Reactivity from Judgments of Learning is Not Due to Memory Forecasting: Evidence from Associative Memory and Frequency Judgments

An individual’s ability to accurately monitor the progress of their own learning is a critical component for successful retention. Effective monitoring allows individuals to adjust their study strategies to maximize memory performance (Nelson & Narens, 1990) and provides insights on how best to allocate memorial resources to optimize learning (Soderstrom, Clark, Halamish, & Bjork, 2015; see also Bjork, 1999, for a review). Empirically, information about learning processes can be obtained through metacognitive judgments. Though these tasks have received significant attention from memory researchers (see Bjork, 2016; Metcalfe, 2000, for a historical overview of metamemory judgments), comparatively few studies have examined whether the act of providing metamemory judgments at study can affect subsequent performance and if so, determine the memorial processes that are affected.

A common type of judgment used to assess online metamemory processes is the judgment of learning (JOL) task. In a standard JOL task, participants study cue-target pairs (e.g., paired associates) and are asked to predict the likelihood a target will be retrieved at test if only provided with 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 typically elicited using a continuous 0 to 100 scale that represents the percent likelihood that the cue-target pair would be successfully recalled at test (e.g., 100% = definitely would remember; 0% = definitely would not remember). The use of a 100-point scale allows for a comparison between predicted recall (via JOLs) and the proportion of target items later recalled at test.

Recently, several studies have examined whether providing JOLs at study is *reactive*. JOL reactivity refers to changes in memory due to providing JOLs at encoding. A simple way to assess whether JOLs produce a reactive effect on learning is to compare recall performance for participants who complete a JOL task at study to those who do not (e.g., Janes, Rivers, & Dunlosky, 2018, Soderstrom et al., 2015). Reactivity could produce a memory benefit (i.e., *positive reactivity*) or a memory cost (i.e., *negative reactivity*) relative to a no-JOL task. While evaluating reactivity simply involves the inclusion of a no-JOL control group, this comparison is often absent in JOL studies, as researchers have either been interested in condition-specific effects on JOLs themselves rather than memory performance or have assumed that the act of providing JOLs at study has no impact on later memory. However, given that no-JOL control groups are often absent, this assumption cannot be confirmed.

The lack of no-JOL controls across studies is surprising given that early evidence for the reactive effects of JOLs on memory was documented by Arbuckle and Cuddy (1969). In one experiment, metacognitive judgments were elicited using a 1-5 Likert scale, and importantly, participants provided metamemory judgements either during both study and test phases, or only at test. Judgments at study were framed as a JOL (i.e., predicted likelihood of recalling the target in the presence of a cue at test), while judgments made at retrieval were elicited as a confidence rating (i.e., confidence that the memory response was 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 (i.e., a no-JOL control). A positive reactivity pattern was found, though it is important to note that participants in both the JOL and no-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 which contained both related and unrelated pairs. After studying each pair, one group of participants was instructed to provide JOLs, while a no-JOL group studied each pair in isolation via silent reading. Participants were then tested on their recall of the target word when presented with the cue without additional metacognitive judgments made at retrieval (cf. Arbuckle & Cuddy, 1969). Overall, target recall was 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 no-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. Furthermore, Witherby and Tauber (2017) found evidence for positive reactivity on related pairs after a 48-hour retention interval, providing evidence for positive reactivity after a delay.

In contrast to the positive reactivity for JOLs associated with related pairs reported by Soderstrom et al. (2015) and Janes et al. (2018), Mitchum, Kelley, 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. initially interpreted this discrepancy as arising from methodological differences between their study and Soderstrom et al., such as differences in experimenter-paced study and the inclusion of a generation task in their second experiment. However, in a subsequent experiment that used experimenter-paced study, Mitchum et al. again found no evidence for positive reactivity on related pairs and negative reactivity on unrelated pairs. Taken together, these studies demonstrate that providing JOLs at study can induce reactivity on target learning, but the direction of the reactivity is mixed, with positive or no reactivity reported when pairs are related and negative or no reactivity reported with unrelated pairs.

**Mechanisms of JOL Reactivity**

Several mechanisms have been proposed to account for JOL reactivity (see Mitchum et al., 2016 and Soderstrom et al., 2015). First, the *positive reactivity hypothesis* states that given monitoring is essential for determining the effectiveness of the learning process (e.g., Nelson & Narens, 1990), retention will benefit from any additional monitoring that occurs as a byproduct of providing JOLs at encoding. Because JOLs are provided for all pairs at study, a global memory improvement should occur across study materials relative to a non-JOL control. Alternatively, the *dual-task* *hypothesis* predicts that the opposite will occur, such that generating JOLs at encoding will produce negative reactivity across study materials versus a no-JOL control, since providing JOLs is resource demanding and may interfere with the learning of word pairs (Hertzog, Dunlosky, Powell-Moman & Kidder, 2002).

Next, the *changed-goal hypothesis* proposes that JOL reactivity occurs due to online changes in participant study goals that arise during encoding. According to this hypothesis, participants set an initial goal of memory mastery and strategically allocate more encoding time and/or effort towards studying items perceived as challenging to remember relative to those perceived as easy. However, certain conditions may induce a change of study goal in which easier items are prioritized. For example, Metcalfe and Kornell (2003) presented participants with English-Spanish vocabulary pairs and found that when study time was limited, participants prioritized learning of pairs perceived as “easy” due to a shared root word (i.e., cognate pairs, *park* - *parque*) versus more difficult pairs that did not contain the same root word (i.e., non-cognate pairs, *dog – perro*).When providing JOLs (specifically those utilizing a 0-100 rating scale), it becomes clear to participants that not all items will be recalled equally. Thus, participants may use perceptions of item difficulty when providing JOLs to shift their study goals towards mastering easier items.

Within the context of JOL reactivity on word pairs, the changed-goal hypothesis assumes that study lists will provide participants with at least two discernable pair types. This hypothesis predicts that providing JOLs will induce positive reactivity for pairs perceived as easy to remember, but negative reactivity for pairs perceived as difficult to remember. This is because when individuals perceive differences in difficulty between pair types, they prioritize encoding of the easier to remember related pairs at a cost of encoding more difficult unrelated pairs. Thus, for related and unrelated pairs, the changed-goal hypothesis predicts a divergent memory pattern when comparing JOL to a non-JOL group due to participant perceptions of pair difficulty.

Finally, Soderstrom et al. (2015) introduced a *cue-strengthening account,* which is based on Koriat’s (1997) cue-utilization theory. This account posits that the JOL task draws attention to certain intrinsic cues about study pairs (e.g., perceived difficulty, pair relatedness, etc.) and that reactivity will occur when those cues are made available at test. Within the context of cued-recall of word pairs, the act of making JOLs at encoding reinforces relatedness cues that are used when participants make JOLs. By further strengthening these cues, the JOL task functions akin to a generation task (e.g., Slamecka & Graf, 1978), boosting recall for pairs that receive JOLs at study. According to this account, JOL reactivity should occur whenever relatedness cues are made easily discernable (as in the case of related pairs), while no reactivity would be expected when relatedness cues are weak or nonexistent (e.g., unrelated pairs). Recent work by Myers, Rhodes, and Hausman (2020) supports this account, as they found positive reactivity on related pairs when participants completed cued-recall and recognition tests, but these patterns did not extend to free-recall testing in which relatedness cues were absent at retrieval.

Although JOL reactivity patterns based on 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 provided no support for the positive reactivity and dual-task hypotheses, showed only partial support for the changed-goal hypothesis, and fully supported a cue-strengthening account. Specifically, providing JOLs yielded a positivity effect for related target recall, but showed no reactivity on recall of unrelated targets relative to no-JOL controls. 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**

While relatedness has been shown to affect JOL reactivity, the associative direction between related word pairs has also been shown to directly influence both how well individuals recall items at test and the accuracy of JOLs made at study. Koriat and Bjork (2005; see too Koriat & Bjork, 2006) demonstrated that across three experiments, 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 compared to strong forward pair associations (e.g., lost-found). For weak forward pairs, JOLs were similar to those given to strong associates, but recall was reduced as weakly related cues were less effective in aiding target retrieval. Thus, 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; Experiment 2) also evaluated the correspondence between JOLs and target recall for pairs associated in the backward direction (e.g., card-credit). Like weak forward associates, backward associates received high JOL ratings, but recall for the target word was considerably lower than 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 (2021) showed 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 extended this pattern to symmetrical associates (e.g., off-on). Thus, the direction of association more so than the associative strength, contributes to the illusion of competence.

The illusion of competence serves as an example of how the associative direction between related pairs can affect the predictive capacity of JOLs on later recall. Regarding JOL reactivity, most studies use forward associate pairs in which the cue is highly predictive of the target. In a notable exception, Mitchum et al. (2016, Experiment 1), compared target recall using forward associates, backward associates, and unrelated pairs that were presented within the same study list. Study latencies were also measured. As reported above, no reactivity was found for either backward or forward paired-associates. Yet, despite this null pattern, the authors concluded that the changed-goal hypothesis was partially supported, as JOL participants spent less time studying unrelated pairs, which suggested that related pairs were being prioritized with additional study time.

Although Mitchum et al. (2016) showed reactivity results that were inconsistent with findings from other JOL reactivity studies (e.g., Janes et al., 2018; Soderstrom et al., 2015), it is worth noting an additional inconsistency in their data—no illusion of competence pattern emerged for backward pairs (cf. Castel et al., 2007; Koriat & Bjork, 2005; Maxwell & Huff, 2021). Though Mitchum et al. reported reduced recall rates for backward than forward pairs across JOL and non-JOL groups, these differences were much smaller than those typically reported, as participants had high percentages of correct recall on both backward and unrelated pairs. This discrepancy may have resulted from how association was measured across these studies. Koriat and Bjork (2005) 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). Maxwell and Huff (2021) similarly utilized the USF norms, as in Mitchum et al., and used pairs that were identical in associative strength (0.37 in both studies); however, a robust illusion of competence pattern was found.

A further possibility for this discrepancy is that while the 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 of cues and targets that may have covaried across pair types. Characteristics such as word length, frequency, and concreteness have each been shown to affect later recall (Balota & Neely, 1980; Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010) and could be confounded with associative direction in these studies. Thus, given discrepancies in recall that occur due to pair direction (i.e., the illusion of competence), it remains unclear whether pair direction could moderate JOL reactivity (i.e., greater reactivity for forward vs. backward pairs).

