Item-Specific and Relational Encoding are Effective at

Reducing the Illusion of Competence

Nicholas P. Maxwell, Emily E. Cates, & Mark J. Huff

The University of Southern Mississippi

Word count: XXXX

**Author Note**

Correspondence concerning this article can be addressed to Mark. J. Huff, School of Psychology, The University of Southern Mississippi, 118 College Dr. #5025, Hattiesburg, MS 39406, United States. Email: mark.huff@usm.edu. Study materials and analyzed data are available via OSF (https://osf.io/x9n4f/). Supplemental Materials have been made available at https://osf.io/svzg8/. This study was completed as part of the Honors Thesis requirements for EEC. Nicholas P. Maxwell is now at Midwestern State University.

Abstract

Metamemory, or the ability to understand the capacities of one’s own memory, is important for learning. A common method for assessing metamemory is the Judgment of Learning (JOL) task in which participants rate their likelihood of correctly recalling the target in a cue-target word pair when shown only the cue at test. The associative direction cue-targets pair can affect JOL accuracy. Unlike forward associate pairs (e.g., credit-card), in which JOLs often accurately predict recall, an illusion of competence has been reported for backward associates (e.g., card-credit), symmetrical associates (e.g., salt-pepper), and unrelated pairs (e.g., artery-bronze) in which JOLs overestimate later recall. The present study evaluates whether the illusion of competence can be reduced when participants use a deep item-specific or relational encoding strategy relative to reading. Across two experiments, we show that both item-specific and relational encoding reduce the illusion of competence for backward and unrelated pairs while improving the calibration between JOLs and recall. However, these encoding strategies largely reduced resolution, except for when pairs were unrelated. Therefore, item-specific and relational encoding strategies are effective at reducing the illusion of competence, and this accuracy benefit primarily reflects improved calibration via improved recall.

Word count: 193

*Keywords:* Judgements of Learning; Illusion of Competence; Item-Specific Encoding; Relational Encoding; Calibration; Resolution

Item-Specific and Relational Encoding are Effective at

Reducing the Illusion of Competence

Successfully monitoring the progress of one’s learning at study is paramount for improving retention, as effective monitoring allows individuals to adjust their encoding strategies to maximize later retrieval (Nelson & Narens, 1990). Metamemory judgments, or having individuals estimate the effectiveness of their memorial abilities, can be used to obtain information about an individual’s knowledge of the learning process and, as a result, can be used to assess how they adjust their encoding when faced with different study materials. Judgments of learning (JOLs) are a common measure of the metamemorial processes individuals use at study. In a standard JOL task, participants study sets of cue-target word pairs and are asked to estimate the likelihood that they can recall a target word in the presence of the cue on a later test. JOLs can be elicited using several types of measurement scales including Likert scales, binary “yes-no” responses (Hanczakowski, Zawadzka, Pasek, & Higham, 2013) or via a continuous 0 to 100 scale representing the percent likelihood that the target item will be successfully recalled at test (e.g., 100% = definitely would remember; 0% = definitely would not remember). The use of a 100-point scale is beneficial as it provides a straightforward computation of judgment accuracy which can be contrasted to the percentage of targets recalled at test.

Although JOL ratings can accurately predict later recall, several factors have been shown to influence their efficacy. These include perception of identical cue-target word pairs as being fluent due to word repetitions (Castel, McCabe, & Roediger, 2007; Mueller, Dunlosky, and Tauber, 2016), increasing the time spent studying word pairs (Koriat & Ma’ayan, 2005), and the direction and strength of the relatedness between cue-target study pairs (Koriat & Bjork, 2005; Maxwell & Huff, 2021). The present study further examines factors affecting the accuracy of JOLs by examining the associative direction between cue-target pairs (i.e., probability that the cue word elicits the target at test or vice versa) and by testing whether encoding tasks that emphasize the shared or distinctive characteristics of the word pairs through relational and item-specific encoding tasks, respectively, can improve the predictive accuracy of JOLs.

**The Effect of Cue-Target Relations on JOL Accuracy**

Interest in the correspondence between memory predictions and retrieval is not new. In an early example, Arbuckle and Cuddy (1969) asked participants to study letter-number pairs (e.g., A-73) and predict whether they would or would not remember the pairs on a later test. At test, participants also provided a postdiction that they were either correct or incorrect regarding their answer. Arbuckle and Cuddy reported that participants correctly predicted later recall for an average of 67% of trials and correctly postdicted their responses for 88% of trials, leading the authors to conclude that participants generally had insight into how difficult each pair would be to remember and adjusted their predictions accordingly.

More recently, Koriat and Bjork (2005) found that aspects of the associative relationship between cue-target study pairs, namely the direction and the strength of the relationship, can affect JOL accuracy. Specifically, the authors delineated between two types of associations thought to influence the relationship between JOLs and recall. First, *a priori* associations refer to associations in the forward direction (e.g., credit-card, stork-baby). The strength of these pair types is based on the likelihood that the cue word elicits the target word at test. The degree of a priori/forward associative strength can be readily assessed using free association norms (e.g., The University of South Florida Free Association Norms; Nelson, McEvoy, & Schreiber, 2004; The Small World of Words Project; De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019), which denote the probability of responding to word A with word B in an A-B pair (i.e., forward associative strength; FAS). Separately, *a posteriori* associations refer to any perceived relatedness between pairs that becomes more apparent to participants when words are presented together, rather than separately. While this type of association has traditionally referred to weakly associated forward pairs (e.g., article-newspaper; see Koriat & Bjork, 2005), they may also include strong forward associates in which the pair order has been flipped (e.g., backward pairs like card-credit, baby-stork, etc.). Thus, backward pairs reflect a specific type of a posteriori pair, as their relatedness is only apparent when these items are presented together. As with forward pairs, free association norms are useful for indexing the degree of backward associative strength (BAS) between pairs (i.e., the probability of responding to word B with word A in A-B pairs; see Nelson, McEvoy, & Dennis, 2000, for a review). Thus, a posteriori pairs could potentially have weak levels of FAS, strong levels of BAS, or both.

To test the correspondence between JOLs and recall for a priori and a posteriori pairs, Koriat & Bjork (2005) evaluated JOL accuracy when participants studied unrelated pairs, weak forward associates, and strong forward associates/a priori pairs (Experiment 1), a priori and a posteriori pairs (e.g., forward and backward paired associates; Experiment 2), and unrelated pairs, a forward associates, and semantically related a posteriori pairs that shared no forward or backward associations based on norms (Experiment 3). Across experiments, all a posteriori pair types demonstrated an *illusion of competence* pattern in which JOLs exceeded later recall rates, indicating that participants overpredicted the likelihood that they would later recall the target word. This pattern was particularly robust for backward pairs, as the cue word, when presented in isolation, does not ostensibly converge upon the studied target word. Thus, though participants predict that backward pairs are highly likely to be subsequently recalled, recall accuracy for this pair type is typically much lower than predicted.

The illusion of competence pattern found on backward pairs was similarly reported by Castel et al. (2007), who examined the correspondence between JOLs and subsequent recall when participants studied and provided JOLs for strongly and weakly related forward associates, unrelated items, and identical cue-target word pairs. Overall, an illusion of competence emerged for identical word pairs in which JOLs exceeded subsequent recall rates. While the authors ascribed this pattern to identical pairs being easier to learn, and therefore, more fluent versus forward and unrelated pairs given identical pairs were repeated items, participants may also hold the belief that identical pairs are better remembered (e.g., Mueller et al. 2016). As a result, participants may not have encoded identical pairs as deeply because they thought they would be easier to recall given the cue word was perfectly predictive of the target.

More recently, Maxwell and Huff (2021) further investigated the correspondence between JOLs and recall rates by assessing symmetrical associates (e.g., on-off) relative to forward, backward, and unrelated pairs. Symmetrical pairs differ from forward and backward pairs in that the associative strength between the cue and target word is equivalent in both directions (i.e., on-off would have approximately the same associative strength as off-on), whereas for forward and backward pairs, the association is stronger is one direction than the other (i.e., tuna-fish is strongly associated in the forward direction and weakly associated in the backward direction, while fish-tuna is weakly associated in the forward direction and strongly associated in the backward direction). Importantly, symmetrical pairs were matched to forward and backward pairs in overall associative strength. Thus, the only difference across pair types was the direction of the association. Across four experiments, Maxwell and Huff found a robust illusion of competence pattern for backward pairs, which occurred regardless of experiment pacing (self-paced vs. experimenter paced) and JOL timing (concurrent vs. immediate vs. delayed). Importantly, this illusion extended to symmetrical associates, suggesting that the bidirectional association found for symmetrical pairs is not sufficient for the cue word to regularly illicit the target word. Maxwell and Huff also suggested that participants may be using both the forward and backward associations when studying symmetrical pairs, even though only the forward association would be beneficial towards recall at test. Thus, the associative direction of a word pair strongly affects JOL accuracy, even when associative strength is matched across pair types, and furthermore, this effect is highly robust, replicating across a variety of manipulations.

