Item-Specific and Relational Encoding, but not Warnings, are Effective at Reducing the Illusion of Competence

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

The University of Southern Mississippi

Abstract

Metamemory, or the ability to understand one’s own memory, is an important part of the learning process. One method in which metamemory can be assessed is through the Judgment of Learning (JOL) task. Prior work has shown that direction of the cue-target pair can influence the accuracy of JOLs. Unlike forward pairs (e.g., credit-card) in which JOLs and recall are well calibrated, an illusion of competence is generally observed for backward associates (e.g., card-credit) in which JOLs are overestimated relative to recall. This finding has been extended to include symmetrical associates (e.g., salt-pepper) and unrelated pairs (e.g., artery-bronze). The present study provides a further test of the illusion of competence for these four pair types and examines strategies that can be used to reduce or eliminate this effect. First, Experiment 1 tests whether item-specific and relational encoding strategies can be used to eliminate the illusion of competence. Next, Experiment 2 tests whether warning participants about this metacognitive illusion can further reduce or eliminate it. Overall, we show that item-specific processing reduces the illusion of competence for backward associates while relational processing reduces the illusion of competence for unrelated pairs. Warnings, however, did not affect JOL overestimation. These findings suggest that the method of encoding can be instrumental in facilitating future recall.

Word count: 211

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

Item-Specific and Relational Encoding, but not Warnings, are Effective at Reducing the Illusion of Competence

Accurately monitoring the progress of one’s learning is paramount for improving the learning process when studying new information. Effective monitoring allows individuals to adjust their encoding strategies to maximize later retention (Nelson & Narens, 1990). Metamemory judgments, or having individuals judge or estimate the effectiveness their memorial abilities can be used to obtain information about an individual’s knowledge of learning process. A common method used to gauge metamemory knowledge is the Judgment of Learning (JOL) task. In a standard JOL task, individuals study a set of cue-target word pairs and are asked to estimate the likelihood that they will be able to recall a target word when only provided with the cue on a later memory test. These estimates can be elicited using several types of measurement scales such as Likert Scales or binary “yes-no” responses (Hanczakowski, Zawadzka, Pasek, & Higham, 2013), however, JOLs are typically elicited using 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 allows for a straightforward comparison between predicted target recall (via JOLs) and the percentage of targets that are correctly recalled at test.

Although JOL ratings can be highly predictive of later recall (i.e., well-calibrated), several factors can affect the efficacy of JOLs. These include perception of identical cue-target word pairs as being highly semantically related (Castel, McCabe, & Roediger, 2007), increasing the time spent studying word pairs (Koriat & Ma’ayan, 2005), and the direction and strength of the relatedness between the cue-target study pairs (Koriat & Bjork, 2005; Maxwell & Huff, in press). The present study further examines factors that affect the accuracy of JOLs by examining the associative direction between cue-target pairs (i.e., probability that the cue item 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 accuracy of JOLs in predicting later recall.

Interest in the relationship between memory predictions and accuracy is not new. In an early example, Arbuckle and Cuddy (1969) asked participants to study paired letter-number associates (e.g. A-73) and report 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. Overall, Arbuckle and Cuddy reported that participants correctly predicted later recall for an average of 67% of trials and correctly postdicted their responses for an average of 88% of trials, leading the authors to conclude that participants had insight into how difficult each pair would be to remember and adjusted their predictions accordingly based on the association between participants’ predictions and subsequent recall.

More recently, research conducted by Koriat and Bjork (2005) has shown that the associative relationship between the cue-target pairs, such as the direction of the relationship and the strength of the relationship, can affect JOL accuracy. Specifically, the authors delineated between types two types of associations thought to influence the relationship between JOLs and recall. *A priori* associations refer to associations in the forward direction (e.g., credit-card, stork-baby). The strength of these pair types is rooted in the likelihood that the cue word will elicit the target word at test. *A priori*/forward association strength can be readily assessed through the use of 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). These norms are generated via free association tasks in which participants are provided with a single cue word and are asked to respond with the first target word that comes to mind. These norms can then be used to compute the probability of responding to word A with word B (i.e., forward associative strength, FAS). Separately, *a posteriori* associations refer to the perceived relatedness between pairs that are only apparent to participants when words are presented together. These pairs can refer to weakly associated pairs (e.g., article-newspaper) or strong associates in which the pair order has been flipped (i.e., backward pairs, card-credit, baby-stork). Similar to *a priori* pairs, free association norms can be useful for indexing the backward associative strength (BAS) between pairs (i.e., the probability of responding to word B with word A in an A-B item pairs; see Nelson, McEvoy, & Dennis, 2000 for a review). Thus, *a posteriori* pairs could have either weak levels of FAS or strong levels of BAS.