Therefore, a goal of the present study was to examine pair associations as a means of testing potential mechanisms that contribute to JOL reactivity. First, Experiment 1 was designed to provide a replication of JOL reactivity patterns reported by Janes et al. (2018) and Soderstrom et al. (2015) to further test the reliability of positive reactivity for related pairs and no reactivity for unrelated pairs while controlling for lexical and semantic characteristics of cues and targets. Furthermore, we compared reactivity effects on four different pair types, including three types of related pairs (forward, backward, and symmetrical) and unrelated pairs.

Next, Experiments 2 and 3 evaluated whether JOL reactivity effects are due to the memorial forecasting that occurs when providing a JOL or due to rating cue-target pairs within the same context, which could encourage relational encoding. This set of experiments compared recall in the JOL and no-JOL groups to a group that completed either the judgment of associative memory task (JAM; Experiment 2) or a frequency of co-occurrence judgment task (Experiment 3). The JAM task was utilized because it is a relational-encoding task (i.e., emphasizing the relations between cue-target pairs) that utilizes a similar rating process as JOLs, whereas the frequency task was designed to mimic this rating process while placing less emphasis on associations between the cue and target. In doing so, both experiments allowed participants to provide ratings while removing the memorial forecasting component associated with JOLs.

Finally, given that previous research has shown JOL reactivity to be contingent upon pair relatedness, Experiment 4 tested the strategic nature of this effect. As evidenced by Soderstrom et al., 2015 and others (e.g., Janes et al., 2018; Myers et al., 2020), when participants are exposed to related and unrelated pairs, reactivity only emerges for related pairs. Because metacognitive processes are thought to operate strategically (see Nelson & Narens, 1990), it is assumed that this pattern occurs because participants selectively emphasize processing of related (but not unrelated) pairs at encoding, leading to their greater recall. To test this assumption, Experiment 4 compared target recall in JOL and no-JOL groups relative to a relational-encoding group in which participants were explicitly instructed to relate all cue-target pairs together. In this latter group, relational encoding is a non-strategic task, as participants are instructed to use relational encoding on all pair types rather than choosing to use relational encoding on different subsets. Thus, Experiment 4 allowed for the comparison of relational encoding that is applied selectively (via JOLs) to relational encoding that is explicitly directed towards all pairs.

To preview, across experiments, we found reliable positive JOL reactivity for all three related pair types, consistent with the general pattern in the literature (cf. Double et al., 2018). We then show that both JAMs and frequency judgments elicit identical patterns of reactivity as JOLs by boosting correct recall of only related pairs, suggesting that participants strategically allocate relational processing to related pairs, even when memory forecasting is not used. Finally, we found that the benefit to related pairs when participants make JOLs is equivalent to the benefit related pairs receive when studied using an explicit relational encoding task, suggesting that when participants provide JOLs, they deploy relational encoding for related, but not unrelated pairs. Collectively, our experiments reveal that reactivity patterns are not unique to JOLs and reflect strategic use of relational encoding directed towards related pairs.

**Experiment 1: JOL Reactivity on Related and Unrelated Pairs**

The purpose of Experiment 1 was to replicate and extend previous JOL reactivity patterns by comparing target recall following study of related and unrelated pairs. The changed-goal hypothesis predicts that JOL reactivity should produce a benefit to related pairs and a cost to unrelated pairs as participants shift their study goals to prioritize the easier related pairs over unrelated pairs. Alternatively, the cue-strengthening account predicts that JOLs will produce a positive benefit to related pairs, but that no reactivity would occur for unrelated pairs. Given that prior studies have generally only shown positive reactivity for related pairs and no effect on unrelated pairs (e.g., Double et al., 2018), we expected that this pattern of reactivity would emerge, and thus we expected our findings would follow predictions from the cue-strengthening account.

An additional goal of Experiment 1 was to evaluate positive reactivity effects across different types of related pairs. We therefore compared forward and backward pairs, but also included symmetrical pairs—a related pair type that has not yet been tested in reactivity experiments. We expected that positive reactivity would be found across all three related pairs despite differences in recall rates (Maxwell & Huff, 2021). Importantly, we also controlled for lexical and semantic item effects that were not equated for across pair types in previous studies (e.g., Janes et al., 2018; Soderstrom et al., 2015). All related and unrelated pairs were matched on word frequency, concreteness, and length and related pairs were further matched on associative strength. Thus, Experiment 1 provides a more precise test of JOL reactivity patterns while controlling for important lexical and semantic item effects.

**Methods**

**Participants**

Seventy-eight participants were recruited online through Prolific (www.prolific.co) and were compensated at a rate of $8.00/hour. Participants were randomly assigned to either the JOL or no-JOL group (39 per group). A sensitivity analysis conducted with G\*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that this sample size provided adequate power (0.80) to detect medium-sized main effects/interactions (Cohen’s *d* = 0.41) or larger. All participants were native English speakers with normal or corrected-to-normal vision who had obtained at least a high school education or equivalent.

**Materials**

All stimuli pairs were taken from Maxwell and Huff (2021). These pairs consisted of 180-word pairs generated from the University of South Florida Free Association Norms (Nelson et al., 2004). Pairs were split into four types consisting 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 distributed across two study lists of 90 items which were used in two separate study/test blocks. Thus, each list contained 20 items of each of the four pair types and 10 buffer items. Pairs are available at https://osf.io/8yvn3/.

Study lists were created such that the 80 tested pairs were always proceeded and followed by five buffer pairs to reduce primacy and recency effects. Additionally, lists were constructed such that pair types were equated on frequency (SUBTLEX; Brysbaert & New, 2009), word length, and concreteness (from the English Lexicon Project; Balota et al., 2007), and related pair types were further equated associative strength (e.g., FAS and BAS values derived from the Nelson et al. (2004) free association norms; see Tables A1-A2 in the Appendix for associative strength and lexical properties for each pair type). Finally, counterbalanced versions of each study list were created that flipped the order of words with each of the four pair types (i.e., king-queen becomes queen-king). While the order within pairs was switched across all pair types, this was especially important for forward and backward pair types given forward pairs were transformed to backward pairs, making these pair types perfect controls. Study pairs were presented in a randomized order. The cued-recall test was generated from all 80 cue items (excluding buffers) by replacing the target item with a question mark (i.e., credit - ?). Test items were presented in a newly randomized 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 groups, 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 group received further instruction 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 began the first study list. Study was self-paced, with both groups pressing the Enter key to advance to the next pair. Additionally, participants in the JOL group were asked to type a JOL rating before advancing to the next study pair. JOL instructions reminding participants to use the full 0 to 100 scale were presented on screen for each trial, and all ratings were provided concurrently with study such that ratings were typed while the pair was displayed.

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 advance to the next pair. Following the first cued-recall test, participants began the second block, which followed the format of the first. Participants were fully debriefed following completion of the second cued-recall test. Each experimental session lasted approximately 30 minutes.

**Results**

A *p* < .05 significance level was used for all analyses. Partial eta-squared (*ηp*2)and Cohen’s *d* effect sizes are reported for all significant analyses of variance (ANOVAs) and *t*-tests. For all *t*-tests, we report standard test statistics, but note that all comparisons hold when using a Bonferroni correction. Additionally, for all non-significant main effects and post-hoc comparisons, we report a Bayesian estimate of the strength of the evidence supporting the null hypothesis (Masson, 2011; Wagenmakers, 2007). This analysis compares two models, one in which a significant effect is assumed, and one that assumes a null effect. From this analysis, a probability estimate is generated, a *p*-value termed *p*BIC (Bayesian Information Criterion), which estimates the probability that the null hypothesis is retained. This estimate is sensitive to the sample size, providing increased confidence in null effects reported. For completeness, encoding durations for experimental groups as a function of pair types are reported in our Supplemental Materials with data available on our OSF page (https://osf.io/xq375/).

Figure 1 plots mean recall rates for participants who made JOLs at study versus those who silently read pairs at study. A liberal scoring criterion was adopted for recall such that misspellings and grammatical errors (i.e., changes in tense) were counted as correct. All comparisons between JOL ratings and correct recall proportions for each pair type are displayed in Appendix Table A3. All analyses have been collapsed across block order1. In our analyses, we first test for an illusion of competence pattern in the JOL group, given this pattern has not been reported consistently in JOL reactivity studies (cf. Mitchum et al., 2016). These analyses were conducted across all experiments, and each demonstrated reliable illusion of competence patterns for backward associates that were consistent with previous findings (i.e., JOLs overpredicted recall of this pair type; Koriat & Bjork, 2005; Maxwell & Huff, 2021). We then test for reactivity patterns across pair types by comparing the JOL and no-JOL groups. Analyses testing for the illusion of competence for all experiments are reported in the Appendix, and all comparisons between correct recall proportions for JOL and no-JOL groups are reported in Table A4.

We tested JOL reactivity patterns by comparing the pair types across study groups using a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 2 (Study Group: JOL vs. No-JOL) mixed ANOVA. A main effect of Pair Type was found, *F*(3, 228) = 512.24, *MSE* = 75.53, *ηp*2 = .87, indicating that across study groups, correct recall was greatest for forward pairs (58.69), followed by symmetrical pairs (46.89), backward pairs (23.88), and unrelated pairs (9.26). Post-hoc *t*-tests indicated that all comparisons differed significantly, *t*s ≥ 7.79, *d*s ≥ 1.27. An effect of Study Group was also found, *F*(1, 76) = 26.01, *MSE* = 623.74, *ηp*2 = .26, in which correct recall in the JOL group (41.89) exceeded the no-JOL group (27.47), indicating an overall JOL reactivity pattern. Importantly however, a significant interaction was found, *F*(3, 228) = 28.71, *MSE* = 75.53, *ηp*2 = .27, and post-hoc tests indicated that positive reactivity was confined to related pairs. Correct recall in the JOL group exceeded that of the no-JOL group for forward pairs (69.29 vs. 48.07), symmetrical pairs (57.78 vs. 36.03), and backward pairs (31.67 vs. 16.09), *t*s ≥ 4.90, *d*s ≥ 1.11. However, for unrelated pairs (8.85 vs. 9.68), no reactivity was found, *t* < 1, *p*BIC = .88. Thus, JOLs only appear to benefit cued-recall performance when item pairs are related.