Because the illusion of competence occurs across multiple pair types and study contexts, the present study further examines methods to potentially improve the correspondence between JOLs and recall. When making JOLs, participants base their judgments on several types of cues, including intrinsic and extrinsic cues (see Koriat, 1997; Rhodes, 2016). Intrinsic cues refer to the inherent properties of the study materials and can include frequency, imageability, and, importantly, word associations, whereas extrinsic cues refer to conditions at study, including study pacing (e.g., experimenter vs. self-paced) and any encoding manipulations participants complete concurrently with study (e.g., deep vs. shallow encoding tasks, Slamecka & Graf, 1978). Because participants base their JOLs on both intrinsic and extrinsic cues, a potential method to improve JOL accuracy, tested in the present study, is to have participants complete different types of encoding tasks that help them more easily process the existing relationships between the paired items. By providing participants with additional extrinsic cues with which to process items, these tasks may improve JOL accuracy. Specifically, these tasks would allow participants to more readily assess cue availability at encoding, potentially modifying JOLs and/or improving overall recall. This possibility may be particularly important given intrinsic and extrinsic cues may interact. For instance, if the encoding tasks draws attention toward or away from intrinsic cues such as pair relatedness, JOL accuracy may be affected. Thus, the qualitative features of how a study task affects processing of cue-target pairs may impact JOLs provided in addition to later recall.

**Item-Specific/Relational Processing on Memory Performance**

Memory researchers have long known that certain study tasks are more successful at improving retention than others. The levels-of-processing (LOP) framework classifies tasks that promote elaborative processing of studied items that generally promote memory as “deep” tasks, while less beneficial tasks that focus on surface or perceptual features of study items constitute “shallow” tasks (Craik & Lockhart, 1972; Craik, 2002). Several deep tasks have been identified and shown to improve retention, including generation (Slamecka & Graf, 1978), production (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), and survival processing (Nairne, Thompson, & Pandeirada, 2007). These deep tasks can be bifurcated further based on the task’s propensity to encourage the processing of item-specific or relational features of study pairs (Einstein & Hunt, 1980).

Regarding LOP effects on JOLs, few studies have assessed changes in JOL accuracy across different LOP tasks. In a recent exception, Tekin and Roediger (2020) compared the effects of deep and shallow LOP tasks on JOL reactivity. Across three experiments, participants studied individual words using deep (vs. shallow) encoding tasks. Specifically, participants in Experiments 1 and 2 completed encoding items while assessing case (shallow), rhyme (intermediate), or category (deep), while participants in Experiment 3 either completed a shallow vowel-counting task or a deep pleasantness rating task. Finally, across encoding groups, half of their participants also made JOLs after applying their deep or shallow strategy. Overall, deep encoding tasks enhanced correct recognition relative to shallow tasks, but critically, the magnitude of JOLs remained unchanged between participants in the deep and shallow encoding groups. Tekin and Roediger’s findings suggest that deep encoding may be an effective method for reducing the illusion of competence by increasing memory performance relative to JOLs (i.e., boosting memory performance to be more aligned with predicted memory). We test this possibility within the context of cued-recall testing (vs. recognition), while also investigating potential interactive effects between JOL accuracy and associative direction. Importantly, we further investigate the effects of deep processing on JOLs by delineating between two types of deep processing tasks that have been shown to benefit cued-recall performance: Item-specific and relational encoding.

According to the item-specific/relational framework (Einstein & Hunt, 1980; Hunt & Einstein, 1981), encoding tasks differ in the likelihood that they can encourage the processing of unique features of study items via item-specific processing, or through the processing of shared characteristics of study items via relational processing. Item-specific processing entails having participants focus on the unique features of items at study (e.g., for the pair cat-turtle, cats are mammals and turtles are reptiles, cats have fur and turtles have shells, etc.) while relational processing has participants focus on the shared features (e.g., cats and turtles are animals, both can be kept as pets, etc.). These types of processing qualitatively affect encoding strategies by changing how information is encoded in memory. Many studies have found differential memory benefits for item-specific and relational encoding tasks. For example, McCurdy, Sklenar, Frankenstein, and Leshikar (2020) showed that relational processing facilitated the generation effect for lower-constraint tasks (i.e., generating a target word in the presence of a cue), potentially because participants had to create a relationship between the two words. Relational processing could therefore be beneficial in studying unrelated word pairs since there is no existing relationship between the words and participants would be directed to create their own. Separately, Huff and Bodner (2014) found that item-specific tasks were more successful at improving recall and recognition when studying strongly than weakly related items. Conversely, relational tasks were more successful at improving recall and recognition when studying weakly than strongly related items revealing that tasks that affect processing of extrinsic mnemonic cues can interact with the intrinsic mnemonic cues of the study materials (e.g., Mulligan, 2011). Collectively then, despite that item-specific and relational processing tasks are generally classified as “deep” tasks based on the LOP framework, their relative memory benefits are affected by the association between study materials (see Huff & Bodner, 2014, for discussion).

**Assessing JOL Accuracy**

Given our interest in investigating how item-specific/relational encoding processes influence the relationship between JOLs and recall, it is important define what is meant by JOL accuracy. Metacognitive research typically draws a distinction between two types of JOL accuracy. First, absolute accuracy or *calibration* describes the overall difference between predicted recall (assessed via JOLs) and actual performance at test. In terms of probabilities, calibration reflects the likelihood that a probabilistic prediction of an event will correctly map onto the event’s true occurrence (Jiang, Osl, Kim, & Ohno-Machadao, 2012). For example, a participant would be said to be perfectly calibrated if items given a JOL rating of 100 were recalled 100% of the time at test. Item calibration has been a topic of extensive research across various domains of psychological research, including clinical psychology (Lindheim, Peterson, Mentch, & Youngstrom, 2020), eyewitness memory (Brewer & Wells, 2006), metacognitive confidence ratings (Double & Birney, 2017; Mitchum & Kelley, 2010), and, importantly, JOLs (Koriat, Sheffer, & May’ayan, 2002; Nelson & Dunlosky, 1991; Van Overschelde & Nelson, 2006). While calibration can be assessed in terms of bias (i.e., signed difference between JOLs and recall, e.g., Undorf & Bröder, 2020), its typically assessed by plotting mean JOL ratings against mean recall proportions, so long as JOLs and recall are measured using the same scale. These *calibration plots* allow researchers to assess whether JOLs are over or underconfident, and furthermore, they can be used to assess whether metacognitive illusions like the illusion of competence uniformly affect recall at all JOL levels or whether the illusion is stronger for high JOL ratings vs low ratings (e.g., Dunlosky & Nelson, 1992; Maxwell & Huff, 2021; Nelson & Dunlosky, 1991).

Next, the relative accuracy between JOLs and recall has commonly been of interest to metacognitive researchers. Relative accuracy or *resolution* refers to the degree to which a person’s JOL rating discriminates between what is and what is not remembered (Rhodes, 2016) Unlike calibration, which can be assessed through plots, resolution is commonly assessed via Goodman-Kruskal gamma correlations. The gamma coefficient represents a measure of association between -1 and +1, with resolution decreasing as gamma approaches zero. Positive values denote the degree that remembered items were given high JOLs and non-remembered items low JOLs, while negative gamma values denote the inverse of this pattern (Nelson, 1984).

Given that calibration and resolution reflect different elements of JOL accuracy, these metrics are influenced by different factors. Calibration is most strongly influenced whenever manipulations affect the magnitude of JOLs and/or the likelihood that encoded information will be successfully recalled at test (Rhodes, 2016). Thus, factors that have been shown to directly influence the magnitude of JOLs such as associative direction (Koriat & Bjork, 2005; Maxwell & Huff, 2021) or font-size (Rhodes & Castel, 2008) would be expected to produce changes in calibration. Similarly, encoding manipulations designed to improve recall (e.g., the item-specific/relational framework; Einstein & Hunt, 1980) would also be expected to influence calibration.

Whereas calibration is strongly influenced by factors that affect the magnitude of JOLs/recall, resolution is primarily impacted by factors influencing retrieval, including testing (King, Zechmeister, & Shaugnessy, 1980), practice (Koriat et al., 2002), and timing (Nelson & Dunlosky, 1991; see Rhodes, 2016 for a comparison of factors influencing resolution). Allowing participants to complete multiple test trials, engage in multiple study-tests cycles, or provide JOLs after a delay generally results in a resolution improvement. Thus, resolution is expected to improve anytime the encoding task affords participants with an opportunity to adjust their JOL ratings based on previous performance. However, given that these factors may also affect encoding, they may similarly affect calibration.

**Experiment 1: Effects of Item-Specific/Relational Encoding on JOL Accuracy**

Given the interactive benefits of item-specific and relational encoding with different associative materials (e.g., Huff & Bodner, 2014), Experiment 1 tested whether these encoding strategies can facilitate the calibration and/or resolution between JOLs and later recall, especially on backward and unrelated pairs in which the illusion of competence is robust (Castel et al., 2007; Koriat & Bjork, 2005; Maxwell & Huff, 2021). Specifically, we assessed JOLs and cued-recall performance for groups of participants who encoded cue-target pairs using either item-specific or relational encoding strategies relative to a standard JOL control task across forward, backward, symmetrical, and unrelated pair types. Additionally, we used calibration plots modeled after Maxwell and Huff (2021) to assess changes in calibration across each item type as a function of encoding strategy. Finally, changes in resolution were assessed using Goodman-Kruskal gamma correlations.