To test the correspondence between JOLs and recall for *a priori* and *a posteriori* pairs, Koriat & Bjork (2005) conducted three experiments in which participants were presented with unrelated and *a priori* study pairs (e.g., strong forward associates; Experiment 1), *a priori* and *a posteriori* pairs (e.g., backward associates; Experiment 2), and unrelated pairs, *a priori* pairs, and a set of semantically related *a posteriori* pairs that shared no association based on norms (Experiment 3). Across experiments, an *illusion of competence* was found for *a posteriori pairs* in which participants’ JOLs exceeded subsequent recall rates, indicating that participants were overpredicting 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 directly converge upon the target word. Thus, although participants predict that backward pairs are highly likely to be recalled, recall accuracy is typically much lower than predicted.

The illusion of competence pattern found with *a posteriori* and backward pairs has similarly been reported by Castel et al. (2007) who examined the correspondence between JOLs and subsequent recall when participants studied identical cue-target pairs. Participants studied a set of strongly and weakly related forward associates, unrelated items, and identical cue-target word pairs and provided JOL ratings. Overall, an illusion of competence emerged for identical word pairs in which JOLs exceeded subsequent recall rates. The authors ascribed this pattern to the identical pairs being easier to learn relative to forward associates and unrelated pairs due to their identical semantic similarity. As a result, participants may not have encoded the 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 (in press), further investigated the correspondence between JOLs and recall rates by looking at symmetrical associates (e.g., on-off), relative to forward, backward, and unrelated pairs. Symmetrical pairs are different from forward and backward pairs in that the associative strength between the cue and target word is the same regardless of direction (i.e., on-off would have the same associative strength as off-on), whereas for forward and backward cue-target pairs there is a stronger strength depending on the direction of the pair (i.e., tuna-fish is strongly associated in the forward direction, but has a weaker association in the backward direction, fish-tuna). Furthermore, symmetrical pairs differ from identical pairs in that they have equivalent levels of forward and backward associative strength without needing to repeat the same word. Across four experiments, Maxwell and Huff (in press) found a strong illusion of competence pattern for backward pairs and, additionally, the illusion of competence was shown to extend to symmetrical associates, suggesting that the weak association found for symmetrical pairs is not strong enough for the cue word to regularly illicit the target word. Maxwell and Huff (in press) also suggested that participants may be using both the forward and backward associations when studying the symmetrical pairs even though only the forward association would be beneficial at test. These findings indicate that the associative direction of a word pair plays a larger role than the associative strength in future recall.

An additional contribution of Maxwell and Huff’s (in press) study was they incorporated sets of calibration plots in which JOL ratings for pairs were rounded to each 10% interval, and then were plotted against their corresponding recall accuracy (see Nelson & Dunlosky, 1991). Calibration plots are useful because they can provide qualitative information regarding specific JOL ratings where participants are well-calibrated (i.e., pairs given a 40% JOL rating should be correctly recalled 40% of the time, pairs with a 60% ratings are recalled 60% of the time, etc.), compared to those ratings that are not. Calibration plots can therefore pinpoint specific ratings where JOLs become less predictive. Across experiments, Maxwell and Huff found that forward and symmetrical pairs were generally well-calibrated until the highest JOL ratings, but for backward and unrelated pairs, participant JOLs became over-predictive at JOL ratings of 30% or greater. Thus, the calibration plots revealed that an illusion of competence pattern emerged for all pair types, however this pattern was only found at the highest JOLs for forward and symmetrical pairs (70% and greater) but occurred at much lower JOL ratings for backward and unrelated pairs (30% and greater).

Given the Maxwell and Huff’s (in press) findings that illusion of competence patterns can be found diffusely across all pair types, the goal of the present study was to examine methods that could potentially be effective at increasing the correspondence between JOLs and recall, and thereby reduce the illusion of competence. One such method, tested in Experiment 1, is by having participants engage in different types of encoding strategies that may help or hinder the processing of the relationship between the cue-target pair, a discussion with which we now turn.

**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 framework classifies tasks that promote elaborative processing of studied items that typically promotes memory as “deep” tasks, while less successful tasks that focus on surface or perceptual features of study items as “shallow” tasks (Craik & Lockhart, 1972; Craik, 2002). Several deep tasks have been identified, including generation (Slamecka & Graf, 1978), production (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), and survival processing (Nairne, Thompson, & Pandeirada, 2007), however deep tasks can be bifurcated further based on a task’s propensity to encourage the processing of item-specific or relational features. According to the item-specific/relational processing 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 through item-specific processing, or they can encourage the processing of shared characteristics of study items through relational processing.