**Discussion**

The results from Experiment 1 are quite clear. Providing JOLs at study greatly increased correct recall of targets for forward, backward, and symmetrical related pairs relative to a no-JOL control. For unrelated pairs, however, providing JOLs had no effect on later recall compared to the no-JOL group. The finding that JOL reactivity effects on related pairs generalize to different types of directional associates that are matched on several lexical and semantic characteristics indicates that JOL reactivity effects occur for related pairs more broadly and are not specific to one associative direction. The JOL reactivity pattern is therefore consistent with patterns reported in other reactivity studies (Double et al., 2018; Janes et al., 2018; Soderstrom et al., 2015), who showed positive JOL reactivity for forward but not unrelated pairs.

The finding that positive reactivity effects are consistently found for related pairs but that negative reactivity is not found for unrelated pairs is inconsistent with a changed-goals account (e.g., Mitchum et al., 2016). As demonstrated in Experiment 1, related pairs, regardless of their associative direction, are prioritized at encoding and thus receive a recall boost. Given this pattern, it is likely that participants are strategically processing related pairs over unrelated pairs leading to only a memory benefit for related pairs. Given the associative relations between the cue and target for related pairs, we argue that JOLs may encourage participants to engage in relational encoding at study, such that participants emphasize shared features or characteristics of a study set (Einstein & Hunt, 1980; Hunt & Einstein, 1981). Because JOLs only produce a recall benefit for related pairs, we suggest that this relational processing is being applied *strategically* based on participant’s perceptions of association. This notion is complimentary to previous research on JOL reactivity conducted by Soderstrom et al. (2015), who proposed that JOLs were reactive because they strengthened cues used at retrieval (e.g., pair relatedness). Though they made no explicit claims regarding the strategic nature of any JOL induced relational encoding, previous work on metacognition (e.g., Nelson & Narens, 1990) has already proposed that metacognitive processes operate in a strategic manner. Therefore, our findings in Experiment 1 provide further support for Soderstrom et al.’s (2015) account while simultaneously providing additional evidence for strategy use regarding reactivity.

Because JOL reactivity appears to be largely driven by relational encoding, it may be the case that other judgment tasks that also encourage relational processing at encoding would produce similar reactivity patterns. While the literature on JOL reactivity has recently experienced an increased focus, to date, no work investigating JOL reactivity effects has explicitly tested whether observed reactivity effects are unique to JOLs or if they can occur extend to other, non-metacognitive judgment paradigms. Because Soderstrom et al.’s (2015) cue-strengthening account predicts that reactivity will occur anytime the encoding task strengthens relatedness cues between study pairs, reactivity should be expected to occur anytime this criterion is met at encoding, regardless of whether participants are engaging in metacognitive processes. Therefore, Experiment 2 openly tested this possibility by comparing JOLs to the judgment of associative memory task (JAM; Maki, 2007; Valentine & Buchanan, 2013). Like JOLs, JAMs encourage participants to attend to the relatedness between items within cue-target pairs. However, unlike JOLs, JAMs do not require participants to make memorial predictions at encoding. Therefore, the goal of Experiment 2 was to test whether the metacognitive aspects of JOLs were a requisite for reactivity to occur and whether this reactivity pattern would still emerge when the predictive component of JOLs was removed.

**Experiment 2: JOLs versus Judgments of Associative Memory**

The goal of Experiment 2 was to test whether JOL reactivity patterns could be induced when participants engage in other, non-predicative judgment tasks at encoding. In doing so, we compared JOL reactivity effects to a JAM task. In the JAM task, participants are presented with a cue-target pair and are asked to estimate the percent likelihood that an individual would respond to the cue with the presented target (Garskof, & Forrester, 1966; Nelson, Dyrdal, & Goodmon, 2005; see Maki, 2007, for a review). These estimates are typically framed as predicting the number of individuals out of 100 who would respond to the cue item with the presented target. In doing so, the JAM task is heavily dependent upon relational cues, as it gauges perceived associations between cue-target pairs. Thus, like JOLs, JAMs should encourage relational encoding, and this encoding may be strategically applied to related pairs as participants are not given explicit relational encoding instructions.

By encouraging participants to process the cue and target together, this task was designed to mimic the processing used by the JOL task. We elected to use JAMs due to this task’s similarity to JOLs, as both require participants to process related aspects of the study pairs (either conceptually or their use together) and assign a judgment value. Further, ratings on both tasks are provided using the same scale, allowing for easy comparison. If participants are using relational encoding strategically on related word pairs, they would be able to use this encoding on both the JOL and JAM tasks. Of course, a key difference between the two tasks is that JOLs require participants to predict later recall at encoding, whereas JAMs do not. Thus, an interesting question regarding JOL reactivity is whether memory predictions are necessary to produce a memory improvement. Because JOL reactivity appears to be driven primarily by selective relational encoding, only the use of relational encoding given to pairs at study will benefit memory, not necessarily whether a memory prediction is made. Thus, we expected memory forecasting via JOLs would not be necessary to produce reactivity effects.

**Methods**

**Participants**

70 participants were recruited from The University of Southern Mississippi’s undergraduate research pool and completed the study online for partial course credit. Additionally, 28 participants were recruited from Prolific and completed the study at a rate of $8.00/hour, leading to a total of 98 participants who completed Experiment 22. Participants were randomly assigned to either the JOL group (*n* = 33), the no-JOL group (*n* = 32), or the JAM group (*n* = 33). A sensitivity analysis conducted using *G\*Power 3* indicated that this sample size provided adequate power (0.80) to detect medium-sized main effects/interactions (Cohen’s *d* = 0.50) or larger. All participants were native English speakers who reported normal or corrected-to-normal vision.

**Materials and Procedure**

Experiment 2 used the same materials and followed the same general proceduredescribed in Experiment 1 with the following exception. In addition to standard JOL and no-JOL groups, participants were also randomly assigned to a JAM task group in which they were asked to rate the likelihood in which the target word would be given as a response to the cue. Like JOLs, JAM ratings were elicited using a continuous 0-100 scale. JAM instructions were modeled after the associative judgment task used by Maxwell & Buchanan (2020; exact instructions are available at https://osf.io/6xgkt/). Specifically, JAMs were framed as the number of individuals out of 100 who would respond with the target word if shown only the cue. As with the JOL task, JAMs were elicited concurrently with study, and study was self-paced across all groups. Thus, only the focal point of the two judgments differed.

**Results**

Figure 2 plots mean recall as function of encoding group and pair type. To test for reactivity effects, we conducted a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 3 (Study Group: JOL vs. JAM vs. No-JOL) mixed ANOVA on correct recall. An effect of Pair Type was found, *F*(3, 285) = 616.18, *MSE* = 81.46, *ηp*2 = .60, in which correct recall was highest for forward pairs (64.92), followed by symmetrical pairs (56.22), backward pairs (33.16), and lowest for unrelated pairs (14.82). All comparisons differed significantly, *t*s ≥ 8.08, *d*s ≥ 0.45. Next, an effect Study Group was found, *F*(2, 95) = 3.90, *MSE* = 827.92, *ηp*2 = .06, in which correct recall was highest when participants made JOLs (45.36) and JAMs (44.85) relative to participants in the no-JOL control group (36.46). All comparisons differed statistically, *t*s ≥ 2.28, *d*s ≥ 0.57, with the exception of the comparison between the JOL and JAM groups, *t* < 1, *SEM* = 3.57, *p* = .88, *pBIC* = .88.

Importantly, a significant interaction between Pair Type and Study Group emerged, *F*(6, 285) = 9.82, *MSE* = 81.46, *ηp*2 = .04. Follow-up *t*-tests revealed that for forward pairs, correct recall in both the JOL (71.74) and JAM (67.58) groups exceeded that of the no-JOL group (55.16). Comparisons across all statistically differed, *t*s ≥ 2.93, *d*s ≥ 0.65, with the exception of the JOL and JAM groups, which did not differ statistically, *t* < 1, *p*bic = .84. A similar pattern was observed for symmetrical pairs. Correct recall was greater for the JOL (60.68) and JAM (61.29) groups versus the no-JOL group (46.41). Again, all comparisons statistically differed *t*s ≥ 3.22, *d*s ≥ 0.80, except between the JOL and JAM groups, *t* < 1, *p*bic = .87. For backward pairs correct recall in the JOL (35.61) and JAM (36.36) groups was again greater relative to the no-JOL group (27.34). Correct recall in the JAM and the No-JOL groups differed significantly, *t*(63) = 2.11, *SEM* = 4.35, *d* = 0.52, while the comparison between the JOL and the No-JOL groups was marginal, *t*(63) = 1.93, *SEM* = 4.37, *p* *=* .06, *p*bic = .56, *d* = 0.48. Recall did not differ between the JOL and JAM groups, *t*< 1, *p*bic = .88. Finally, for unrelated pairs, recall rates were statistically equivalent across the JOL (13.41), JAM (14.68), and no-JOL (16.95) groups, *t*s ≤ 1.23, *p*s ≥ .22, *p*bics ≥ .79. Taken together, both JOL and JAM tasks resulted in equivalent reactivity on correct recall for related pairs and no reactivity on unrelated pairs.

**Discussion**

The goal of Experiment 2 was to examine whether JOL reactivity pattens would extend to other non-predictive judgment tasks by comparing the standard JOL task to a JAM task. In both tasks, participants processed the cue-target relations prior to providing a judgment using the same 0-100 scale. Although the judgment type differs (recall forecasting vs. relatedness estimates), the reactivity patterns observed for related and unrelated pairs did not differ, suggesting that similar processing occurred between the two task types. Compared to the no-JOL control group, both the JOL and JAM groups showed increased correct recall of targets across forward, backward, and symmetrical pairs—a positive reactivity pattern, but produced no recall benefit on unrelated targets.

The similarity in recall rates between the JOL and JAM groups yields several important findings regarding reactivity effects in recall of cue-target pairs. First, similar reactivity patterns observed for the JOL and JAM tasks indicate that type of task employed at encoding may not be a critical factor as to whether or not a reactivity pattern emerges. Instead, the qualitative processing given to the cue and target by the task may be more impactful. Second, providing a memory prediction does not appear to be a requisite for positive reactivity on related pairs given the similarity between the JOL and JAM groups. This finding is important in reference to other studies that have reported JOL reactivity patterns (e.g., Soderstrom et al., 2015; Mitchum et al., 2016) which have only compared JOL and no-JOL groups and have not measured recall differences relative to additional, non-JOL encoding tasks. Finally, the finding that reactivity does not operate globally across all pair types (regardless of judgment task) further suggests that reactivity processes are applied strategically, with an emphasis on related over unrelated pairs.