Overall, we sought to replicate the illusion of competence pattern for backward, symmetrical, and unrelated pairs for participants completing a silent reading intentional encoding control task while testing whether item-specific/relational encoding tasks could reduce the illusion of competence, either by lowering JOL ratings, increasing correct recall, or both. We expected that having participants engage in item-specific/relational encoding tasks would reduce the illusion of competence by improving correct recall relative to the read control task, but that the relative reduction in the illusion of competence would depend upon the associative pair type. Specifically, because relational encoding encourages participants to generate associations between cue-target pairs, we expected that relational encoding would be especially beneficial for unrelated pairs where the cue is ineffective at prompting target retrieval. Separately, because item-specific (vs. relational) processing has been shown to be more beneficial to memory when pairs are strongly related (Huff & Bodner, 2014), we expected that the item-specific task would be most beneficial for improving calibration on related pairs and, as a result, would be most effective at reducing and/or eliminating the illusion of competence for backward and symmetrical pairs. For forward pairs, which typically do not show an illusion of competence pattern (Maxwell & Huff, 2021), we predicted that the item-specific task could increase recall rates higher than the initial JOL ratings resulting in a situation in which JOLs *underpredict* subsequent recall.

Finally, although calibration is strongly affected by encoding manipulations designed to benefit recall, resolution was not expected to differ as a function of processing task, given that this type of JOL/recall relationship is primarily influenced by factors such as timing and practice, rather than changes in encoding strategy (Rhodes, 2016). Therefore, we expected that gammas measuring resolution would remain unchanged as a function of encoding strategy.

**Methods**

**Participants**

Eighty-eight University of Southern Mississippi undergraduates participated for partial course credit. Participants were randomly assigned to either the item-specific encoding group (*n* = 29), the relational encoding group (*n* = 31), or the read-only control group (*n* = 28). All participants were native English speakers with normal or corrected-to-normal vision. Sample sizes for each group were based on Maxwell & Huff (2021), and a sensitivity analysis conducted using *G\*Power 3.1* (Faul, Erdfelder, Lang, & Buchner, 2007) confirmed that this sample was sufficient to detect a small-to-medium main effects and interactions (Cohen’s *d* = 0.28) or larger.

**Materials**

The stimuli used were 180 associative word pairs initially used by Maxwell and Huff (2021). Pairs were taken from the University of South Florida Free Association Norms (Nelson et al., 2004). These consisted of 40 forward pairs (e.g., credit-card), 40 backward pairs (e.g., card-credit), 40 symmetrical pairs (e.g., salt-pepper), 40 unrelated pairs (e.g., art-lion), and 20 weakly related, non-tested buffer pairs used to control for primacy and recency effects. Pairs were divided evenly into two study blocks, each containing 20 of each forward, backward, symmetrical, and unrelated pairs, and 10 buffer pairs, for a total of 90 pairs in each list. All participants saw both lists presented in separate study-test blocks, the order of which was counterbalanced across participants. Each list began and ended with five buffer pairs, with the other pairs randomized anew for each participant.

Pair types were equated on associative strength (i.e., FAS and BAS) using the Nelson et al. (2004) free-association norms. Additionally, these pairs were designed to control for lexical and semantic properties that could potentially influence recall rates, including word length, SUBTLEX frequency (Brysbaert & New, 2009), and concreteness values derived from the English Lexicon Project (Balota et al., 2007; Maxwell & Huff, 2021). Further, both study blocks were matched on these properties. Thus, mean associative overlap and lexical/semantic properties were equivalent between direction types and study blocks. Finally, counterbalanced versions of the study lists were created that switched the order of the word pairs (i.e., forest-tree vs. tree-forest). As a result, forward pairs from one counterbalance became backward pairs on another and vice versa. Alternating pair direction allowed for greater control of item differences, particularly on forward and backward pairs, as the same items were used in the forward and backward directions across counterbalances. Pair order was similarly flipped and counterbalanced across unrelated and symmetrical pairs. Semantic and lexical characteristics for each pair type are reported in the Appendix (Tables A1-A2).

The cued-recall test in each block contained all 80 cue words from the studied pairs minus the buffer pairs which were not tested. The cue word was shown next to a question mark that had replaced the target word. Test order was newly randomized for each participant.

**Procedure**

The experimental procedure followed the general procedure used by Maxwell and Huff (2021). All participants completed the study individually on computers using *E-Prime* 3 software (Psychology Software Tools, Pittsburgh, PA). Participants were randomly assigned to one of the three encoding groups. For each study group, participants were instructed that they would study a series of cue-target word pairs and that their memory for the target word in these pairs would be tested later with the cue word present. The cue word was always presented on the left and the target on the right. Participants were instructed to rate (via JOL) how likely they were to remember the target word if they were only presented with the cue at test. JOL ratings were made using a 0 to 100 scale, with 0 being “I am certain I WILL NOT REMEMBER the word pair” and 100 being “I am certain I WILL REMEMBER the word pair.” Participants were instructed to use the full range of the scale to help reduce anchoring at points on the scale.

For the read group, participants were instructed to study the word pairs by reading them silently to themselves. For the relational group, participants were instructed to study the word pairs by thinking about how the pair of words were related to each other. Relational participants were similarly given the example of the word pair “Cat-Turtle” and were instructed to think about how cats and turtles are both animals and can both be pets. For the item-specific group, participants were instructed to study the word pairs by thinking about how the words in each pair were unique with the example that for the pair “Cat-Turtle”, participants might think about how cats have fur, but turtles have shells and how cats are mammals, but turtles are reptiles. Item-specific and relational groups similarly completed encoding tasks silently. Participants only saw one type of task instruction. After receiving the encoding instructions, participants completed a ten-word practice set using their assigned encoding task. Following completion of the practice trials, participants were required to describe their study strategy to the experimenter in their own words prior to starting the experiment. Participants were then given their first block of word pairs to study at their own pace and provided their JOL ratings while the word pair was displayed. Finally, after studying half of the pairs, participants were presented with a quick reminder to use their respective encoding strategy.

Following the first study list, participants completed an arithmetic filler task for 2 min followed by a cued-recall test in which only the cue word was presented and participants were tasked with providing the target word from memory. Participants were encouraged not to leave test answers blank and to try their best to retrieve the target word from memory. After the first cued-recall test was finished, participants completed a second study/test block using the same encoding instructions as the first. Once participants completed the second study/test block, they were debriefed and granted participation credit. Participants typically completed the experiment in under 1 hour.

**Results**

Prior to conducting analyses, study items that were missing JOL ratings or had ratings that were outside of the 0-100 range were removed. These responses were rare and fewer than 0.5% of items were removed. When scoring recall responses, test items that were skipped were scored as incorrect, and a liberal criterion for scoring correct items was adopted such that misspellings or pluralizations were scored as correct. All analyses were collapsed across block (analyses split by block are available in the Supplemental Materials; https://osf.io/svzg8/), and we note that the data patterns were similar between blocks. Partial-eta squared (*η*p2) and Cohen’s *d* eﬀect sizes were included for signiﬁcant analyses of variance (ANOVAs) and *t*-tests, respectively. For all analyses, a *p* < .05 signiﬁcance level was used unless noted otherwise. For all reported non-significant comparisons, we further analyzed the strength of the evidence supporting the null hypothesis using a Bayesian estimate (Masson, 2011; Wagenmakers, 2007). In this analysis, a model that assumes an effect is compared to a model that assumes a null effect. This process yields a probability estimate that the null hypothesis is retained (termed *p*BIC; Bayesian Information Criterion). The *p*BIC estimate is advantageous in that it is sensitive to sample size, increasing confidence in null effects reported. This Bayesian analysis is therefore supplementary to null effects detected using standard null-hypothesis-significance testing.

Mean JOL and recall rates as a function of pair type are reported in Figure 1. For completeness, all comparisons are reported in Appendix Table A3. A 2 (Measure: JOL vs. Recall) × 3 (Encoding Group: Item-Specific vs. Relational vs. Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA compared differences between mean JOL ratings and recall rates across pair types and encoding groups. An effect of Measure was found, *F*(1, 85) = 18.79, *MSE* = 694.46, *η*p2 = .18, such that collapsed across encoding groups and pair types, mean JOL ratings exceeded later recall rates (62.66 vs. 54.19). Next, an effect of Encoding Group was detected, *F*(2, 85) = 5.40, *MSE* = 814.98, *η*p2 = .11, in which JOL ratings/recall rates were significantly higher for the relational (61.44) and item-specific (60.12) groups relative to the read-only group (53.33). All comparisons differed significantly, *t*s ≥ 2.96, *d*s ≥ 0.78, except for the relational and item-specific groups, which were equivalent, *t* < 1, *p*BIC = .87. Finally, an effect of Pair Type was found, *F*(3, 255) = 766.58, *MSE* = 107.66, *η*p2 = 0.90, in which JOL ratings/recall rates were higher for symmetrical pairs (74.22), followed by forward pairs (72.29) backward pairs (59.01), and unrelated pairs (27.55). Comparisons across all pair types differed statistically, *t*s ≥ 2.69, *d*s ≥ 0.17.

A significant two-way interaction between Measure and Pair Type confirmed an illusion of competence pattern across encoding groups, *F*(3, 255) = 56.94, *MSE* = 87.42, *η*p2 = .40. Collapsed across encoding groups, mean JOLs approximated later recall for forward pairs (70.62 vs. 73.95, respectively; *t*(87) = 1.46, *SEM* = 2.33, *p* = .15, *p*bic = .77) and symmetrical pairs (75.82 vs. 72.63; *t*(87) = 1.27, *SEM* = 2.56, *p* = .21, *p*bic = .81). However, an overconfidence pattern occurred for backward pairs, as mean JOLs exceeded recall (69.96 vs. 49.24; *t*(87) = 8.32, *SEM* = 2.52, *d* = 1.17), a pattern which similarly extended to unrelated pairs (34.21 vs. 20.89; *t*(87) = 5.22, *SEM* = 2.59, *d* = 0.72)

Critically, a significant three-way interaction was also found, *F*(6, 255) = 15.56, *MSE* = 87.42, *η*p2 = .27, in which the magnitude of the illusion of competence differed as a function of encoding group. Starting with backward pairs, reliable illusion of competence patterns were detected across encoding groups, though at different rates. In the read-control group, a robust illusion of competence was detected in which JOLs greatly exceeded later recall accuracy (68.62 vs. 37.78), *t*(27) = 9.44, *SEM* = 3.41, *d* = 2.19. For the item-specific group, JOLs also exceeded recall (69.55 vs. 59.01), *t*(28) = 2.16, *SEM* = 5.12, *d* = 0.58, though at a lesser magnitude relative to the read condition. A similar pattern was observed in the relational group, where JOLs exceeded recall, but again at a lower rate than the read group (71.55 vs. 50.49), *t*(30) = 5.41, *SEM* = 4.05, *d* = 1.18.