An example of item-specific processing would be studying the word pair cat-turtle by thinking about how cats and turtles are different, such as cats are mammals and turtles are reptiles. An example of relational processing would be studying the same pair by thinking about how cats and turtles are the same, such as both animals can be pets. These types of processing qualitatively affect encoding strategies by changing what information is being encoded, and thus later remembered. There have been several studies recently that have focused on item-specific and relational processing. In a study done by McCurdy, Sklenar, Frankenstein, and Leshikar (2020), they found that relational processing facilitated the generation effect for lower-constraint tasks (i.e. participants were given a cue word and asked to provide a target word) 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. Another study done by Gretz and Huff (2020) found that having participants use a deep, or distinctive, encoding task (i.e. participants were asked to rate the pleasantness of words) when studying word lists facilitated correct recognition. Item-specific processing could therefore be beneficial in studying cue-target word pairs because it is a deep processing task. Item-specific and relational processing have also been shown to produce memory benefits outside of correct memory, such as reducing associative false memory illusions (Huff & Bodner, 2013). Huff and Bodner (2013) found that item-specific encoding helped to reduce false recognition for words in the Deese-Roediger-McDermott (DRM) paradigm.

Given the documented benefits of item-specific and relational processing on memory, the present study tested whether these encoding strategies can improve the calibration 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, in press; Soderstrom, Clark, Halamish, & Bjork, 2015). Specifically, Experiment 1 compares JOLs and cued-recall performance for groups of participants who either encode cue-target pairs using an item-specific task, a relational task, or a standard read-only control task across forward, backward, symmetrical, and unrelated pair types. In Experiment 2, we then examined whether tests whether combining item-specific and relational encoding tasks with an explicit warning about the illusion of competence and the deceptive nature of different pair types could further reduces the illusion of competence and improve JOL calibration. Finally, across both experiments, we follow analyses used by Maxwell and Huff (in press) by plotting participants JOL ratings against their recall rates using a series of calibration plots to examine specific JOL ratings where participants may over/under predict subsequent recall.

**Experiment 1: Item-Specific Versus Relational Encoding Instructions**

The goals of Experiment 1 were twofold. First, this experiment sought to replicate the illusion of competence for backward, symmetrical, and unrelated pairs for participants completing the silent reading task. Next, it tested whether the encoding manipulations modeled after the Item-Specific/Relational framework (Hunt & Einstein, 1981) could reduce the illusion of competence by either lowering JOL ratings, increasing correct recall, or both. Overall, it was expected that having participants engage in these additional processing tasks at encoding would reduce the illusion of competence by improving correct recall relative to the control group. Additionally, because relational encoding encourages participants to generate an association between the cue target pairs, it was expected that this encoding manipulation would be beneficial across pairs given only the cue-word is available at test, but especially beneficial for backward and unrelated pairs where the cue is less effective at promoting target retrieval. Finally, because item-specific (vs. relational) processing has been shown to be more beneficial to memory when pairs are related (Huff & Bodner, 2014), it was expected that this encoding strategy would be most beneficial for improving JOL calibration overall and reducing the illusion of competence for backward and symmetrical pairs. Thus, the qualitative differences in item-specific and relational encoding were expected to produce differential benefits on improving JOL calibration depending on the pair type that was studied.

**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. . A sensitivity analysis using *G\*Power* (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that the sample had sufficient power (.80) to detect a small-to-medium effect size (Cohen’s *d* = 0.27) or larger.

**Materials**

The stimuli used were 180 associative word pairs originally used by Maxwell and Huff (in press). Pairs were taken from the University of South Florida Free Association Norms (Nelson et al., 2004) and 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 that were used to control for primacy and recency effects. Pairs were divided evenly into two study blocks, each containing 20 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 (Table 1). 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 from derived from the English Lexicon Project (Balota et al., 2007; Maxwell & Huff, in press; Table 2). Further, the two study blocks were also matched on each of 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 both the forward and backward directions across counterbalances. Pair order was similarly flipped and counterbalanced across unrelated and symmetrical pairs.

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. The order of the test was randomized anew for each participant.

**Procedure**

The experimental procedure followed of the general procedure used by Maxwell and Huff (in press). 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 three encoding groups: A read-only control, item-specific encoding, or relational encoding. 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 also instructed to use the full range of the scale to help reduce anchoring on the ends of 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 also given the example of the word pair “Cat-Turtle”, and how they might 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. Participants only saw one type of task instruction. After the encoding instructions, participants completed a ten-word practice set. Participants were then given their first block of word lists to study at their own pace and provided their JOL ratings while the word pair was displayed.