While the JAM task does not explicitly instruct participants to relate study pairs together at encoding, relatedness is still a focal point of this task as participants are required to estimate the association strength between two words. Because of this, JAMs may be more likely to induce relational encoding relative to JOLs. As such, a stronger test of whether JOL reactivity extends to other encoding tasks would be to compare JOLs to a judgment task that less overtly directs attention to the relational characteristics between items. To this end, Experiment 3 introduced a frequency of co-occurrence judgment task in which participants were instructed to rate the likelihood that two words would be used together in everyday language. Like JAMs, frequency judgments emphasize the correspondence between cues and targets, but do not explicitly instruct participant to relate items together at encoding. However, relative to JAMs, the frequency judgment task places less emphasis on pair relatedness, providing an encoding task that is more comparable to JOLs.

**Experiment 3: JOLs versus Frequency Judgments**

The primary goal of Experiment 3 was to provide an additional test of whether JOL reactivity effects found with JOLs extend to other encoding tasks by comparing JOLs to a frequency of co-occurrence judgment task. In this task, participants are asked to estimate the likelihood that the cue and target words would appear together contextually within the English language. We note that while the frequency task is still sensitive to pair relatedness, unlike the JAM task, it is less likely to explicitly direct participants to process pair relations. Overall, we expected that any observed reactivity would adhere to the patterns previously reported in Experiments 1 and 2. Specifically, we anticipated that the JOL group would again show positive reactivity for related pairs (forward, backward, and symmetrical), and recall would not differ on unrelated pairs relative to a no-JOL control. Furthermore, consistent with findings for JAMs in Experiment 2, we also expected that this pattern of reactivity would extend to the frequency judgment group, such that positive reactivity would be observed for related, but not unrelated pairs. Finally, we expected that any reactivity patterns observed for frequency judgments would be equivalent to the JOL group due to relational encoding of related pairs being fostered by both tasks.

**Methods**

**Participants**

A total of 118 participants completed Experiment 3 and were randomly assigned to either the JOL group (*n* = 40), the no-JOL group (*n* = 39), or the frequency judgment group (*n* = 39). A sensitivity analysis conducted with *G\*Power 3* indicated that this sample size provided adequate power (0.80) to detect medium main effects/interactions (Cohen’s *d* = 0.45) or larger. All participants were recruited from The University of Southern Mississippi’s undergraduate research pool and completed the study online in exchange for partial course credit. Participants were all native English speakers and reported normal or corrected-to-normal vision.

**Materials and Procedure**

Experiment 3 used the same materials and followed the general procedure of Experiment 1 with one exception. In addition to the JOL and no-JOL groups, Experiment 3 included a frequency-judgment group 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, with higher ratings corresponding to more frequent occurrences. Both JOLs and 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 tasks was the focus of the judgment.

**Results**

Figure 3 reports mean recall rates as function of encoding group and pair type. We conducted a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 3 (Study Group: JOL vs. Frequency vs. No-JOL) ANOVA to evaluate reactivity effects. First, an effect of Pair Type was detected, *F*(3, 348) = 590.71, *MSE* = 99.13, *ηp*2 = 0.84, indicating that correct recall was highest for forward pairs (62.94), followed by symmetrical pairs (56.13), backward pairs (29.97), and lowest for unrelated pairs (15.31). Differences were significant across all comparisons, *t*s ≥ 10.80, *d*s ≥ 0.79. An effect Study Group was also found, *F*(2, 116) = 6.00, *MSE* = 1205.07, *p* = .003, *ηp*2 = .12, indicating that correct recall was highest when participants made JOLs (47.13) and frequency judgments (43.30) relative to the no-JOL control group (32.66). All comparisons were significant, *t*s ≥ 2.97, *d*s ≥ 0.67, except for the JOL and frequency groups, *t* < 1, *p*bic = .86.

Critically, a significant interaction was found, *F*(6, 348) = 12.34, *MSE* = 1205.07, *ηp*2 = .17. Follow-up tests indicated that for forward pairs, correct recall in both the JOL (72.57) and frequency judgment (66.58) groups exceeded that of the no-JOL group (49.42). All comparisons differed, *t*s ≥ 3.91, *d*s ≥ 0.88, except for the JOL and frequency judgment groups, *t*(76) = 1.50, *SEM* = 4.07, *p* *=* .14, *p*bic = .74. Symmetrical pairs displayed a similar pattern. Recall was greater in the JOL (62.91) and frequency judgement (62.05) groups relative to the no-JOL group (43.27), and again, all comparisons differed *t*s ≥ 4.23, *d*s ≥ 0.96, with the exception of the JOL and frequency judgment groups, *t* < 1, *p*bic = .85. For backward pairs, correct recall in the JOL (35.44) and frequency judgment (31.23) groups were greater than the no-JOL group (23.01). All comparisons differed significantly, *t*s ≥ 1.96, *p*s < .05, except for the JOL and frequency judgment group, which did not differ, *t* < 1, *p*bic = .90. Finally, for unrelated pairs, recall rates were equivalent across the JOL (17.53), frequency judgment (13.34), and no-JOL (14.94) groups, *t*s ≤ 1.02, *p*s ≥ .31, *p*bic ≥ .88. Thus, both JOL ratings and frequency judgments produced equivalent reactivity on correct recall for related pairs but no reactivity on unrelated pairs.

**Discussion**

The primary goal of Experiment 3 was to provide an additional test of whether JOL reactivity patterns could be produced by other, non-metacognitive encoding tasks. Specifically, we assessed whether reactivity patterns observed for JOLs and JAMs in Experiment 2 would replicate when participants engaged in a frequency judgment task at encoding. We selected the frequency judgment task because it provided a closer comparison to the JOL task by reducing the emphasis on pair relatedness that is inherent to JAMs. Consistent with Experiment 2, reactivity patterns emerged for both JOLs and frequency judgments. Relative to the no-JOL group, participants completing either the JOL or frequency judgment task at encoding showed increased correct recall for each of the three types of related pairs. These tasks, however, produced no reactivity when participants studied unrelated pairs, indicating that reactivity effects operated selectively as a function of pair relatedness. Importantly, frequency judgments produced reactivity patterns that were comparable to those observed for JAMs in Experiment 2, providing further evidence that memory forecasting is not a requirement for reactivity to occur.

Experiments 2 and 3 showed that JOL reactivity patterns can be reproduced using other, non-metacognitive judgment tasks, as both JAMs and frequency judgments each selectively boosted recall of related pairs relative to unrelated pairs, mimicking previously observed JOL reactivity patterns (e.g., Janes et al., 2018; Soderstrom et al., 2015). Although Soderstrom et al. (2015) did not make explicit claims regarding the strategic nature of JOL reactivity, it is assumed that this pattern emerges because the JOL task selectively emphasizes the processing of related pairs over unrelated pairs. To test this possibility, Experiment 4 compared JOLs to an explicit relational encoding task in which participants were instructed to relate all pairs together at study, regardless of relatedness. In doing so, Experiment 4 provided a test of this strategy use account by comparing JOL reactivity, which may operate strategically, to an explicit relational encoding task that is globally applied to all pair types and, therefore, does not operate strategically.

**Experiment 4: JOLs versus Relational Encoding**

In Experiment 4 we tested whether positive reactivity found for related pairs following JOLs versus no-JOLs was due to the strategic use of relational processing at encoding. We tested this possibility by comparing standard JOL and no-JOL groups to a relational-encoding group which was given intentional encoding instructions to relate all pairs together at study. We reasoned that if the JOL group employs relational encoding strategically on related pairs leading to reactivity, then this pattern of reactivity should be equivalent to related pair recall rates for participants who are engaging in explicit relational encoding at study. Furthermore, because recall is typically greater following relational encoding relative to standard read-only instructions (Huff & Bodner, 2014; 2019), we expected that recall would be increased following relational encoding instructions relative to the no-JOL group. Because our previous experiments showed that JOL reactivity selectively increased recall of related pairs relative to unrelated, only unrelated pairs encoded using the relational instructions were expected to receive a memory benefit.

The inclusion of an explicit relational encoding group was designed to contrast the strategic relational encoding processes induced by JOLs. Whereas JOLs encourage relational processing only when pairs are related, the relational encoding instructions in Experiment 4 were designed to encourage participants to apply relational encoding to all pair types, regardless of relatedness. Having participants in the relational group apply this task across all pairs (vs. a subset of related pairs) was used because explicit relational encoding instructions have been shown to spill over into other encoding tasks when encoding processes are manipulated within-subjects (Huff, Bodner, & Gretz, 2021). Given these carryover issues, it was reasonable to have participants utilize relational encoding for all pair types. In addition to the relational encoding group, we also included a shallow levels-of-processing group (i.e., vowel-counting task) to serve as an additional control. The inclusion of this group allowed us to ensure that any recall benefits found in the relational encoding group were due to relational encoding and not due to the use of an explicit encoding task.

Consistent with the previous experiments, we again expected a positive reactivity pattern for the JOL versus no-JOL group. Additionally, we expected that relational encoding would also produce a recall benefit that would mimic positive reactivity in the JOL group on related pairs, consistent with reactivity patterns observed for JOLs. However, we also expected that recall of unrelated pairs would be greater in the relational-encoding group relative to the JOL group. This is because the explicit relational task forces participants to utilize relational encoding regardless of pair type, which will likely benefit memory for unrelated pairs. Finally, we expected that the shallow group would produce lower levels of recall, possibly even lower than the no-JOL group since shallow processing is ineffective for promoting long-term memory.

**Methods**

**Participants and Stimuli**

A total of 167 participants were recruited for Experiment 4. Participants were recruited from two sources. First, we recruited 84 undergraduate psychology students from The University of Southern Mississippi who completed the study online for partial course credit. The remaining 83 participants were recruited online via Prolific and were compensated at a rate of $8.00/hour3. Participants were randomly assigned to the JOL group (*n* = 39), the no-JOL group (*n* = 40), the relational encoding group (*n* = 45), and the shallow group (*n* = 43). A sensitivity analysis conducted with *G\*Power 3* indicated that this sample size provided adequate power (0.80) to detect a medium main effects/interactions (Cohen’s *d* = 0.40) or larger. All participants were native English speakers with normal or corrected-to-normal vision.