For forward pairs, an illusion of competence pattern was not found for any of the encoding groups, with JOLs matching later recall for both the read group (70.04 vs. 65.23), *t*(27) = 1.32, *SEM* = 3.42, *p* = .19, *p*BIC = .69), and the relational group (72.96 vs 77.22), *t*(30) = 1.15, *SEM* = 3.86, *p* = .26, *p*BIC = .74. For the item-specific group, however, JOLs were lower than later recall rates (68.67 vs. 78.84), *t*(28) = 2.42, *SEM* = 4.41, *d* = 0.65—a situation in which JOLs underestimated later recall.

For symmetrical pairs, the illusion of competence was moderated by encoding task. For the read group, JOLs exceeded later recall accuracy (80.22 vs. 64.85), *t*(27) = 3.59, *SEM* = 4.48, *d* = 1.06; however, for both the item-specific and relational groups, the illusion of competence did not emerge as JOLs were equivalent to subsequent recall rates (71.62 vs 78.24), *t*(28)= 1.41, *SEM* = 4.90, *p* = .17, *p*BIC = .66, and (75.77 vs 74.41), *t* < 1, *SEM* = 3.46, *p* = .67, *p*BIC = .83, respectively.

Finally, for unrelated pairs, the illusion of competence was observed in both the read group (24.78 vs 14.73), *t*(27) = 3.23, *SEM* = 3.26, *d* = 0.76 and the item-specific group (40.64 vs 14.35), *t*(28) = 5.71, *SEM* = 4.81, *d* = 1.56, as JOLs exceeded later recall. However, the illusion of competence was not found in the relational group (36.59 vs. 32.52), *t* < 1, *SEM* = 4.52, *p* = .35, *p*BIC = .78), indicating that relational encoding provides a unique benefit on unrelated pairs by improving the correspondence between JOLs and subsequent recall.

Taken together, item-specific and relational processing tasks were each found to reduce and/or eliminate the illusion of competence pattern, but these reductions depended upon the pair type studied. Both item-specific and relational tasks were successful at eliminating the illusion of competence for symmetrical pairs and reducing it for backward pairs relative to reading. However, the two tasks interacted with pair type—relational encoding eliminated the illusion of competence on unrelated pairs, while the illusion of competence persisted on unrelated pairs for item-specific encoding. The different task effects on reducing the illusion of competence appeared to reflect an increase in correct recall rather than an adjustment of JOL ratings. Indeed, recall rates were greater in the item-specific (*M* = 57.62) and relational groups (*M* = 58.67), relative to the read group (*M* = 45.68; *t*s ≥ 3.13, *d*s ≥ 0.57), with the item-specific and relational groups being equivalent, *t* < 1, *pBIC* = .88. Mean JOL rates, however, were stable across the three encoding groups, *F*(2, 85) < 1, *MSE* = 147.50, *p* = .59, *p*BIC = .98, though encoding group interacted with pair type *F*(6, 255) = 7.27, *MSE* = 107.67, *η*p2 = .15. For unrelated pairs, JOLs were increased when participants used item-specific (*M* = 40.64) and relational encoding (*M* = 36.59) strategies relative to the read group (*M* = 24.85; *t*s ≥ 2.82, *d*s ≥ 0.73). Differences between the item-specific and relational encoding groups were not statistically reliable, *t* < 1, *pBIC* = .84. For related pairs, however, JOLs did not differ as a function of encoding strategy, with the exception of symmetrical associates in which mean JOLs were higher for participants in the read group (*M* = 80.20) compared to the item-specific encoding group (*M* = 71.65; *t*(55) = 2.49, *SEM* = 3.52, *p* = .02, *d* = 0.66). All other JOL comparisons involving related pairs were non-significant, *t*s ≤ 1.34, *pBICs* ≥ .76.

**Calibration Plots**

We further assessed the absolute accuracy between JOLs and recall for each pair types using a series of calibration plots (cf. Maxwell and Huff, 2021). To generate these plots, JOLs were first rounded to the nearest 10% increment, which resulted in 11 JOL bins ranging from 0% to 100%. For example, the 0% JOL increment contains the proportion of correct recall for items given a judgment of 0%, the 10% increment contains the proportion of correct recall for items given a judgment of 10%, etc. Mean correct recall for each JOL bin was then plotted. By plotting mean recall as function of JOL bin, these calibration plots allowed us to qualitatively assess whether the illusion of competence uniformly affected recall across JOL levels (e.g., Maxwell & Huff, 2021) and whether changes in the illusion of competence occurred as a function of encoding type (i.e., whether overestimations emerged at different JOL increments based on encoding instructions).

Figure 2 displays calibration plots for each encoding group as a function of pair direction. Plots are structured such that they include a calibration line denoting a perfect correspondence between JOL ratings and mean correct recall (e.g., a 40% JOL and a 40% correct recall rate would be perfectly calibrated). Overestimations are reflected by data points falling below the calibration line. Underestimations are represented by data points falling above the calibration line. These plots revealed important qualitative differences regarding specific JOL increments in which item-specific and relational encoding tasks start to reduce the illusion of competence pattern. For forward and symmetrical pairs, where illusions of competence are generally not found, all encoding groups showed similar calibration patterns. For unrelated and backward pairs, the illusion of competence pattern emerged at higher JOL increments in the item-specific/relational encoding groups relative to the read group. Item-specific encoding was most effective at increasing the JOL increment in which the illusion of competence pattern was detected for backward pairs (> 80%), while relational encoding was most effective at increasing the JOL increment for unrelated pairs (> 50%), again demonstrating qualitative differences in how item-specific and relational encoding at improving JOL accuracy.

These patterns were confirmed using a 3 (Encoding Group: Item-Specific vs. Relational vs Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL Increment) mixed ANOVA. Overall, this analysis yielded significant main effects of Pair Type, *F*(3, 255) = 41.12, *MSE* = 623.35, *η*p2 = .33, and JOL Increment, *F*(10, 850) = 50.15, *MSE* = 917.76, *η*p2 = .37. Importantly, a significant interaction was detected between Pair Type and JOL Increment, *F*(30, 2550) = 27.40, *MSE* = 448.77, *η*p2 = .24, confirming the presence of the illusion of competence. The interaction between JOL Increment and Encoding Group was non-significant, *F*(20, 850) < 1, *MSE* = 623.35, *p* = .75, *pBIC* > .99, and the three-way interaction was marginal, *F*(60, 2520) = 1.05, *MSE* = 1.30, *p* = .06, *pBIC* > .99.

**Resolution**

Finally, we assessed whether item-specific or relational encoding instructions affected the resolution between JOLs and recall. Following the procedure used by Nelson and colleagues (Dunlosky & Nelson, 1992; 1994; Nelson, 1984), we computed Goodman-Kruskal gamma correlations (*G*) between JOLs and recall for each participant for each of the four pair types (forward, backward, symmetrical, and unrelated; see Table 1 for mean *G*s and 95% *CI*s). A 3 (Encoding Group: Item-Specific vs. Relational vs Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA tested for differences in mean *G* across encoding groups and pair types. Overall, main effects/interactions were only marginally significant, *F*s ≥ 1.94; *p*s ≤ .07, *p*bics > .99; however, planned follow-up analyses were still carried out.

For forward pairs, both item-specific and relational encoding resulted in reduced resolution compared to silent reading (.10 vs. .13 vs. .35, respectively). All comparisons differed significantly (*t*s ≥ 2.56, *d*s ≥ 0.64), except for the comparison between item-specific and relational encoding, *t* < 1, *p* = .97, *p*bic = .88. This pattern subsequently extended to backward pairs (.12 vs. .07 vs. .24), though only the comparison between the relational encoding and read groups was significant, *t*(57) = 2.34, *SEM* = .07, *d* = 0.60, and all other comparisons for backward pairs were non-significant, *t*s ≤ 1.63, *p*s ≥ .11, *p*bics ≥ .67. For symmetrical pairs, *G* was again lower for item-specific and relational encoding relative to the read group (.15 vs. .13 vs. .23), however, all comparisons failed to reach conventional significance, *t*s ≤ 1.53, *p*s ≥ .13, *p*bics ≥ .70. Finally, for unrelated pairs, resolution was increased for participants who completed item-specific (.26) and relational encoding tasks (.33) relative to participants in the read group (.20). However, again, all comparisons failed to reach significance, *t*s ≤ 1.06, *p*s ≥ .29, *p*bics ≥ .81. Thus, while item-specific and relational encoding strategies are effective at reducing the illusion of competence, this reduction appears to occur primarily due to changes in calibration rather than resolution.