After the first study block was completed, participants were given two minutes to complete an arithmetic filler. Participants then completed a cued-recall task in which only the cue word was presented and were asked to provide 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, data were screened for missing responses and outliers (i.e., JOLs outside of the 0-100 range). Recall responses that were skipped were scored as incorrect. A liberal criterion for scoring correct items was adopted such that misspellings or pluralizations were scored as correct. All analyses were collapsed across block (see the supplemental section for analyses split by block). 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 non-significant comparisons reported, 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 and 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 with standard null-hypothesis-significance testing.

Mean JOL and recall rates as a function of pair type are reported in Figure 1. 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 was tested for differences between mean JOL ratings and recall rates across the four pair types across the three encoding groups. An effect of measure was found, *F*(1, 85) = 18.79, *MSE* = 694.46, *η*p2 = .07, such that overall, JOL ratings exceeded later recall rates (62.66 vs. 54.19), *t*(87) = 4.18, *SEM* = 2.06, *d* = 0.60. An effect of encoding group was also found, *F*(2, 85) = 5.40, *MSE* = 814.98, *η*p*2* = .05, 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, with the exception of the comparison between the relational and item-specific groups, which was non-significant, *t* < 1, *p*BIC = .87. Finally, a significant effect of pair type was found, *F*(3, 255) = 766.58, *MSE* = 107.66, *η*p2 = 0.58, in which JOL ratings/recall rates were higher for symmetrical pairs (74.22), followed by forward pairs (72.29) backward pairs (59.60), and unrelated pairs (27.55). Comparisons across all pair types differed statistically, *t*s ≥ 2.69, *d*s ≥ .17.

A significant two-way interaction between measure and pair type confirmed that the illusion of competence replicated across encoding groups, *F*(2, 85) = 5.21, *MSE* = 107.66, *η*p2 = 02. Critically, however, a significant three-way interaction was found, *F*(6, 255) = 15.56, *MSE* = 87.42, *η*p2 = .04, in which the magnitude of the illusion of competence differed as a function of encoding group (See Figure 1 for comparison across encoding groups).

Starting with backward pairs, a reliable illusion of competence pattern was found across each of the three encoding groups, though at different rates. A robust illusion of competence was detected in the read group in which JOLs greatly exceeded later recall accuracy (68.58 vs. 37.78), *t*(27) = 9.44, *SEM* = 3.41, *d* = 2.19. For the item-specific group, JOLs also exceeded recall (69.57 vs 58.97), *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 the JOLs exceeded recall, but again at a lower rate (71.54 vs 50.49), *t*(30) = 5.41, *SEM* = 4.05, *d* = 1.18.

Next, for forward pairs, an illusion of competence pattern was not found for any of the three encoding groups with JOLs matching later recall for both the read group (70.11 vs. 65.33), *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 actually lower than later recall rates (68.65 vs 78.85), *t*(28) = 2.42, *SEM* = 4.41, *d* = 0.65, revealing a situation in which JOLs can underestimate later recall.

For symmetrical pairs, the illusion of competence was moderated by encoding manipulation. For the read group, JOLs exceeded later recall accuracy (80.20 vs. 64.84, *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 (71.65 vs 78.23), *t*(28)= 1.41, *SEM* = 4.90, *p* = .17, *p*BIC = .66, and (75.81 vs 74.39), *t*(30) < 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 (40.65 vs 14.35), *t*(28) = 5.71, *SEM* = 4.81, *d* = 1.56, as JOLs exceeded later recall. However, the illusion of competence was eliminated in the relational group (36.62 vs. 32.51), *t*(30) < 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 both found to reduce and/or eliminate the illusion of competence pattern, but these reductions depended upon the type of pair type studied. Item-specific encoding was most successful at reducing the illusion of competence when participants studied backward associates. Relational encoding, however, was most beneficial for reducing the illusion of competence for unrelated item pairs.

We next assessed the correspondence between JOLs provided at study and correct recall for each of the pair types using a series of calibration plots (cf. Maxwell and Huff, in press). In these plots, JOLs were first rounded to the nearest 10% increment which were then plotted against the proportion of correct recall for items that were rated at that increment. For instance, the 0% JOL increment contains the proportion of correct recall for items given an initial judgment of 0%, the 10% increment contains the proportion of correct recall for items given an initial judgment of 10%, and so on.

Calibration plots for each of the four pair types are reported in Figure 2 as a function of encoding group. Plots are structured such that they include a calibration line which depicts a perfect one-to-one correspondence between JOL ratings and correct recall percentage (e.g., a 30% JOL and a 30% correct recall rate would be perfectly calibrated). Using these plots, overestimations reflected data points falling below the calibration line whereas underestimations reflected data points falling above the calibration lines.