**Materials and Procedure**

The same materials and general procedure from Experiment 1 were again used in Experiment 4, with the exception of two additional encoding tasks. Participants in the relational-encoding group 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 group 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, etc.). In the vowel-counting group, participants were instructed to report the number of vowels in both the cue and target items by typing their response into a text box. Both the relational-encoding and vowel counting groups did not provide JOL ratings at study as in the no-JOL group and were instructed to apply their encoding strategy to all study pairs. After viewing each pair and studying it using their respective encoding strategy, participants pressed the enter key to move to the next pair. Participants in the JOL and no-JOL groups followed the same procedure used in Experiment 1, and all groups completed a 2-min filler task and a cued-recall test following the study phase.

**Results**

Mean cued-recall rates for each of the four encoding strategies as function of pair type are reported in Figure 4. To examine reactivity effects across encoding tasks, we used a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 4 (Study Group: JOL vs. No-JOL vs. Relational Encoding vs. Shallow Encoding) mixed ANOVA. An effect of Pair Type, *F*(3, 489) = 691.11, *MSE* = 78.13, *ηp*2 = .81, indicated that correct recall was highest for forward pairs (52.17), followed by symmetrical pairs (42.95), backward pairs (22.28), and lowest for unrelated pairs (13.73), which all differed statistically from each other, *t*s ≥ 10.72, *d*s ≥ 0.44. A main effect of Study Group was also found, *F*(1, 163) = 10.56, *MSE* = 1166.90, *ηp*2 = .16, in which correct recall was highest in the relational encoding group (41.06), followed by the JOL group (38.61), the no-JOL group (28.11), and shallow group (23.18). Post-hoc *t*-tests indicated that cued-recall rates in the JOL and relational encoding groups differed significantly from the no-JOL and shallow groups tasks (*t*s ≥ 4.14, *d*s ≥ 0.93), but did not differ between each other, *t* < 1, *p*BIC = .88. Additionally, there was no difference between the no-JOL and shallow groups, *t*(69) = 1.48, *SEM* = 3.39, *p* = .14, *p*BIC = .76.

The effects of Pair Type and Study Group were qualified by a significant interaction, *F*(9, 489) = 13.29, *MSE* = 78.13, *ηp*2 = .03. Beginning with forward pairs, correct recall was highest in the JOL group (63.78), followed by the relational group (58.17), the no-JOL control group (48.06), and the shallow group (39.19). All comparisons differed significantly (*t*s ≥ 2.13, *d*s ≥ 0.47), with the exception of the JOL and relational groups, *t*(75) = 1.37, *SEM* = 4.18, *p* = .18, *p*BIC = .79. This same pattern was also found with symmetrical pairs: Correct recall was highest in the JOL group (54.17), followed by the relational group (50.06), the no-JOL group (38.13) and shallow group (29.83). All comparisons differed significantly, *t*s ≥ 2.06, *d*s ≥ 0.45, again with the exception of the JOL and relational groups, *t* < 1, *p*BIC = .79. For backward pairs, correct recall was highest in the relational group (30.89), followed by the JOL group (26.60), the no-JOL group (17.13), and the shallow group (14.13). Follow up *t*-tests showed that recall rates in the JOL and relational groups differed from both the no-JOL and shallow groups (*t*s ≥ 3.24, *d*s ≥ 0.77). Recall did not differ between the JOL and relational group (26.60 vs. 30.89), or between no-JOL and shallow groups (17.13 vs. 14.13), *t*s < 1, *p*s ≥ .33, *p*BICs ≥ .85. Finally, for unrelated item pairs, recall rates were highest for the relational group (25.11) relative to the JOL task (9.87), the no-JOL group (9.13), and the shallow group (9.59), *t*s ≥ 3.73, *d*s ≥ 0.74). All other comparisons were non-significant, (*t*s < 1, *p*s ≥ .73, *p*BICs ≥ .90).

**Discussion**

Experiment 4 produced two notable outcomes. First, a JOL reactivity pattern was again found in which, relative to the no-JOL group, providing JOLs increased recall for related but not unrelated targets. Second, and more importantly, the JOL reactivity pattern found for related pairs mimicked related pairs in the relational encoding group that was instructed to explicitly associate pairs together at encoding. This similarity suggests that JOL participants are engaging in deep relational encoding of related pairs despite not receiving explicit instruction to do so. Positive reactivity was similarly found when comparing the JOL and relational groups to the shallow-encoding group, indicating that reactivity effects hold relative to a shallow task. As expected, recall differed between the JOL and relational group for unrelated pairs. This pattern is likely due to relational participants employing their encoding task across all pair types as was instructed, rather than strategically limiting it to only related pairs as is likely occurring in the JOL group. Thus, these patterns provide evidence that reactivity processes are strategic in nature and are directed towards processing related over unrelated pairs.

**General Discussion**

The primary goals of this study were twofold. First, Experiment 1 sought to replicate previous work showing that JOLs produce a reactive effect on cued-recall of related targets while comparing these reactivity patterns on forward, backward, and symmetrical paired associates—a novel contribution. Second, and more importantly, Experiments 2-4 were designed to test whether reactivity patterns that have been found with JOLs can occur in other tasks that do not require memorial forecasting. In Experiment 2, we gauged JOL reactivity effects relative to the JAM task in which participants made relational, non-metacognitive frequency judgments. Next, Experiment 3 provided an additional test of whether JOL reactivity patterns generalize to other judgment tasks by comparing JOL reactivity to a frequency-judgment task. Finally, Experiment 4 compared JOL reactivity to a deep relational encoding strategy. Collectively, our results indicate that reactivity is not unique to JOLs and that enhanced relational encoding applied to related but not unrelated pairs primarily contributes to these reactivity benefits.

Results from Experiment 1 found positive JOL reactivity on forward pairs that was consistent with previous work by Soderstrom et al. (2015) and Janes et al. (2018), while extending this pattern to include backward and symmetrical pairs. Importantly, these reactivity patterns occurred using pairs that were engineered to control for lexical and semantic item effects, including associative strength that could potentially influence correct recall. The positive reactivity pattern found across each of the three related pair types indicates that the associative direction of cue-target pairs does not have an effect on reactivity. Instead, the mere presence of association is likely sufficient to facilitate additional encoding of related pairs. For unrelated pairs, however, no reactivity pattern was found as recall was equivalent between the JOL and no-JOL groups. The discrepancy in reactivity for related and unrelated pairs provides further evidence that JOLs cause participants to engage in selective relational encoding of related pair types, which is consistent with Soderstrom et al. (2015) and Myers et al. (2020).

Next, to test whether reactivity effects were unique to JOLs, Experiment 2 compared JOL and no-JOL groups to participants completing a JAM task, which required participants to provide relatedness judgments for cue-target pairs. This task was selected because, like JOLs, it allowed for processing of the relational characteristics of study pairs without explicit instruction to encode all study pairs using a relational strategy. Moreover, the JAM task utilized the same rating scale as the JOL task. The JAM task therefore resembled the JOL task but did not require that participants forecast later recall performance. This provided and novel comparison, as to date, studies investigating the reactive effects of JOLs have not compared reactivity to other, non-metacognitive judgment tasks. Overall, Experiment 2 found equivalent positive reactivity on related pairs when compared to the JOL task and critically, no reactivity was found on unrelated pairs, indicating that reactivity patterns are not exclusive to JOLs and that likely reflect use of strategic relational encoding.

Experiment 3 then compared the JOL and no-JOL groups to a frequency-judgment task in which participants were required to estimate the frequency in which the cue-target pair would co-occur in the English language. The frequency-judgment task provided a stronger test of whether JOL reactivity extended to other judgment tasks, as relative to JAMs, frequency judgments place less emphasis on the associative characteristics of cue-target pairs, making them more akin to the JOL task. Like the JAM task used in Experiment 2, frequency judgments showed the same positive reactivity on related pairs as the JOL task, and critically, no reactivity was found on unrelated pairs. The extension of this finding to frequency judgments provides further evidence that reactivity patterns are not limited to JOLs and that memory forecasting is not a requirement for reactivity to occur.

Finally, Experiment 4 compared JOLs to a relational encoding task in which participants were explicitly instructed to relate all cue-target pairs together at study. Relative to both the no-JOL control group and a group of participants completing a shallow vowel-counting task, relational encoding produced the same positive reactivity pattern on related pairs as participants who completed the JOL task. However, unlike the JOL task, positive reactivity induced by relational processing was not restricted to related targets, as recall of unrelated targets was also greater relative to the no-JOL control group. This latter pattern was unsurprising given participants were instructed to utilize relational encoding for all pair types. Finally, the shallow vowel counting task did not induce reactivity, suggesting that the qualitative aspects (i.e., deep relational processing) of the encoding task were a driving factor of reactivity rather than merely having participants engage in an additional task at study.

Across all experiments, positive reactivity consistently emerged on related pairs when participants engaged in encoding tasks that implicitly encouraged relational processing. However, negative reactivity effects on unrelated pairs as reported by Mitchum et al. (2016) continuously failed to occur, regardless of whether participants made JOLs, JAMs, or frequency judgments at encoding, a finding that is consistent with previous work on reactivity (e.g., Soderstrom et al., 2015; Janes et al., 2018). However, given that participants generally performed poorly across experiments when recalling unrelated pairs (mean recall of unrelated pairs was < 18% in the no-JOL groups across experiments), negative reactivity may not have occurred because participants’ lack of success left little room for further decreases in performance in the judgment groups. Though these levels of recall performance are in line with findings from other reactivity studies that have shown positive reactivity for related pairs (e.g., Janes et al., 2018, Soderstrom et al., 2015), we note that Mitchum et al. (2016) reported higher recall rates for unrelated pairs in their control groups, with mean correct recall for these pairs exceeding 40% across experiments. Thus, whether negative reactivity occurs on unrelated pairs may be at least partially contingent on participant performance on this pair type.