**Discussion**

Experiment 1 tested whether item-specific and relational encoding manipulations could reduce the illusion of competence by improving the correspondence between JOLs and recall. Relative to the read-only control group, both item-specific and relational encoding strategies were effective at reducing the illusion of competence for backward and symmetrical pairs. Furthermore, consistent with our predictions, relational encoding was also effective at reducing the illusion of competence for unrelated pairs. These patterns were confirmed via calibration plots, which provided qualitative information regarding the JOL bin in which overestimations emerge. For backward pairs, item-specific encoding increased this bin relative to relational encoding and the read-only control group, while unrelated pairs showed a similar increased when relational encoding was utilized. However, although these manipulations benefited calibration, they decreased resolution, particularly on related pairs. Taken together, our findings in Experiment 1 suggest that item-specific and relational encoding manipulations are each effective at reducing the illusion of competence, though these improvements reflect improvements to calibration rather than resolution.

While Experiment 1 demonstrated that the item-specific/relational framework can effectively reduce the illusion of competence for backward and symmetrical paired associates and unrelated pairs, we note that given item-specific and relational participants applied their encoding strategies silently, we are unable to verify separately that these strategies were being appropriately and consistently applied. Several experiments have reported consistent item-specific/relational processing differences using this procedure (e.g., Huff & Bodner, 2013; 2014), however, we sought to confirm the use of item-specific and relational processing though the use of a “think aloud” protocol. In this protocol, participants are not only instructed and required to practice their given study strategy, but also required to state aloud the item-specific or relational characteristics aloud for each word pair with an experimenter present to provide feedback. We apply this procedure in Experiment 2 to ensure the application of item-specific and relation encoding processes while also providing a replication of Experiment 1.

**Experiment 2: Think-Aloud Manipulation**

Experiment 2 sought to replicate previous findings from Experiment 1 showing that item-specific and relational encoding strategies improve the calibration between JOLs and recall. In doing so, participants in the item-specific and relational encoding groups applied their strategies using a vocal, think-aloud procedure in which they verbalized their thought processes. We expected that findings from Experiment 1 would replicate. Specifically, both item-specific and relational encoding strategies were expected to reduce the illusion of competence, with item-specific encoding being most effective on backward pairs and relational encoding most effective on unrelated pairs. Finally, because of the additional encoding due to participants vocalizing their encoding strategies versus completing the task silently, it is likely that the inclusion of the think-aloud protocol could further benefit to recall (i.e., a production effect; McLeod et al. 2010), but it is unclear whether this task would affect JOL magnitudes. To account for this possibility, we also included a cross-experimental analysis which tested for differences in JOLs and recall rates between participants who encoded pairs silently (Experiment 1) versus aloud (Experiment 2).

**Methods**

**Participants**

A total of 102 undergraduate psychology students were recruited from the University of Southern Mississippi and completed Experiment 2 for partial course credit. Consistent with Experiment 1, participants were randomly assigned to one of three encoding groups: Item-specific encoding (*n* = 34), relational (*n* = 32), or the read-only group (*n* = 36). This sample size was modeled after Experiment 1, with the constraint that each group contain at least 30 participants. A sensitivity analysis conducted via *G\*Power 3.1* (Faul et al. 2007) confirmed that our sample had sufficient power (.80) to detect small-to-medium main effects/interactions (Cohen’s *d* = 0.26).

**Materials and Procedure**

Experiment 2 used the same materials as Experiment 1 and followed the same general procedure with the following two exceptions. First, after receiving their respective encoding strategies, participants in the item-specific and relational encoding groups received additional instructions to vocalize their thought processes aloud during encoding. For example, a participant encoding the pair “mouse – cheese” with a relational strategy might state that these concepts are related because mice eat cheese, while a participant encoding the same pair with an item-specific strategy might instead state that these items differ because mice are animals while cheese is a type of food. Second, given the additional time needed for participants to implement this think-aloud procedure, Experiment 2 only consisted of one study-test block. All other procedures, including the practice trials and filler task, were identical to Experiment 1.

**Results**

Figure 3 displays mean JOL and recall percentages as functions of pair type and encoding group. For completeness, all comparisons are reported in Table A4. Like Experiment 1, differences between mean JOLs and recall were tested via a 2 (Measure: JOL vs. Recall) × 3 (Encoding Group: Item-Specific vs. Relational vs. Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated). A significant effect of measure was not detected, *F*(1, 99) < 1, *MSE* = 981.87,  *p* = .59, *p*bic = .90), as overall, mean JOLs did not differ from mean recall (60.85 vs. 59.39, respectively)., An effect of encoding group occurred, *F*(2, 99) = 23.48, *MSE* = 749.85, *η*p2 = .32, as mean JOLs/recall percentages were highest for the participants in the relational encoding group (69.73), followed by participants in the item-specific (59.83) and read groups (52.12; *t*s ≥ 3.05, *d*s ≥ 0.73). Additionally, an effect of Pair Type was detected, *F*(3, 297) = 359.85, *MSE* = 181.82, *η*p2 = .78, in which JOLs/recall rates were highest for symmetrical pairs (73.60), followed by forward (70.15), backward (62.86), and unrelated pairs (33.82). All comparisons differed statistically, *t*s ≥ 2.04, *d*s ≥ 0.29.

A significant Measure × Direction interaction was also found, *F*(3, 297) = 22.64, *MSE* = 168.97, *η*p2 = .19, which confirmed the presence of an illusion of competence pattern. Across encoding groups, mean JOLs were underconfident for forward pairs (66.58 vs. 73.72; *t*(101) = 2.59, *SEM* = 2.79, *d* = 0.38) but were overconfident for backward pairs (66.55 vs. 59.16; *t*(101) = 2.54 *SEM* = 2.92, *d* = 0.35) and unrelated pairs (39.01 vs. 28.64; *t*(101) = 3.41, *SEM* = 3.08, *d* = 0.43). However, for symmetrical pairs, JOLs did not differ from recall (71.22 vs. 75.99; *t*(101) = 1.68, *SEM* = 2.84, *p* = .10, *p*bic = .76).

Importantly, a significant three-way interaction confirmed that illusion of competence pattern differed as a function of Encoding Group. Beginning with backward pairs, a robust illusion of competence was detected in the read group, such that JOLs greatly exceeded later recall (65.86 vs. 45.26), *t*(35) = 4.09, *SEM* = 5.20, *d* = 1.03. However, this pattern did not extend to the item-specific group, as JOLs and recall did not significantly differ (62.22 vs. 67.70), *t*(33) = 1.34, *SEM* = 4.26, *p* = .19, *p*BIC = .70. Relational encoding similarly eliminated the illusion of competence (71.92 vs. 65.73), *t*(31) = 1.28, *SEM* = 5.03, *p* = .21, *p*BIC = .71.

Regarding forward pairs, no illusion of competence patterns emerged across any of the encoding groups. JOLs did not statistically differ from recall for participants in the read group (65.11 vs. 61.22), *t* < 1, *SEM* = 4.30, *p* = .35, *p*BIC = .79. Furthermore, consistent with Experiment 1, JOLs in the item-specific group were lower than subsequent recall (62.96 vs. 80.71), *t*(33) = 3.53, *SEM* = 5.24, *d* = 0.97. Finally, JOLs in the relational group marginally underestimated later recall (72.06 vs 80.38), *t*(31) = 1.84, *SEM* = 4.61, *p* = .08, *p*BIC = .52.

The illusion of competence was again moderated by encoding task for symmetrical pairs. Starting with the read group, JOLs exceeded later recall accuracy (74.32 vs. 64.06), *t*(35) = 2.27, *SEM* = 4.71, *d* = 0.60. However, item-specific encoding produced a noticeable under confidence effect on this pair type, with JOLs below recall (65.13 vs 82.52), *t*(33) = 4.03, *SEM* = 4.49, *d* = 0.96, while relational encoding produced JOLs that were marginally under confident (74.19 vs 82.44), *t*(31) = 1.80, *SEM* = 4.65, *p* = .08, *p*bic = .53.

Finally, the illusion of competence was observed on unrelated pairs for participants in both the read (28.58 vs 16.99), *t*(35) = 2.71, *SEM* = 4.42, *d* = 0.64, and the item-specific groups (39.78 vs 17.63), *t*(33) = 4.33, *SEM* = 5.33, *d* = 1.17. However, these JOL overestimations did not extend to participants in the relational group, as JOLs and recall did not statistically differ (49.84 vs. 53.29), *t* < 1, *SEM* = 5.89, *p* = .54, *p*BIC = .82), replicating patterns observed in Experiment 1 and providing further evidence that relational encoding uniquely benefits unrelated pairs.

**Calibration Plots**

Figure 4 displays calibration plots for Experiment 2. Starting with related pairs, both item-specific and relation encoding strategies influenced the correspondence between JOLs and recall. For forward and symmetrical pairs were generally well calibrated for participants in the read group, however participants completing the item-specific/relational encoding tasks greatly underestimated later recall, with recall approximating JOLs at only the highest JOL increments (> 80%). Consistent with Experiment 1, the illusion of competence pattern occurred at higher JOL increments for participants in the item-specific and relational groups relative to the read group. Furthermore, item-specific and relational encoding were equally effective at increasing the JOL increment in which the illusion of competence emerged on backward pairs (> 70% for both groups), while relational encoding was most effective for unrelated pairs (> 60%).