Calibration plots were initially analyzed 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, however, the 3-way interaction was non-significant, *F*(60, 2520) = .81, *MSE* = 919.81, *p* = .86, *pBIC* = 1. We, next analyzed calibration plots separately for each of the encoding groups.

Starting with the read group, for unrelated pairs, JOLs were found to overestimate later recall at all JOL increments (JOLs > 30%). However, for associative pairs overestimations emerged at higher JOL ratings. For backward pairs, overestimations occurred at JOLs greater than 50%, while overestimations of symmetrical and forward associates each occurred at the highest JOL ratings (< 90%). Using a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL increment) mixed ANOVA, these patterns were confirmed by effects of Pair Type, *F*(3, 81) = 32.19, *MSE* = 50758.57, *η*p2= .51, JOL Increment, *F*(10, 270) = 9.74, *MSE* = 14084.99, *η*p2 = .27, and a significant interaction, *F*(30, 810) = 2.50, *MSE* = 2084.56, *η*p2 = .09.

Next, for the item-specific group, overestimations of unrelated pairs were observed for JOL ratings above 40%. For backward pairs, calibration of JOLs and recall was improved relative to silent reading, as overestimations occurred at JOL ratings greater than 80%. Finally, for symmetrical and forward associates, overestimation again occurred only for JOLs greater than 90%. These patterns were again confirmed by effects of Pair Type, *F*(3, 84) = 36.92, *MSE* = 57849.302, *η*p2= .57, JOL Increment, *F*(10, 280) = 8.00, *MSE* = 16024.10, *η*p2 = .22, and a significant interaction, *F*(30, 840) = 3.37, *MSE* = 2932.80, *η*p2 = .11.

Finally, for the relational group, JOL overestimations of unrelated pairs were reduced relative to the read and item-specific groups, as overestimations emerged JOL ratings above 50%. However, overestimations of associative pairs followed similar patterns as observed for the item-specific and read groups. Specifically, overestimations of backward pairs emerged at JOLs ratings greater than 60%, while overestimations of symmetrical and forward associates again occurred at JOLs greater than 90%. These patterns were confirmed by effects of Pair Type, *F*(3, 87) = 23.86, *MSE* = 31563.43, *η*p2= .45, JOL Increment, *F*(10, 290) = 10.14, *MSE* = 19751.25, *η*p2 = .26, and a significant interaction, *F*(30, 870) = 2.73, *MSE* = 2894.75, *η*p2 = .09.

Collectively, the calibration plots reveal 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. However, for unrelated and backward pairs, the illusion of competence pattern emerged at higher JOL increments relative to the read group. In particular, item-specific encoding was most effective at increasing the JOL increment in which the illusion of competence pattern was detected for backward pairs (> 80%), whereas relational encoding was most effective at increasing the JOL increment for unrelated pairs (> 50%), again demonstrating the differential benefits of item-specific and relational encoding at improving JOL accuracy.

**Discussion**

The goal of this Experiment 1 was to examine whether item-specific and relational encoding strategies would reduce the illusion of competence found with JOLs in the backward, unrelated, and symmetrical pairs. Overall, the illusion of competence replicated such that JOLs generally exceeded that of later recall. However, this effect was moderated by both pair direction and encoding group. First, an effect of pair direction was found such that JOL ratings/recall rates were highest overall for symmetrical pairs followed by forward, backward, and unrelated pairs in descending order, meaning that the associative direction of the pair types facilitated recall rates. Moreover, an effect of encoding emerged in which JOL ratings/recall rates were higher for the item-specific and relational groups compared to the read group, meaning that the encoding groups did improve recall rates overall compared to the control group. Though the illusion of competence emerged across encoding groups, the magnitude of the effect differed as a function of encoding strategy.

Starting with backward pairs, there was a strong illusion of competence found in the read group. There was still an illusion of competence found for the relational and item-specific groups, but it was greatly reduced, with the item-specific group reducing the illusion of competence the most. This shows that the item-specific group had the most benefit for the backward pairs, and this benefit could be because the participants are forced to create a new association for the cue-target pair instead of relying on the weak association present. For forward pairs, there was no illusion of competence found in any of the encoding groups, and JOL ratings were underestimated relative to recall when participants used item-specific encoding. These findings largely replicate previous work (e.g., Koriat & Bjork, 2005; Maxwell & Huff, in press) showing that forward pairs are generally immune to the illusion of competence. Next, for symmetrical pairs, an illusion of competence was found in the read group, but the effect did not replicate when participants studied using item-specific or relational encoding strategies, suggesting that these strategies were effective at reducing the illusion of competence, due to the encoding strategies strengthening the existing forward and backward association between symmetrically paired items. Finally, for the unrelated pairs, there was an illusion of competence found in both the read and item-specific groups. However, the illusion of competence was eliminated for the relational group. This shows that there is a unique benefit that relational encoding provides to improve recall rates, and this could be due to the relational encoding forcing participants to create an association for the unrelated cue-target pair.