**JOLs are not a Requisite for Reactivity**

An important finding from this set of experiments is that reactivity patterns are not unique to JOLs. Because JOLs call attention to pair relatedness (which is a strong predictor of cued-recall performance; Maxwell & Buchanan, 2020), relatedness cues may become more salient relative to participants in a no-JOL control. Based on this account, reactivity would be expected to occur whenever participants engage in tasks that encourage the use of a relational strategy at encoding and when these tasks include study items that differ in their relatedness, regardless of whether participants engage in metacognitive processes at encoding. Results from Experiments 2-4 support this claim, as JAMs (Experiment 2), frequency judgments (Experiment 3), and relational encoding (Experiment 4) each produced equivalent reactivity patterns for related pairs relative to the JOL group. Furthermore, the similarity in reactivity patterns between JOLs and both JAMs and frequency judgments suggests that each task taps into similar underlying relational encoding processes. Based on Koriat’s (1997) cue-utilization framework, each judgment type tunes participants to specific *intrinsic* cues about the study pairs, providing them with information about inherent properties of the studied material (i.e., pair relatedness). Thus, cued-recall performance is enhanced whenever an encoding task draws attention to the relatedness between studied items, regardless of whether this is done explicitly (e.g., relational study instructions) or implicitly (e.g., JOLs, JAMs, frequency judgments, etc.). However, because this occurred indirectly in Experiments 2 and 3 (as neither the JOL, JAM, or frequency judgment tasks explicitly instructed participants to relate items together at study), only related items receive a memory boost when judged. As such, reactive effects are not generally observed for unrelated items unless the task explicitly instructs participants to relate all pairs together.

The findings that reactivity repeatedly occurs only when pairs are related suggests that making JOLs, JAMs, and frequency judgments, are not merely “deep” encoding tasks. Within the levels of processing framework (Craik & Lockhart, 1972), tasks that facilitate deeper processing are those that encourage participants to elaborate on characteristics of items at encoding. However, a deep encoding task in the present experiments should operate across all pair types irrespective of relatedness, as observed in the Experiment 4 relational task. The observation that JOLs do not operate globally across pair types suggests that they are not functioning as a depth of processing task. Rather, JOL reactivity is consistently moderated by pair relatedness, a pattern which was extended to both JAMs and frequency judgments. Thus, while JOLs improve retention relative to silent reading, this increase is not simply due to depth of processing but to the selective nature of the processing induced by this task.

**A Case for Strategic Relational Encoding**

Soderstrom et al. (2015) proposed that JOLs will induce reactivity when two criteria are met. First, the JOL task must strengthen cues that inform JOLs (i.e., such as pair relatedness) and second, the same cues must be available at test (i.e., such as a cued-recall test in which the desired target can be triggered by the presentation of the cue). Consistent with this account, Myers et al. (2020) showed that positive reactivity on related pairs only occurred when cues used to inform the JOL were available at test. JOLs were reactive when using cued-recall and recognition testing, but not when using free-recall testing. Myer et al.’s extension of this pattern to recognition memory but not free-recall provides support for Soderstrom et al.’s first criterion that the JOL task strengthens cue-target associations that are subsequently used at retrieval. Additionally, the present study further supports this account, and suggests that JOLs encourage participants to engage in relational encoding, which is applied selectively to pairs as a function of pair relatedness. Therefore, the present study is consistent with previous studies which have indicated that JOL reactivity is found on related pairs and further establishes that the selective use of relational processing contributes to JOL reactivity and provides further evidence supporting Soderstrom et al.’s cue strengthening account.

The strategic nature of this relational encoding is consistent with previous work on metamemory and strategy use. For example, in their metamemory framework, Nelson and Narens (1990) posited that participants can adjust their encoding strategies based on cues inherent to the stimuli as participants monitor their study. Because pair relatedness is a salient cue of future recall performance, it is likely that these relatedness cues may trigger changes in study strategies. Thus, only related pairs are processed using a relational encoding strategy, as participants modify their study strategy whenever they encounter this pair type. This results in a memory boost for related items that receive additional relational processing at encoding while unrelated pairs show no benefit.

While our findings can be explained via a strategy use account, enhanced recall of related pairs may also reflect greater processing of related pairs due to intrinsic relatedness cues being used as a basis for JOLs, rather than participants strategically altering their study strategies as a function of relatedness. For related pairs, participants must distinguish between differing levels of relatedness when assigning a JOL (e.g., weak vs strong paired associates), which in turn leads to enhanced relational encoding of this pair type. However, because relatedness cues are absent for unrelated pairs, participants may simply assign them a low JOL by default, rather than attempting to discern between varying levels of relatedness. This lack of additional processing of unrelated pairs results in only related pairs showing a memory improvement at test.

Two methods can be used to assess this possibility: Assessing changes in standard deviation of judgments as a function of pair relatedness and assessing differences in encoding durations between related and unrelated pairs. First, if participants are defaulting to a low JOL for unrelated pairs, one may predict that variability would be relatively low for unrelated pairs relative to related pairs. Somewhat consistent with this possibility, JOL standard deviations were lower for unrelated pairs relative to each of the related pair types in three of the four experiments (Experiments 1, 3, and 4). Second, if low JOLs are a default response for unrelated pairs, latencies to provide JOLs for unrelated pairs should be faster relative to related pairs. Analyses of JOL latencies were mixed however, as latencies for unrelated pairs were sometimes faster than related pairs (Experiments 1 and 2), but sometimes not (Experiments 3 and 4; see Supplemental Materials for both standard deviations of judgment values and JOL latency analyses). Thus, these secondary analyses do not provide a clear pattern and cannot discern whether providing JOLs informs strategy use at encoding.

Relatedly, reactivity effects may partially represent increased encoding durations for participants who completed judgment tasks at study relative to silent reading. Though encoding was self-paced in the present study, previous research has often attempted to control for this via experimenter paced study (e.g., Janes et al., 2018; Soderstrom et al., 2015). These studies, however, have repeatedly shown that reactivity effects still emerge even after encoding durations are held constant between JOL and no-JOL groups. Further, Janes et al. (2018) showed that positive reactivity effects on related pairs only emerged when experimenter pacing was used. Therefore, although the self-paced encoding used in the present study resulted in longer encoding durations for participants in the judgment groups (see Supplemental Materials for all RT analyses), these differences likely reflected participants in the judgment groups multi-tasking the encoding of each pair with simultaneously completing their respective judgment task. Consistent with this notion, reactivity effects repeatedly failed to emerge for unrelated pairs, even though across experiments, participants in the judgment groups spent significantly longer encoding this pair type compared to participants in the control groups. Collectively, reactivity effects in the present study appear to be driven primarily by relational encoding rather than increased study durations.

Additionally, although prior research on JOL reactivity has largely suggested that relatedness cues are a primary factor driving reactivity effects, recent work conducted by Senkova and Otani (2021) proposes that JOL reactivity effects are not due to the use of relational encoding and instead reflect the effects of item-specific processing. According to this account, JOLs modify memory by calling attention to the item and modifying its distinctiveness. While Senkova and Otani showed that recall following JOLs was equivalent to recall for item-specific processing tasks (i.e., ratings of pleasantness and imagery), we note one methodological discrepancy between their study and the present that may account for this. Whereas the majority of studies investigating JOL reactivity have tested for these effects using mixed lists of related and unrelated word pairs (e.g., Janes et al., 2018; Soderstrom et al., 2015), Senkova and Otani instead had participants study lists of single words. Because participants studied single words as opposed to word pairs, participants could not access relational information from a cue to inform JOL strategy use. Instead, both the JOL and item-specific tasks operated as deep encoding tasks which participants applied universally across all items in the study list (Craik & Lockhart, 1972). Our findings in Experiment 4 lend support to this notion, as participants applied relational encoding globally across pair types when explicitly instructed to engage in relational encoding rather than selectively as when making JOLs.

Finally, while the present study provides further support that JOL reactivity results from participants selectively engaging in relational strategies at encoding, we did not directly assess the type of encoding participants engage in while providing JOLs. Instead, we rely upon comparisons to similar relational tasks in Experiments 2-4 as a means of triangulating encoding processing (see Huff & Bodner, 2013; Meade, Klein, & Fernandes, 2020, for similar approaches). Additionally, our experiments did not include any online measures of strategic encoding at either study or test. While it has been well documented within the metacognitive literature that participants engage in strategic encoding both when acquiring new knowledge and when processing metamemorial information (e.g., Hertzog & Dunlosky, 2004; Nelson & Narens, 1990), our study did not explicitly assess whether participants were altering study strategies as a function of pair type. Rather, strategic changes of encoding strategy were inferred based on differences in cued-recall rates. Future research could utilize more direct measures such as having participants report the type of encoding strategy used during study as a function of pair type, which could also indicate any encoding changes consistent with a strategy-use account.

**Conclusion**

The present study provides a further examination of JOL reactivity and its underlying mechanisms. The use of multiple associative pair types within each experiment provided us with a more precise test of reactivity, the changed-goal and cue-strengthening accounts, and allowed us to test whether different associative pair types produce the same reactive benefits as forward associates. Overall, we found that the reactive benefits of JOLs can extend to both backward and symmetrical pairs (Experiment 1). Importantly, our findings from Experiments 2 and 3 indicate that the reactive effects associated with JOLs are not exclusive to JOLs and extend to other types of judgment tasks that both do and do not emphasize the associative characteristics of cue-target pairs. Finally, Experiment 4 provided further evidence that JOL reactivity occurs as a function of selective relational encoding of related pairs. Overall, our experiments demonstrate that memory forecasting from JOLs is not a prerequisite for reactivity and that JOL reactivity is primarily driven by selective encoding of related pairs.

**Open Practices Statement**

The data for all experiments have been made available at https://osf.io/8yvn3/. None of the experiments were preregistered.

References

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, & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6* (5), 576-587.

Balota, D. A., Yap, M. J., Hutchison, 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.

Bjork, R. A. (2016). Prologue: Some metacomments on metamemory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 1–3). Oxford: Oxford University 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.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6* (5), 588-598.

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.

Garskof, B. E., & Forrester, W. (1966). The relationship between judged similarity, judged association, and normative association. *Psychonomic Science, 6*, 504.

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. (2004). Aging, metacognition, and cognitive control. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 215−251). San Diego, CA, US: Academic Press.

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.

Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39* (4), 1246-1256.

Huff, M. J., & Bodner, G. E. (2014). All varieties of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language, 73*, 43-58.

Huff, M. J. & Bodner, G. E. (2019). Item-specific and relational processing both improve recall accuracy in the DRM paradigm. *Quarterly Journal of Experimental Psychology, 72* (6), 1493-1506.

Huff, M. J., Bodner, G. E., & Gretz, M. R. (2021). Distinctive encoding of a subset of DRM lists yields not only benefits, but also costs and spillovers. *Psychological Research, 85*, 280-290.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20* (5), 497-514.

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. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experiment Psychology: General, 126* (4), 349-370.

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.

Maki, W. S. (2007). Judgments of associative memory. *Cognitive Psychology, 54*(4), 319-353.

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*, 679-690.

Maxwell, N. P., & Buchanan, E. M. (2020). Investigating the interaction of direct and indirect relation on memory judgments and retrieval. *Cognitive Processing, 21*, 41-53.

