOL | Forward | 62.56 | 4.66 |  |  |  |
|  | Backward | 57.06 | 4.95 | 0.45 |  |  |
|  | Symmetrical | 66.66 | 4.19 | 0.25 | 0.79\* |  |
|  | Unrelated | 18.91 | 6.08 | 3.07\* | 2.60\* | 3.45\* |
| Recall | Forward | 70.62 | 6.34 |  |  |  |
|  | Backward | 34.69 | 6.80 | 2.06\* |  |  |
|  | Symmetrical | 59.87 | 6.62 | 0.62\* | 1.42\* |  |
|  | Unrelated | 10.01 | 3.25 | 4.54\* | 1.75\* | 3.61\* |

*Note.* Mean JOL and recall rates for each associative direction condition in Experiment 1’s JOL encoding condition. The three right-most columns indicate Cohen’s *d* effect sizes for post-hoc comparisons, \* = *p* < .05.

Table A4

*Comparison of Mean Recall Percentages for each Associative Direction Pair Type Across Each of the Three Experiments*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Experiment | Task | Pair Type | *M* | *95% CI* | F | B | S |
| Exp. 1 | JOL | Forward | 70.62 | 6.34 |  |  |  |
|  |  | Backward | 34.69 | 6.80 | 2.06\* |  |  |
|  |  | Symmetrical | 59.87 | 6.62 | 0.62\* | 1.42\* |  |
|  |  | Unrelated | 10.01 | 3.25 | 4.54\* | 1.75\* | 3.61\* |
|  | Read | Forward | 51.39 | 6.32 |  |  |  |
|  |  | Backward | 16.57 | 4.26 | 2.43\* |  |  |
|  |  | Symmetrical | 36.67 | 5.72 | 0.92\* | 1.50\* |  |
|  |  | Unrelated | 11.94 | 4.26 | 2.76\* | 0.41 | 1.85\* |
| Exp. 2 | JOL | Forward |  |  |  |  |  |
|  |  | Backward |  |  |  |  |  |
|  |  | Symmetrical |  |  |  |  |  |
|  |  | Unrelated |  |  |  |  |  |
|  | Relational | Forward |  |  |  |  |  |
|  |  | Backward |  |  |  |  |  |
|  |  | Symmetrical |  |  |  |  |  |
|  |  | Unrelated |  |  |  |  |  |
|  | Vowel | Forward |  |  |  |  |  |
|  |  | Backward |  |  |  |  |  |
|  |  | Symmetrical |  |  |  |  |  |
|  |  | Unrelated |  |  |  |  |  |
|  | Read | Forward |  |  |  |  |  |
|  |  | Backward |  |  |  |  |  |
|  |  | Symmetrical |  |  |  |  |  |
|  |  | Unrelated |  |  |  |  |  |
| Exp. 3 | JOL | Forward | 72.58 | 5.20 |  |  |  |
|  |  | Backward | 35.45 | 6.52 | 1.95\* |  |  |
|  |  | Symmetrical | 62.90 | 6.21 | 0.52\* | 1.33\* |  |
|  |  | Unrelated | 17.60 | 7.15 | 3.25\* | 0.80\* | 2.09\* |
|  | Frequency | Forward | 66.56 | 5.87 |  |  |  |
|  |  | Backward | 31.31 | 6.14 | 1.85\* |  |  |
|  |  | Symmetrical | 62.03 | 6.21 | 0.23 | 1.56\* |  |
|  |  | Unrelated | 13.33 | 4.06 | 3.31\* | 1.08\* | 2.91\* |
|  | Read | Forward | 49.42 | 6.29 |  |  |  |
|  |  | Backward | 23.01 | 5.60 | 1.39\* |  |  |
|  |  | Symmetrical | 43.27 | 6.06 | 0.31 | 1.09\* |  |
|  |  | Unrelated | 14.94 | 4.09 | 2.04\* | 0.52\* | 1.72\* |

*Note.* Mean recall rates for each associative direction condition for each encoding strategy. The three right-most columns indicate Cohen’s *d* effect sizes for post-hoc comparisons, \* = *p* < .05.