Following the design of Maxwell & Huff (in press), calibration plots were computed in order to further explore the correspondence between JOLs and recall for each pair direction split by encoding task. Across all groups, participants were generally well calibrated for the forward and symmetrical pair types. For the read group, participants were overconfident for unrelated pairs at all JOL increments and for backward pairs above all JOL increments over half. This shows that participants in the read group were not very good at predicting their own recall rates for pairs that did not readily converge on the target. For the item-specific group, participants were overconfident for unrelated pairs at almost all of the JOL increments and overconfident for backward pairs above the higher JOL increments. This shows that participants in the item-specific group improved slightly at predicting their own recall over the read group. For the relational group, participants were overconfident for unrelated pairs above all JOL increments over half and overconfident for backward pairs above all JOL increments slightly above half. This shows that there was a significant improvement in participants’ abilities in the relational group to predict their own recall for the unrelated pairs and suggests that there is a benefit to studying unrelated word pairs with the relational encoding strategy. It is possible that item-specific and relational processing improve the calibration between JOLs and recall because item-specific and relational processing improve correct recall rates.

Though it is evident that both item-specific and relational encoding tasks can be beneficial for improving JOL accuracy and reducing the illusion of competence, neither of the study tasks were able to eliminate the illusion of competence completely. As such, Experiment 2 sought to further eliminate the illusion of competence through the use of a warning manipulation.

**Experiment 2**

Given the benefit found for item-specific and relational processing at improving JOL calibration, the purpose of Experiment 2 was to evaluate whether JOL calibration could be improved further by testing whether participants can adjust their JOL ratings in response to performance-related feedback. In the literature, there are several demonstrations that participants are able to adjust their memory responses in the presence of experimenter-provided instructions. For example, in the false memory literature, participants are often able to reduce their suggestibility when exposed to misleading details (e.g., Chambers & Zaragoza, 2001; Eakin, Schreiber, & Sergent-Marshall, 2003; see Blank & Launay, 2014, for a meta-analysis). Moreover, in the highly potent Deese/Roediger-McDermott (DRM) paradigm, the false memory illusion can similarly be reduced (though not eliminated) when participants are warned about the critical lure, especially when the warning is presented prior to study (Gallo, Roediger, & McDermott, 2001; McCabe & Smith, 2002; see Gallo, 2006 for review).

In a recent study done by Karanian, Rabb, Wulff, Torrance, Thomas, and Race (2020), they examined if providing participants with a warning could potentially reduce the effect of misinformation on memory for a witnessed event. Participants were shown a video of a crime being committed, were tested for their memory of the video, and were then asked to listen to a recording of someone retelling the crime they had watched. However, the audio retelling of the crime had some details altered and thus introduced misinformation. Some participants were given a warning about the misinformation in the audio recording before listening to it, some were given a warning after listening, and some were not given a warning. Karanian et al. (2020) found that regardless of when the warning was given, that those exposed to a warning showed fewer memory errors when tested on the original event than those who did not receive a warning. This is not surprising considering the effectiveness of warnings shown in past studies, but Karanian et al. (2020) also looked at neural activation and found that those who had been given a warning, regardless of timing, had more activation in visual areas and less activation in auditory ones. This shows that warnings may be able to change what parts of the brain participants use to access the information they’re trying to remember. Collectively, then, participants can improve their memory accuracy in response to experimenter instructions, though an important question is whether participants can also show similar accuracy benefits on metamemory judgments.

Unlike the false memory literature, there are fewer studies that have examined the effects of feedback/warnings on metamemory judgments. In one exception, Koriat and Bjork (2006) examined the effects of using a debiasing procedure to reduce the illusion of competence found for backward pairs. In this study, all participants completed one study-test block in which they provided JOL ratings for a mix of forward, backward, and unrelated cue-target pairs (e.g. rain-umbrella). Before completing the second study-test block, participants were split into either a theory-based or mnemonic-based group (Koriat & Bjork, 2006). The theory-based group was asked to evaluate a series of cue-target word pairs and say how likely they thought an individual would be to provide the target word when presented with the cue word. The experimenter then showed the participants the true percentages, pointed out cue-target pairs that showed an illusion of competence, and explained to them that participants often overestimate their JOLs for these pairs because they perceive the pairs to have a stronger association than is actually present. The mnemonic-based group simply completed filler tasks for the same amount of time that the theory-based group used to complete their evaluations. Koriat and Bjork (2006) found that when participants were given a new list of cue-target pairs that participants in the theory-based group showed a reduced illusion of competence compared to those in the mnemonic-group who did not show a significant change. A second experiment conducted for this study also showed that theory-based group did a better job of pacing themselves when allowed to self-pace their study of the cue-target pairs (Koriat & Bjork, 2006). For Experiment 2, we propose to cross warnings with the item-specific and encoding groups to see if warnings are more effective under these conditions.