Maxwell, N. P., & Huff, M. J. (2021). The deceptive nature of associative word pairs: Effects of associative direction on judgments of learning. *Psychological Research, 85*, 1757-1775.

Meade, M. E., Klein, M. D, & Fernandes, M. A. (2020). The benefit (and cost) of drawing as an encoding strategy. *Quarterly Journal of Experimental Psychology, 73* (2), 199-210.

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.

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition, 48*, 745-758.

Nelson, D. L., Dyrdal, G. M., & Goodmon, L. B. (2005). What is preexisting strength? Predicting free association, similarity ratings, and cued recall probabilities. *Psychonomic Bulletin & Review, 12*, 711-719.

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.

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition, 49*, 955-967.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

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.

Valentine, K. D., & Buchanan, E. M. (2013). JAM-boree: An application of observation oriented modeling to judgements of associative memory. *Journal of Cognitive Psychology, 25*(4), 400-422.

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*, 779-804.

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.

Footnotes

1 For completeness, we further analyzed the effect of block order on reactivity in Experiments 1-4. No interactions with block were found in Experiment 1 or Experiment 3 (*F*s < 2.50, *p*s > .06, *p*BICs > .99), however block did interact with pair type in Experiment 2, *F*(3, 285) = 4.41, *MSE* = 95.71, *ηp*2 = 0.01, and Experiment 4, *F*(3, 489) = 3.50s, *MSE* = 83.64, *ηp*2 = 0.01. We note, however, that all other interactions with block were not significant across either experiment (*F*s < 1.63, *p*s > .10, *p*BICs > .99). Post-hoc testing revealed that correct recall of symmetrical pairs in Experiment 2 was lower in block 1 (58.57) relative to block 2 (53.88), *t*(97) = 2.25, *SEM* = 2.11, *p* = .02. All other comparisons were non-significant (*t*s < 1, *p*BICs > .86). In Experiment 4, correct recall of backward pairs was numerically higher in block 1 (23.66) than block 2 (21.07), however, this comparison failed to reach conventional significance, *t*(165) = 1.77, *SEM* = 1.46, *p* = .08, *p*BIC = .81. All other comparisons were non-significant (*t*s < 1, *p*BICs > .89). Furthermore, the same general patterns of reactivity were detected in Experiments 2 and 4 after controlling for block order, indicating that block order did not contribute to the reactivity patterns reported.

2 Due to the COVID-19 pandemic, data collection was shifted online to Prolific partway through Experiment 2. In addition to the 70 participants recruited through the University of Southern Mississippi’s undergraduate pool, 28 participants were recruited through Prolific, with 11 completing the JOL task, 10 completing the JAM task, and 7 assigned to the no-JOL control group. Overall, mean recall did not differ between the Prolific or USM groups for the JOL task (44.06 vs 47.95), JAM task (46.09 vs 42.00), or the no-JOL control task (35.85 vs 38.66), all *t*s < 1, *p*s ≤ .48, *p*BICs ≥ .78. Thus, participant performance did not appear to be influenced by recruitment source.

3 As with Experiment 2, data collection in Experiment 4 was shifted online to Prolific midway through data collection in response to COVID-19. The forty participants in the no-JOL group were recruited through Prolific. Additionally, 20 participants in the relational group, 19 participants in the shallow group, and 2 participants in the JOL group were recruited via Prolific. For completeness, we note that mean correct recall did not differ between the no-JOL group in Experiment 4 and the undergraduate sample completing the same task in Experiment 2 (28.11 vs. 32.66; *t*(69) = 1.50, SEM = 3.08, *p* = .14, *p*BIC = .74). Additionally, within Experiment 4, recall did not differ between the undergraduate and Prolific samples in the relational group (44.81 vs. 38.05; *t*(43) < 1, *SEM* = 7.11, *p* = .33, *p*BIC = .79) or the vowel counting group (36.56 vs. 30.47; *t*(43) = 1.07, *SEM* = 5.87, *p* = .29, *p*BIC = .80). Thus, recall performance and JOL responses did not appear to differ as a function of participant source.

*Figure 1.* Comparison of mean recall rates in the JOL and No-JOL groups in Experiment 1. Bars = 95% CIs.

*Figure 2.* Comparison of mean recall rates in the JOL, JAM, and No-JOL groups in Experiment 2. Bars = 95% CIs.

*Figure 3.* Comparison of mean recall rates in the JOL, Frequency Judgment, and No-JOL groups in Experiment 3. Bars = 95% CIs.

*Figure 4.* Comparison of mean recall rates in the JOL, Relational Encoding, Vowel-Counting, and No-JOL groups in Experiment 4. Bars = 95% CIs.

**Appendix**

Across each of our four experiments, we tested for an illusion of competence pattern in the JOL group, given this pattern has not been reported consistently in JOL reactivity studies (cf. Mitchum et al., 2016). Given the prevalence with which this pattern occurs for backward pairs (e.g., Koriat & Bjork, 2005; Maxwell & Huff, 2021), this provided us with an additional test of the integrity of our dataset. Comparisons across all Experiments are reported in Table A4.

**Experiment 1**

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 group. A main effect of Pair Type was found, *F*(3, 114) = 421.81, *MSE* = 99.94, *ηp*2 = .92, in which JOLs/recall rates were highest for forward pairs (65.10), followed by symmetrical pairs (61.32), backward pairs (43.40), and unrelated pairs (14.14). Post-hoc *t*-tests showed that JOLs/recall rates significantly differed across all comparisons, *t*s ≥ 4.42, *d*s ≥ 0.32. Next, a significant effect of measure was observed, *F*(1, 38) = 10.02, *MSE* = 521.91, *ηp*2 = .21, in which JOL ratings (50.07) exceeded later recall rates (41.90). Importantly, a significant interaction between Pair Type and Measure, *F*(3, 114) = 68.55, *MSE* = 49.40, *ηp*2 = .64, confirmed the presence of an illusion of competence pattern. Follow-up *t-*tests indicated a robust illusion of competence for backward pairs whereby JOLs greatly exceeded later recall accuracy (55.18 vs. 31.67), *t*(38) = 7.59, *SEM* = 3.21, *d* = 1.56. Additionally, the illusion of competence extended to unrelated pairs (19.43 vs. 8.85), *t*(38) = 3.97, *SEM* = 2.75, *d* = 0.87, and symmetrical pairs (64.83 vs. 57.78), *t*(38) = 2.32, *SEM* = 3.14, *d* = 0.47, though the difference between judgments and recall smaller than backward pairs. Finally, for forward pairs, this pattern reversed—JOL ratings were significantly lower than cued-recall rates (60.87 vs. 69.34), *t*(38) = 2.93, *SEM* = 2.98, *d* = 0.57, indicating that participants underestimated their performance for this pair type and performed better than predicted at test.

**Experiment 2**

Using a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 2 (Measure: JOL vs. Recall) repeated measures ANOVA, we tested for the illusion of competence in the JOL group. Consistent with our predictions, this analysis yielded a significant effect of Pair Type, *F*(3, 96) = 269.87, *MSE* = 127.66, *ηp*2 = .89 that closely followed the patterns reported across the previous experiments. Specifically, mean JOLs/recall rates were highest for forward pairs (69.02), followed by symmetrical pairs (65.36), backward pairs (47.76), and were lowest for unrelated items (18.61). Comparisons differed statistically across each pair type, *t*s ≥ 3.04, *d*s ≥ 0.29. Next, the effect of measure was also significant, *F*(1, 32) = 10.32, *MSE* = 693.79, *ηp*2 = .24, in which JOL ratings were greater than cued-recall (55.16 vs. 45.36). Finally, a significant interaction between Pair Type and Measure confirmed that the illusion of competence pattern, *F*(3, 96) = 38.71, *MSE* = 64.82, *ηp*2 = .55. Starting with backward pairs, post-hoc analyses revealed that JOLs greatly exceeded subsequent later recall (60.15 vs. 35.61), *t*(32) = 6.92, *SEM* = 3.78, *d* = 1.54, a pattern that was echoed in unrelated pairs, (23.94 vs. 13.41), *t*(32) = 2.77, *SEM* = 3.71, *d* = 0.59, and symmetrical pairs, (70.14 vs. 60.68), *t*(32) = 2.89, *SEM* = 4.15, *d* = 0.61. Finally, for forward pairs, JOLs and recall did not significantly differ (66.25 vs. 71.74), *t*(32) = 1.44, *SEM* = 3.58, *p* = .16, *pBIC* = .67.

**Experiment 3**

First, to test for the illusion of competence in the JOL group, a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 2 (Measure: JOL vs. Recall) repeated measures ANOVA was used. As expected, this analysis revealed a main effect of Pair Type, *F*(3, 117) = 293.33, *MSE* = 151.31, *ηp*2 = .88, following the same pattern reported in the previous two experiments. JOLs/recall rates were highest for forward pairs (68.29), followed by symmetrical pairs (65.73), backward pairs (47.56), and lowest for unrelated items (17.14). All comparisons differed statistically, *t*s ≥ 2.38, *d*s ≥ 0.18. JOL ratings were only marginally greater than cued-recall rates (52.25 vs. 47.11), *F*(1, 39) = 3.56, *MSE* = 590.62, *p* = .07, *ηp*2 = .08, *p*BIC = .53, however a significant interaction confirmed the presence of an illusion of competence, *F*(3, 117) = 57.32, *MSE* = 68.40, *ηp*2 = .59. For backward pairs, JOLs greatly exceeded subsequent cued-recall rates (59.69 vs. 35.44), *t*(39) = 6.79, *SEM* = 3.69, *d* = 1.27. However, for unrelated pairs, the illusion of competence did not occur, as JOLs and recall were equivalent (16.77 vs. 17.53), *t* < 1, *p*BIC = .86, and this equivalence was also found on symmetrical pairs, (68.54 vs. 62.91), *t*(39) = 1.69, *SEM* = 3.44, *p* = .10, *p*BIC = .61. Finally, as found in Experiment 1, an underestimation pattern was found for forward pairs in which JOLs were generally lower than subsequent recall (64.03 vs 72.57), *t*(39) = 2.90, *SEM* = 3.04, *d* = 0.52.