These patterns were tested via a 3 (Encoding Group: Item-Specific vs. Relational vs. Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL Increment) mixed ANOVA. This analysis yielded significant effects of Pair Type, *F*(3, 294) = 67.33, *MSE* = 2191.31, *η*p2 = .03, and JOL Increment, *F*(10, 980) = 33.02, *MSE* = 917.76, *η*p2 = .03. Additionally, a significant effect of Encoding Group was detected, *F*(2, 98) = 6.07, *MSE* = 6404.05, *η*p2 = .11. Next, a significant interaction Pair Type × JOL Increment interaction confirmed the presence of the illusion of competence, *F*(30, 2940) = 8.56, *MSE* = 1433.22, *η*p2 = .08. The JOL Increment x Encoding Group interaction was marginally significant, *F*(20, 980) = 1.53, *MSE* = 2191.31, *p* = .06, *p*BIC> .99, while the three-way interaction was non-significant, *F*(60, 2940) = 1.11, *MSE* = 1008.06, *p* = .26, *pBIC* > .99.

**Resolution**

Finally, we again assessed whether item-specific or relational encoding instructions influenced the resolution between JOLs and recall (see Table 2 for Mean *G*s and 95% *CI*s for all comparisons). A 3 (Encoding Group: Item-Specific vs. Relational vs Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA was used to test for differences in resolution as functions of encoding group and pair type. Overall, this analysis yielded a significant main effect of encoding group, *F*(2, 99) = 3.59, *MSE* = .24, *η*p2 = .07. Collapsed across pair types, resolution was greater for participants in the read group (.19) relative to the item-specific (.10) and relational encoding groups (.03). All comparisons were non-significant, *t*s ≤ 1.04, *p*s ≥ .30, *p*bics ≥ .82, except for the comparison between the read-only and relational encoding groups, *t*(66) = 3.01, *SEM* = .05, *d* = 0.74. Additionally, this analysis revealed a significant effect of pair type, *F*(3, 297) = 4.29, *MSE* = .19, *η*p2 = .04. Post-hoc testing indicated that resolution was greatest for unrelated pairs (.19), followed by symmetrical pairs (.17), forward pairs (.08), and backward pairs (.01). Resolution for backward pairs was significantly lower relative to symmetrical and unrelated pairs, *t*s ≥ 3.22, *d*s ≥ 0.37, though comparison between all other pair types were non-significant, *t*s ≤ 1.49, *p*s ≥ .14, *p*bics ≥ .72. Additionally, the Encoding Group × Pair Type interaction was non-significant, *F*(6, 297) = 1.69, *MSE* = .19, *p* = .12, *pBIC* > .99. Thus, like Experiment 1, item-specific and relational encoding reduced the illusion of competence primarily through improved calibration than resolution.

**Cross-Experimental Analysis**

Because participants in the item-specific and relational encoding in Experiment 2 were required to verbalize their encoding processes, it is possible that this task affected the magnitude of the JOLs and/or their recall performance. We tested this possibility using a 2 (Experiment) × 2 (Measure: JOL vs. Recall) × 3 (Encoding Group: Item-Specific vs. Relational vs. Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA. The only reliable interaction that emerged was the Experiment × Measure × Direction interaction, *F*(3, 552) = 3.94, *MSE* = 128.35, *η*p2 = .02. All other interactions with Experiment, including the four-way interaction, were non-significant, *F*s ≤ 2.02 *p*s ≥ .06, *p*bics ≥ .64.

Overall, collapsed across encoding groups, mean JOL ratings did not differ between Experiments 1 and 2 for forward pairs (70.23 vs. 66.58, respectively), *t*(188) = 1.67, *SEM* = 2.23, *p* = .10, *p*bic = .77, or backward pairs (69.26 vs. 66.55), *t*(188) = 1.19, *SEM* = 2.29, *p* = .24 *p*bic = .87. For symmetrical pairs, JOLs in Experiment 1 were marginally greater than Experiment 2 (75.35 vs. 71.22), *t*(188) = 1.81, *SEM* = 2.32, *p* = .07 *p*bic = .73, while JOLs for unelated pairs were marginally lower in Experiment 1 relative to Experiment 2 (33.69 vs. 39.01), *t*(188) = 1.81, *SEM* = 2.94, *p* = .07 *p*bic = .72. Thus, across pair types, having participants engage in the think-aloud procedure in Experiment 2 did not affect their JOLs.

Regarding recall, no differences emerged between experiments for forward pairs (73.92 vs. 73.72), *t* < 1, *SEM* = 2.87, *p* = .92 *p*bic = .93, or symmetrical pairs (72.70 vs. 75.99), *t*(188) = 1.22, *SEM* = 2.64, *p* = .22 *p*bic = .87. However, for backward pairs, recall was greater in Experiment 2 than Experiment 1 for backward pairs (49.27 vs. 59.16), *t*(188) = 3.01, *SEM* = 3.33, *d* = 0.44, and unrelated pairs (20.91 vs. 28.64), *t*(188) = 2.27, *SEM* = 3.41, *d* = 0.33. Thus, the additional encoding afforded by the think-aloud task boosted recall, but only for more challenging backward and unrelated pairs. Importantly however, the item-specific and relational encoding effects produced similar effects on reducing the illusion of competence on both experiments, demonstrating that participants were indeed applying item-specific and relational processing tasks effectively in Experiment 1 when encoding was completed silently.

Additionally, we examined experiment differences in calibration plots and resolution. First, cross-experimental differences in calibration plots were assessed via a 2 (Experiment) × 3 (Encoding Group: Item-Specific vs. Relational vs. Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL Increment) mixed ANOVA. Overall, this analysis yielded a significant Experiment × Direction interaction, *F*(3, 546) = 12.57, *MSE* = 1640.37, *η*p2 = .12. However, all other interactions, including the four-way interaction, failed to reach significance, *F*s ≤ 1.69, *p*s ≥ .08, *p*bics > .99. Regarding resolution, a 2 (Experiment) 3 × (Encoding Group: Item-Specific vs. Relational vs Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA confirmed that mean *G* did not differ as a function of experiment, as no interactions with Experiment were detected, *F*s ≤ 1.72, *p*s ≥ .16, *p*bics > .99. Thus, changes in calibration and resolution across pair types/encoding groups did not differ between experiments.

**Discussion**

Experiment 2 tested whether our findings in Experiment 1 that item-specific and relational encoding strategies reduce the illusion of competence would replicate. Importantly, participants in Experiment 2 completed the item-specific and relational tasks aloud rather than silently, which allowed us to ensure that they were consistently and correctly applying their respective strategies at encoding. Consistent with Experiment 1, item-specific and relational encoding tasks each reduced the illusion of competence relative to participants in the read-only control group. Specifically, item-specific encoding was effective at reducing the illusion on backward and symmetrical pairs, but not on unrelated pairs. Relational encoding, however, was similarly effective on backward and symmetrical pairs while also reducing the illusion of competence on unrelated pairs. These patterns were further explored via calibration plots, which qualified these patterns. We also found that reductions in the illusion of competence primarily reflected increased calibration rather than resolution, as these encoding strategies often reduced gamma on related pairs.

Finally, we included a set of cross-experimental analyses to assess whether JOLs and/or cued-recall rates differed between experiments, given that the think-aloud procedure in Experiment 2 may have influenced how participants encoded items and/or the cues they used to inform their judgments. Overall, these analyses revealed that while the think-aloud procedure improved recall relative to Experiment 1, this finding was moderated by pair type, such that only the more difficult backward and unrelated pairs showed this improvement. Furthermore, the magnitude of JOLs did not differ between experiments. Analysis of calibration plots and mean *G* similarly revealed that calibration and resolution patterns were consistent across experiments. Thus, our inclusion of the think-aloud procedure in Experiment provided a useful replication which ensured that participants remained attentive to their respective encoding tasks, but the inclusion of this additional task did not alter JOL accuracy patterns.

**General Discussion**

The goal of the present study was to reduce the illusion of competence by improving the predictive efficacy of JOL ratings on recall of forward, backward, symmetrical, and unrelated cue-target word pairs. Previous research has consistently found that JOLs tend to be over predictive on unrelated and deceptive backward pairs, resulting in an illusion of competence pattern (Koriat & Bjork, 2005; Maxwell & Huff, 2021). Because previous work has shown memory benefits when deep processing is used in conjunction with JOLs (Tekin & Roediger, 2020), Experiment 1 first sought to further qualify deep-processing effects by comparing item-specific and relational encoding—separate encoding tasks that promote deep processing. Experiment 2 was then designed to replicate Experiment 1 while also including a think-aloud procedure to ensure that participants applied their instructed encoding strategies. Additionally, across experiments, we evaluated the correspondence of JOLs/recall to test calibration and included gammas as a measure of JOL resolution.

Across experiments, forward pairs did not produce an illusion of competence pattern across any of the three encoding groups. However, consistent with previous research (e.g., Koriat & Bjork, 2005; Maxwell & Huff, 2021), illusions of competence emerged for backward and symmetrical paired associates and unrelated pairs, though these effects were moderated by encoding task. Starting with Experiment 1, participants in the standard, read-only JOL group showed a robust illusion of competence on backward pairs, which extended to symmetrical and unrelated pairs. Next, for the item-specific group, JOLs underpredicted later recall of forward pairs (cf. Koriat & Bjork, 2005; Castel et al. 2007). The illusion of competence again occurred for backward pairs, though it was reduced compared to both the read and relational encoding groups, a pattern consistent with our prediction that item-specific encoding would be most beneficial in reducing the illusion of competence for related pairs (cf. Huff & Bodner, 2014). For symmetrical pairs, item-specific encoding eliminated the illusion of competence. Unrelated pairs also showed an illusion of competence pattern, though this was reduced relative to the read group. Finally, for the relational encoding task, the illusion of competence pattern was eliminated for unrelated pairs, but as with item-specific encoding, the pattern was only reduced but not eliminated for backward pairs compared to participants in in the read group. For symmetrical pairs, the illusion of competence was again reduced.