The purpose of Experiment 2 was therefore to examine whether JOL accuracy could be improved further if participants were warned about the deceptive nature of word pairs—especially backward pairs—prior to studying a list of pairs. Like Experiment 1, 2 blocks containing separate lists of cue-target pairs were studied and immediately tested. Prior to study of Block 2, participants in the warning group were explicitly informed about the illusion of competence and highlighting that association between cue-target backward pairs are particularly deceptive given the cues are ineffective at promoting retrieval of the target at test. To enhance the effectiveness of the warning, we also showed participants a figure (taken from Maxwell & Huff, in press) which depicted the illusion of competence pattern, a procedure that was adopted from Koriat and Bjork (2005). Immediately following the warning instructions and presentation of the figure, participants then studied the second list of word pairs followed by a cued-recall test. We reasoned that warnings would be most effective if 1) participants were initially exposed to the different pair types in study/test formats and thus completed a study/test block before the warning, 2) if warnings were presented prior to study (vs. test; cf. Gallo, 2006), and 3) if warnings were accompanied by a figure depicting the illusion of competence found in an empirical study. The warning group was compared to a no warning group who was not informed of the illusion of competence prior to Block 2.

To further examine JOL calibration improvements, the effects of warning (vs. no warning) were also crossed with the read, item-specific, and relational encoding instructions in Experiment 1. Experiment 2 was therefore designed to assess whether calibration benefits for item-specific and relational instructions that improved recall rates, could be enhanced further with warnings that may improve JOL ratings.

**Methods**

**Participants**

A total of 216 participants were recruited for Experiment 2. Of these participants, 129 (17 in lab; 112 online[[1]](#footnote-1)) were recruited from The University of Southern Mississippi and were compensated with partial course credit, and 84 were recruited from Prolific (www.prolific.co) and were compensated with $4.50 for participation. All participants were randomly assigned to one of six between-subject groups. Of these participants, 12 were eliminated due to floor recall performance (15% or less across pair types), leaving 204 available for analysis. Removed participants were similarly distributed across encoding groups, leaving 37 in the read no-warning group, 33 in the read warning group, 37 in the item-specific no-warning group, 34 in the item-specific warning group, 34 in the relational no-warning group, and 29 in the relational warning group. All participants reported fluency in the English language and had normal or corrected-to-normal vision.

**Materials and Procedure**

All materials and procedures were identical to those used in Experiment 1 with one exception. Specifically, participants in the Warning groups were provided with information regarding the illusion of competence immediately prior to the start of the second study block. Specifically, participants were informed that JOL ratings could overpredict later recall, and this pattern was more likely to occur for backward pairs than other pair types. To ensure participants understood the illusion of competence pattern, they were provided with examples of backward, forward, symmetrical, and unrelated pair types, and were also shown a data figure (taken from Maxwell and Huff, in press) which plotted JOLs against later recall for each of the four pair types (see Figure 3). Verbatim warning instructions for the warning groups are available at https://osf.io/x9n4f/. All other procedural details from Experiment 1, including the use of read, item-specific, and relational instructions, and JOL instructions remained the same.

**Results**

Data were initial screened for missing responses and outliers as in Experiment 1. In the following analyses, because the warning manipulation was only applied to the second block in the warning group, JOL and recall analyses only used participant data on the second block in both the warning and no-warning groups.

In the analyses, we first examined the effects of warning on JOLs and recall rates. However, no main effect of warning was found, *F* < 1, *p*BIC = .92, and warning did not interact with any other factor, largest *F* = 2.03, *p* = .16, *p*BIC = .83. We report means across warning and no warning groups in Experiment 2 (see Figure 4), but for concision, do not include warning as a factor in the analyses below.