**Experiment 4**

To test for the illusion of competence, we first conducted a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 2 (Measure: JOL vs. Recall) repeated measures ANOVA, assessing only participants who completed JOL encoding task. Consistent with Experiment 1, a main effect of Pair Type was found, *F*(3, 114) = 363.39, *MSE* = 112.72, *ηp*2 = .91, in which JOLs/recall rates were highest for forward pairs (65.68), followed by symmetrical pairs (63.15), backward pairs (44.43), and unrelated pairs (16.06). All comparisons differed significantly, *t*s ≥ 2.48, *d*s ≥ 0.22. A significant effect of Measure was also found, *F*(1, 38) = 50.54, *MSE* = 464.04, *ηp*2 = .57, such that JOL ratings (56.03) exceeded cued-recall rates (38.69). Finally, a significant interaction between Pair Type and Measure was found, indicating the presence of an illusion of competence, *F*(3, 114) = 56.41, *MSE* = 61.67, *ηp*2 = .60. Post-hoc tests indicated that an illusion of competence occurred for backward pairs such that JOLs greatly exceeded later recall rates (62.18 vs. 26.67), *t*(38) = 12.02, *SEM* = 3.05, *d* = 2.63. This pattern also occurred on unrelated pairs (22.30 vs. 9.87), *t*(38) = 4.07, *SEM* = 3.16, *d* = 0.96, and symmetrical pairs, (71.89 vs. 54.17), *t*(38) = 6.49, *SEM* = 2.79, *d* = 1.18. The illusion of competence, however, was not found on forward pairs, (67.63 vs. 63.78), but unlike Experiment 1, JOLs were equivalent to recall rates *t*(38) = 1.38, *SEM* = 2.91, *p* = .17, *p*BIC = .71.

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 |

*Mean Associative Strength Summary Statistics for Forward, Backward, and Symmetrical Pairs.*

*Note.* FAS (forward associative strength) and BAS (backward associative strength) values for unrelated pairs as these items share zero associative overlap.

Table A2

*Summary Statistics for Cue and Target Concreteness, Length, and Frequency Item Properties as a Function of Pair Type in Experiments 1A and 1B.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pair Type | 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 | Concreteness | 4.93 | 1.36 |
|  |  | Length | 5.05 | 1.62 |
|  |  | Frequency | 3.27 | 0.61 |
|  | Target | Concreteness | 4.44 | 1.37 |
|  |  | Length | 5.38 | 2.23 |
|  |  | Frequency | 3.18 | 0.73 |
| Unrelated | Cue | Concreteness | 4.59 | 1.40 |
|  |  | Length | 5.13 | 1.56 |
|  |  | Frequency | 3.20 | 0.80 |
|  | Target | Concreteness | 4.67 | 1.15 |
|  |  | Length | 5.30 | 1.49 |
|  |  | Frequency | 3.18 | 0.90 |

*Notes.* 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

*Comparisons of Mean Recall Percentages for each Pair Type in Experiments 1-4.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Experiment | Encoding Task | Pair Type | *M* | *95% CI* | F | B | S |
| Exp. 1 | JOL | Forward | 69.34 | 5.39 |  |  |  |
|  |  | Backward | 31.67 | 5.30 | 2.21\* |  |  |
|  |  | Symmetrical | 57.78 | 5.59 | 0.66\* | 1.51\* |  |
|  |  | Unrelated | 8.85 | 2.50 | 4.51\* | 1.72\* | 3.54\* |
|  | No-JOL | Forward | 48.08 | 5.21 |  |  |  |
|  |  | Backward | 16.09 | 3.30 | 2.30\* |  |  |
|  |  | Symmetrical | 36.03 | 4.97 | 0.74\* | 1.48\* |  |
|  |  | Unrelated | 9.68 | 3.16 | 2.80\* | 0.66\* | 1.99\* |
| Exp. 2 | JOL | Forward | 71.74 | 5.53 |  |  |  |
|  |  | Backward | 35.61 | 5.75 | 2.22\* |  |  |
|  |  | Symmetrical | 60.68 | 5.93 | 0.67\* | 1.46\* |  |
|  |  | Unrelated | 13.41 | 3.75 | 4.32\* | 1.56\* | 3.25\* |
|  | JAM | Forward | 67.58 | 6.74 |  |  |  |
|  |  | Backward | 36.36 | 5.71 | 1.71\* |  |  |
|  |  | Symmetrical | 61.29 | 6.42 | 0.32 | 1.40\* |  |
|  |  | Unrelated | 14.68 | 3.65 | 3.36\* | 1.58\* | 3.08\* |
|  | No-JOL | Forward | 55.16 | 6.28 |  |  |  |
|  |  | Backward | 27.34 | 6.13 | 1.55\* |  |  |
|  |  | Symmetrical | 46.41 | 6.36 | 0.48 | 1.06\* |  |
|  |  | Unrelated | 16.95 | 4.24 | 2.28\* | 0.68\* | 1.89\* |
| Exp. 3 | JOL | Forward | 72.57 | 5.20 |  |  |  |
|  |  | Backward | 35.44 | 6.52 | 1.95\* |  |  |
|  |  | Symmetrical | 62.91 | 6.21 | 0.52\* | 1.33\* |  |
|  |  | Unrelated | 17.53 | 7.15 | 3.25\* | 0.80\* | 2.09\* |
|  | Frequency | Forward | 66.58 | 5.87 |  |  |  |
|  |  | Backward | 31.23 | 6.14 | 1.85\* |  |  |
|  |  | Symmetrical | 62.05 | 6.21 | 0.23 | 1.56\* |  |
|  |  | Unrelated | 13.34 | 4.06 | 3.31\* | 1.08\* | 2.91\* |
|  | No-JOL | 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\* |
| Exp. 4 | JOL | Forward | 63.78 | 4.49 |  |  |  |
|  |  | Backward | 26.60 | 4.21 | 2.68\* |  |  |
|  |  | Symmetrical | 54.17 | 5.06 | 0.63\* | 1.85\* |  |
|  |  | Unrelated | 9.87 | 2.85 | 4.50\* | 1.46\* | 3.39\* |
|  | Relational | Forward | 58.17 | 6.69 |  |  |  |
|  |  | Backward | 30.89 | 7.56 | 1.12\* |  |  |
|  |  | Symmetrical | 50.06 | 6.73 | 0.35 | 0.78\* |  |
|  |  | Unrelated | 25.11 | 7.49 | 1.36\* | 0.22 | 1.02\* |
|  | Vowel | Forward | 39.19 | 6.72 |  |  |  |
|  |  | Backward | 14.13 | 5.68 | 1.20\* |  |  |
|  |  | Symmetrical | 29.83 | 6.37 | 0.42 | 0.78\* |  |
|  |  | Unrelated | 9.59 | 5.47 | 1.44\* | 0.24 | 1.02\* |
|  | No-JOL | Forward | 48.06 | 4.63 |  |  |  |
|  |  | Backward | 17.13 | 3.45 | 2.34\* |  |  |
|  |  | Symmetrical | 38.13 | 4.65 | 0.66\* | 1.59\* |  |
|  |  | Unrelated | 9.13 | 3.16 | 3.04\* | 0.75\* | 2.26\* |

*Note.* The three right-most columns indicate Cohen’s *d* effect sizes for post-hoc comparisons, \* = *p* < .05.

Table A4

*Comparison of Mean JOL Ratings and Correct Recall Percentages across Pair Types for the JOL Group in Experiments 1-4.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Experiment | Task | Pair Type | *M* | *95% CI* | F | B | S |
| Exp. 1 | JOL | Forward | 60.87 | 3.85 |  |  |  |
|  |  | Backward | 55.18 | 4.07 | 0.45\* |  |  |
|  |  | Symmetrical | 64.84 | 3.75 | 0.33\* | 0.77\* |  |
|  |  | Unrelated | 19.43 | 4.76 | 3.00\* | 2.53\* | 3.33\* |
|  | Recall | Forward | 69.34 | 5.39 |  |  |  |
|  |  | Backward | 31.67 | 5.30 | 2.21\* |  |  |
|  |  | Symmetrical | 57.78 | 5.59 | 0.66\* | 1.51\* |  |
|  |  | Unrelated | 8.85 | 2.50 | 4.51\* | 1.72\* | 3.54\* |
| Exp. 2 | JOL | Forward | 66.25 | 5.68 |  |  |  |
|  |  | Backward | 60.15 | 5.75 | 0.36\* |  |  |
|  |  | Symmetrical | 70.14 | 5.49 | 0.23\* | 0.59\* |  |
|  |  | Unrelated | 23.94 | 8.34 | 1.99\* | 1.70\* | 2.20\* |
|  | Recall | Forward | 71.74 | 5.33 |  |  |  |
|  |  | Backward | 35.61 | 5.71 | 2.22\* |  |  |
|  |  | Symmetrical | 60.68 | 5.93 | 0.67\* | 1.46\* |  |
|  |  | Unrelated | 13.41 | 3.75 | 4.32\* | 1.56\* | 3.25\* |
| Exp. 3 | JOL | Forward | 64.03 | 4.98 |  |  |  |
|  |  | Backward | 59.69 | 5.17 | 0.26\* |  |  |
|  |  | Symmetrical | 68.54 | 5.16 | 0.28\* | 0.53\* |  |
|  |  | Unrelated | 16.77 | 4.42 | 3.11\* | 2.77\* | 3.34\* |
|  | Recall | Forward | 72.57 | 5.20 |  |  |  |
|  |  | Backward | 35.44 | 6.52 | 1.95\* |  |  |
|  |  | Symmetrical | 62.91 | 6.21 | 0.52\* | 1.33\* |  |
|  |  | Unrelated | 17.53 | 7.15 | 3.25\* | 0.80\* | 2.09\* |
| Exp. 4 | JOL | Forward | 67.63 | 3.98 |  |  |  |
|  |  | Backward | 62.18 | 4.24 | 0.39\* |  |  |
|  |  | Symmetrical | 71.89 | 4.21 | 0.31\* | 0.72\* |  |
|  |  | Unrelated | 22.30 | 4.98 | 2.99\* | 3.30\* | 3.98\* |
|  | Recall | Forward | 63.78 | 4.49 |  |  |  |
|  |  | Backward | 26.60 | 4.21 | 2.68\* |  |  |
|  |  | Symmetrical | 54.17 | 5.06 | 0.63\* | 1.85\* |  |
|  |  | Unrelated | 9.87 | 2.85 | 4.50\* | 1.46\* | 3.39\* |

*Note.* The three right-most columns indicate Cohen’s *d* effect sizes for post-hoc comparisons, \* = *p* < .05.