These patterns extended to Experiment 2, in which participants completed the item-specific/relational encoding tasks aloud. Both strategies eliminated the illusion on backward and symmetrical pairs, and, consistent with Experiment 1, only relational encoding was effective on unrelated pairs. Furthermore, relative to Experiment 1, each encoding strategy was more effective, often producing underconfidence patterns in which JOLs underestimated recall. These patterns likely resulted due to the additional encoding afforded by the think-aloud procedure in Experiment 2, which further elevated recall rates relative to JOLs. Cross-experimental analyses confirmed that while recall rates improved from Experiment 1 to Experiment 2, JOL ratings were equivalent. Thus, any changes in the illusion of competence patterns between experiments were primarily driven by additional encoding due to participants vocalizing their encoding processes. Furthermore, our finding that patterns in Experiment 1 extended to Experiment 2 suggests that participants were effectively applying their respective encoding strategies in Experiment 1, even when tasked with completely these tasks silently. Taken together, findings from both experiments indicate both item-specific and relational encoding tasks each improve JOL accuracy versus a read task, though their relative effectiveness depends upon the associative direction of the pair type.

Following our analysis of mean JOL and recall rates, we constructed a series of calibration plots which assessed differences in absolute accuracy between JOLs and recall for each pair type as a function of encoding task. These plots reflected qualitative differences in JOL overestimation between encoding groups, particularly for backward and unrelated pairs. Starting with Experiment 1, read group participants were overconfident for unrelated pairs at all JOL increments and for backward pairs at all increments above 50%, a pattern consistent with Maxwell & Huff (2021). Thus, overestimation was most likely to occur for pairs in which relatedness cues used at encoding were not readily available at retrieval, replicating previous work on the illusion of competence (e.g., Koriat & Bjork, 2005; Maxwell & Huff). For the item-specific group, participants were again overconfident for unrelated pairs at almost all JOL increments, but overconfidence of backward pairs occurred at higher JOL increments relative to reading (80% vs. 50%, respectively). Finally, for the relational group, overconfidence for backward pairs emerged at JOL increments greater than 60%, and for unrelated pairs at increments greater than 50%. Thus, compared to the read and item-specific tasks, relational encoding greatly improved participants’ abilities to accurately predict their own recall for unrelated pairs, suggesting that unrelated pairs are particularly benefitted by relational encoding strategies. Finally, across all groups, participants were generally well-calibrated for forward and symmetrical pairs.

These patterns then extended to Experiment 2. Relative to participants in the read-only group, item-specific encoding again improved the correspondence between JOLs and recall for backward and unrelated pairs, thus reducing the illusion of competence. Similarly, relational encoding was again most effective at improving calibration on unrelated pairs. Therefore, across experiments, item-specific and relational encoding strategies primarily benefitted calibration whenever the target word did not readily converge upon the cue at retrieval.

While calibration plots revealed that item-specific and relational encoding strategies generally improved calibration across pair types, the effects of these strategies on resolution were limited by pair type. Across experiments, item-specific strategies improved resolution on unrelated pairs relative to participants completing the read control task. Additionally, relational encoding was effective at improving resolution for this pair type in Experiment 1. However, across related pair types, resolution primarily decreased for the item-specific and relational groups relative to the read control group. Therefore, item-specific and relational encoding strategies can produce a benefit on relative accuracy, but these improvements appear to be restricted to unrelated pairs. For related pairs, these encoding strategies produced a dissociation between calibration and resolution, such that calibration was improved at the cost of resolution. This disconnect may have resulted from how these encoding manipulations reduced the illusion of competence. When pairs were related, both item-specific and relational encoding reduced the illusion of competence by increasing cued-recall relative to the read-control task, but these manipulations did not affect the magnitude of JOLs. For unrelated pairs, however, item-specific and relational encoding strategies increased both recall and JOLs relative to reading, suggesting that these encoding strategies functioned as a control process that allowed participants to continuously modify their JOLs as they progressed through the study task. Thus, unrelated pairs showed improvements to both resolution and calibration, while calibration only benefited when pairs were related.

Our finding that item-specific and relational encoding strategies decreased resolution for related pairs aligns with previous work showing that this type of accuracy primarily benefits from manipulations occurring at test rather than encoding (e.g., Arial & Dunlosky, 2011; see Rhodes, 2016 for review). However, because resolution also benefits from manipulations affecting recall, the improved resolution on unrelated pairs, especially for the relational encoding group, may reflect improvements in correct recall performance. Consistent with this notion, recall of unrelated pairs in Experiment 1 was highest for participants in the relational encoding group, and further, this group showed greater resolution improvements for unrelated pairs relative to participants in the item-specific group. This pattern did not extend to Experiment 2, likely because recall of unrelated pairs in the relational group exceeded JOLs due to the think-aloud procedure.

While the present study is the first to employ the item-specific/relational framework to improve JOL accuracy, we note that Senkova and Otani (2021) compared recall performance for words receiving JOLs relative to lists studied using two item-specific encoding tasks (pleasantness ratings and single-mental imagery) that did not provide concurrent JOLs. Overall, neither item-specific encoding task increased recall relative to JOLs on either related or unrelated word lists. However, compared to a read-only control task that did not provide JOLs, both JOLs and item-specific encoding tasks boosted correct recall, leading the authors to conclude that the act of providing JOLs at study recruits item-specific processing. The present study, however, showed that when combined with JOLs, item-specific encoding strategies boost correct recall relative to standard, read-only JOLs. Thus, it is possible that item-specific encoding may produce an additional memory benefit when combined with JOLs. Of course, it is important to note that the test type differed between Senkova and Otani and our experiment (free-recall vs. cued-recall), so it is unknown whether the benefit of JOLs combined with item-specific encoding compared to JOLs alone would occur in other test types.

Finally, we note that our results complement work by Tekin and Roediger (2020), who showed that JOLs facilitated recognition memory for levels-of-processing (LOP) encoding tasks. Specifically, JOLs were particularly beneficial to memory when participants engaged in them alongside shallow encoding tasks (e.g., an e-counting task), such that the LOP effect (i.e., memory benefits of deep vs. shallow encoding; see Craik & Lockhart, 1972) was eliminated. While the present study was not designed to test the separate effects of JOLs on memory (i.e., JOL reactivity; Maxwell & Huff, 2022; Rivers, Janes, & Dunlosky, 2021; Soderstrom, Clark, Halamish, & Bjork, 2015) or test recall within the LOP framework as we did not include a shallow encoding group, we note that both item-specific and relational encoding strategies constitute deep-encoding tasks. Our finding that recall was greater when participants combined JOLs with these encoding strategies relative to making JOLs alone is consistent with previous research on the benefits item-specific/relational strategies (e.g., Hunt & Einstein, 1981) and, furthermore, suggests that Tekin and Roediger’s findings may extend to cued recall. However, given that the present study did not include no-JOL control groups who only engaged in only item-specific or relational encoding, more research will be needed to test the potential for additivity with JOLs. Regardless however, the benefits of item-specific/relational encoding appear to interact with different types of associative pairs which indicates that there may be boundaries in which relational and item-specific encoding tasks can reduce illusions of competence.

**Conclusion**

The present study found that the illusion of competence can be reduced when participants are directed to engage in item-specific or relational strategies at encoding. Across experiments, we found that the illusion of competence for backward and symmetrical associates can be reduced via item-specific encoding and that overestimation of unrelated pairs is reduced when participants use a relational encoding strategy. Calibration plots revealed that item-specific and relational encoding tasks generally improved the correspondence between JOLs and recall across pair types. However, resolution benefits were only found on unrelated pairs. Taken together, item-specific and relational tasks can be used to reduce, but not eliminate, the illusion of competence for backward, symmetrical, and unrelated word pairs which appears to be the produce of enhanced calibration between JOLs and subsequent recall.

**Open Practices Statement**

The data for all experiments have been made available at https://osf.io/x9n4f/. Neither experiment was pre-registered.

**Compliance with Ethical Standards:**

The studies reported were approved by the University of Southern Mississippi Institutional Review Board (Protocol #IRB-18-15) and found to be in accordance with the 1964 Helsinki Declaration ethical principles. Informed consent was obtained from all individuals who participated in this study. The authors report no competing interests.

**References**

Arial, R, & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition, 39*, 171-184.

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

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

Brewer, N. & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*(1), 11-30.

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*(4), 977-990.

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. (2002). Levels of processing: Past, present … and future? *Memory, 10*(5-6). 305-318.

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.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods,* *51*(3), 987-1006.

Double, K. S., & Birney, D. P. (2017). Are you sure about that? Eliciting confidence ratings may influence performance on Raven’s progressive matrices. *Thinking & Reasoning, 23*(2), 190-206.

Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*(4), 374-380.

Dunlosky, J. & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*, 545-565.

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.

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*(3), 429–444.

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.

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.

Jiang, X., Osl, M., Kim, J., & Ohno-Machado, L. (2012). Calibrating predictive model estimates to support personalized medicine. *Journal of the American Medical Informatics Association, 19*, 263-274.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93*, 329-343.

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

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., & Ma’Ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language,* *52*(4), 478-492.

Koriat, A., Sheffer, L., & Ma’Ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147-162.

Lindhiem, O., Peterson, I. T., Mentch, L. K., & Youngstrom, E. A. (2020). The importance of calibration in clinical psychology. *Assessment, 27*(4), 840-854.

Macleod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *36*(3), 671-685.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods,* *43*(3), 679-690.

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*(4), 1757-1775.

Maxwell, N. P., & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning, 17*, 589-625.

McCurdy, M. P., Sklenar, A.M., Frankenstein, A. N., & Leshikar, E. D. (2020). Fewer generation constraints increase the generation effect for item and source memory through enhanced relational processing. *Memory,* *28*(5), 598-616.

Mitchum, A. L., & Kelly, C. M. (2010). Solve the problem first: Constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 699-710.

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *The Quarterly Journal of Experimental Psychology, 64*(8), 1543-1562.

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people’s metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *The Quarterly Journal of Experimental Psychology, 69*(4), 781–799.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *33*(2), 263-273.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition,* *28*(6), 887-899.

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. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychonomic Bulletin, 95*(1), 109-133.

Nelson, T. O., & Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect.” *Psychological Science, 2*, 267-270.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation,* *26*, 125-173.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from <https://www.pstnet.com>

Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory, 29*(10), 1342–1353.

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.) *The Oxford Handbook of Metamemory* (pp. 90-117). Oxford, Oxford University Press.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions: *Journal of Experimental Psychology: General, 137*(4), 615-625.

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

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 & Memory,* *4*(6), 592-604.

Tekin, E. & Roediger, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift für Psychologie, 228*(4), 278-290.

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgments is strategic. *Quarterly Journal of Experimental Psychology, 73*(4), 629-642.

Van Overschelde, J. P., & Nelson, T. O. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy (resolution). *Memory & Cognition, 34*(7), 1527-1538.

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review,* *14*(5), 779-804.

Table 1

*Mean (± 95% CI) Goodman-Kruskal Gamma Correlations Between JOLs and Recall for each Encoding Group as a Function of Pair Type in Experiment 1*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Encoding Group | Forward | Backward | Symmetrical | Unrelated |
| Read | .35 (.12)\* | .24 (.13)\* | .23 (.10)\* | .20 (.18)\* |
| Item-Specific | .10 (.11) | .12 (.13) | .15 (.16) | .26 (.16)\* |
| Relational | .13 (.11)\* | .07 (.07) | .13 (.14) | .33 (.10)\* |

\* = significant from zero, *p* < .05

Table 2

*Mean (± 95% CI) Goodman-Kruskal Gamma Correlations Between JOLs and Recall for each Encoding Group as a Function of Pair Type in Experiment 2*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Encoding Group | Forward | Backward | Symmetrical | Unrelated |
| Read | .20 (.13)\* | .07 (.12) | .30 (.11)\* | .21 (.17)\* |
| Item-Specific | .02 (.18) | .11 (.15) | .22 (.18)\* | .30 (.19)\* |
| Relational | .02 (.20) | .01 (.14) | .04 (.17) | .12 (.14) |

\* = significant from zero, *p* < .05



*Figure 1.* Mean JOL and recall rates as a function of pair type in the Item-Specific group (top panel), Relational group (middle panel), and the Read group (bottom panel) in Experiment 2. Bars represent 95% confidence intervals.

Chart

Description automatically generated with medium confidence

Diagram

Description automatically generated

Chart

Description automatically generated with low confidence

*Figure 2.* Calibration plots as a function of pair direction in the Read Group (top panel), Item-Specific Group (middle panel), and Relational Group (bottom panel) in Experiment 1. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence interval.



*Figure 3* Mean JOL and recall rates as a function of pair type in the Item-Specific group (top panel), Relational group (middle panel), and the Read group (bottom panel) in Experiment 2. Bars represent 95% confidence intervals.

Engineering drawing

Description automatically generated with low confidenceDiagram, engineering drawing

Description automatically generated

Diagram, engineering drawing

Description automatically generated

*Figure 4.* Calibration plots as a function of pair direction in the Read Group (top panel), Item-Specific Group (middle panel), and Relational Group (bottom panel) in Experiment 2. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence interval.

**Appendix**

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 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.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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/Target | Concreteness | 4.70 | 1.38 |
|  |  | Length | 5.21 | 1.94 |
|  |  | Frequency | 3.23 | 0.67 |
| Unrelated | Cue/Target | Concreteness | 4.63 | 128 |
|  |  | Length | 5.21 | 1.52 |
|  |  | Frequency | 2.49 | 0.85 |

*Notes.* Frequency is measured using SUBTLEX word frequency measure (Brysbaert & New, 2009). Concreteness and length were taken from the English Lexicon Project (Balota et al., 2007).

Table A3

*Comparison of mean JOL ratings and correct recall percentages across pair directions for each encoding group in Experiment 1.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Encoding Group | Task | Direction | *M* | | *95% CI* | F | B | S |
| Item-Specific | JOL | Forward | 68.67 | | 5.95 |  |  |  |
|  |  | Backward | 69.55 | | 6.41 | 0.05 |  |  |
|  |  | Symmetrical | 71.62 | | 5.24 | 0.19 | 0.13 |  |
|  |  | Unrelated | 40.64 | | 7.49 | 1.51\* | 1.51\* | 1.74\* |
|  | Recall | Forward | 78.84 | | 5.47 |  |  |  |
|  |  | Backward | 59.01 | | 6.85 | 1.16\* |  |  |
|  |  | Symmetrical | 78.24 | 6.05 | | 0.04 | 1.08\* |  |
|  |  | Unrelated | 14.35 | 4.35 | | 4.75\* | 2.83\* | 4.42\* |
| Relational | JOL | Forward | 72.96 | | 4.86 |  |  |  |
|  |  | Backward | 71.55 | | 5.52 | 0.08 |  |  |
|  |  | Symmetrical | 75.77 | | 4.82 | 0.20\* | 0.29\* |  |
|  |  | Unrelated | 36.59 | | 5.90 | 2.37\* | 2.15\* | 2.66\* |
|  | Recall | Forward | 77.22 | | 6.09 |  |  |  |
|  |  | Backward | 50.49 | | 6.96 | 1.44\* |  |  |
|  |  | Symmetrical | 74.41 | 5.94 | | 0.16 | 1.30\* |  |
|  |  | Unrelated | 32.52 | 8.08 | | 2.07\* | 0.71\* | 1.95\* |
| Read | JOL | Forward | 70.04 | | 3.89 |  |  |  |
|  |  | Backward | 68.62 | | 4.39 | 0.13 |  |  |
|  |  | Symmetrical | 80.22 | | 4.20 | 0.93\* | 1.00\* |  |
|  |  | Unrelated | 24.85 | | 5.68 | 3.44\* | 3.19\* | 4.11\* |
|  | Recall | Forward | 62.23 | | 6.96 |  |  |  |
|  |  | Backward | 37.78 | | 5.91 | 1.40\* |  |  |
|  |  | Symmetrical | 64.85 | 6.34 | | 0.15 | 1.64\* |  |
|  |  | Unrelated | 14.76 | 3.96 | | 3.11\* | 1.69\* | 3.51\* |

*Note.* Mean JOL and recall rates for each associative direction condition across each encoding group. 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 all pair directions for each encoding group in Experiment 2.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Encoding Group | Task | Direction | *M* | | *95% CI* | F | B | S |
| Item-Specific | JOL | Forward | 62.96 | | 6.87 |  |  |  |
|  |  | Backward | 62.23 | | 6.85 | 0.03 |  |  |
|  |  | Symmetrical | 65.13 | | 7.59 | 0.10 | 0.13 |  |
|  |  | Unrelated | 39.78 | | 7.41 | 1.08\* | 1.06\* | 1.13\* |
|  | Recall | Forward | 80.71 | | 5.47 |  |  |  |
|  |  | Backward | 67.70 | | 6.26 | 0.74\* |  |  |
|  |  | Symmetrical | 82.52 | 4.05 | | 0.12 | 0.94\* |  |
|  |  | Unrelated | 17.63 | 5.14 | | 3.99\* | 2.94\* | 4.71\* |
| Relational | JOL | Forward | 72.03 | | 4.76 |  |  |  |
|  |  | Backward | 71.92 | | 4.82 | 0.01 |  |  |
|  |  | Symmetrical | 74.19 | | 4.92 | 0.15 | 0.16 |  |
|  |  | Unrelated | 49.84 | | 6.36 | 1.37\* | 1.36\* | 1.49\* |
|  | Recall | Forward | 80.38 | | 5.98 |  |  |  |
|  |  | Backward | 65.73 | | 8.31 | 0.69\* |  |  |
|  |  | Symmetrical | 82.45 | 5.39 | | 0.13 | 0.82\* |  |
|  |  | Unrelated | 53.29 | 9.73 | | 1.15\* | 0.47\* | 1.28\* |
| Read | JOL | Forward | 65.11 | | 4.28 |  |  |  |
|  |  | Backward | 65.86 | | 4.15 | 0.05 |  |  |
|  |  | Symmetrical | 74.32 | | 4.50 | 0.68\* | 0.64\* |  |
|  |  | Unrelated | 28.58 | | 5.84 | 2.34 | 2.41\* | 2.87\* |
|  | Recall | Forward | 61.23 | | 7.33 |  |  |  |
|  |  | Backward | 45.26 | | 8.19 | 0.67\* |  |  |
|  |  | Symmetrical | 64.06 | 6.60 | | 0.13 | 0.82\* |  |
|  |  | Unrelated | 16.99 | 5.94 | | 2.17\* | 1.30\* | 2.45\* |

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