Mean JOL ratings and recall rates as a function of encoding task and pair type are reported in Figure 5. A 3 (Task Type: Read vs. Item-Specific vs. Relational) × 2 (Measure: JOL vs. Recall) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed measures ANOVA was conducted. An effect of measure was found, *F*(1, 198) = 58.71, *MSE* = 654.06, *η*p2 = .23, in which JOL ratings were greater than recall rates (55.82 vs. 46.14). An effect of task type was also found, *F*(2, 198) = 3.60, *MSE* = 1361.38, *η*p2 = .04, in which JOL/recall rates were lower in the read than the relational group (47.94 vs. 53.88), *t*(131) = 2.48, *SEM* = 2.38, *d* = 0.43, but equivalent in the read and item-specific groups (47.94 vs. 51.39), *t*(139) = 1.61, *SEM* = 2.14, *p* = .11, *p*BIC = .76. There was no difference between the relational and item-specific groups (53.88 vs. 51.39), *t*(138) = 1.16, *SEM* = 2.14, *p* = .25, *p*BIC = .85. An effect of pair type was also found, *F*(3, 594) = 1253.93, *MSE* = 168.01, *η*p2 = .86, which reflected greater JOL/recall rates for forward pairs (71.22), followed by symmetrical pairs (68.78), backward pairs (52.04), and unrelated pairs (18.22), all of which differed significantly from each other, *t*s > 3.60, *d*s > 0.18.

The main effects were qualified by a significant measure × pair type interaction, *F*(3, 639) = 134.27, *MSE* = 112.44, *η*p2 = .39, which confirmed the presence of the illusion of competence for backward, symmetrical, and unrelated pairs (but not forward pairs, which were well-calibrated), and a significant task type × pair type interaction, *F*(6, 639) = 298.36, *MSE* = 186.55, *η*p2 = .09. Importantly, the three-way interaction was also reliable, *F*(6, 639) = 298.36, *MSE* = 112.44, *η*p2 = .02. An illusion of competence pattern was found across all three encoding groups for both backward and symmetrical pairs, though again, the illusion was greater for backward (all *t*s > 9.13, *d*s > 1.38) than symmetrical pairs (all *t*s > 3.24, *d*s > 0.51). Additionally, forward pairs were well-calibrated: JOLs were equivalent to recall rates across encoding groups, all *t*s < 1.51, *p*s > .14, *p*BICs > .72. For unrelated pairs however, JOLs and recall rates were well-calibrated for the item-specific, *t*(70) = 1.69, *SEM* = 2.20, *p* = .10, *p*BIC = .68) and relational groups, *t* < 1, *p*BIC = .89, but not for the read group, in which an illusion of competence was found, *t*(69) = 3.36, *SEM* = 2.92, *d* = 0.48. Thus, relative to the read group, item-specific and relational processing eliminated the illusion of competence, but only for unrelated pairs.

We again constructed a series of calibration plots to assess the correspondence between the JOLs provided at study and correct recall for each of the four pair types (Figures 6-8). Consistent with our first experiment, calibration plots were initially analyzed 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. This analysis yielded a reliable 3-way interaction, *F*(60, 6120) = 62.26, *MSE* = 1221.10, *η*p2= .02. We then analyzed calibration plots separately for each of the encoding groups.

Starting with participants who completed the silent reading task in the no warning, overestimations were observed at nearly all JOL ratings (JOLs > 30%). Next, overestimation of backward pairs occurred at JOLs greater than 50%. For symmetrical associates, overestimations occurred for JOLs greater than 80%. Finally, overestimation of forward associates occurred only at the highest JOL ratings (< 90%). Next, for participants who received the warning, overestimations were observed for all JOL ratings above 40%. For paired associates, overestimation of backward pairs occurred at JOLs greater than 40%, overestimations of symmetrical associates emerged at JOLs above 70%, and overestimation of forward associates occurred for JOL ratings above 80%. Using a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL increment) repeated measures ANOVA, these patterns were confirmed by effects of Pair Type, *F*(3, 216) = 62.26, *MSE* = 77092.57, *η*p2= .46, JOL Increment, *F*(10, 720) = 10.03, *MSE* = 12526.10, *η*p2 = .12, and a significant interaction, *F*(30, 2160) = 3.99, *MSE* = 3069.50, *η*p2 = .05.

Next, for participants in the item-specific encoding group who did not receive the warning instructions, overestimations of unrelated pairs reduced relative to the read group, with overestimations emerging for JOL ratings above 60%. For backward pairs, overestimations occurred at JOL ratings greater than 60%. Next, for symmetrical associates, overestimations were observed at JOL ratings above 80%. Finally, for forward associates, overestimation again occurred only for JOLs greater than 90%. When participants who engaged in item-specific encoding received a warning, overestimations of unrelated pairs were observed for JOL ratings above 30%. For backward pairs, overestimations occurred at JOL ratings greater than 50%. Next, for symmetrical associates, overestimations occurred for JOLs greater than 80, while for forward associates, they were detected at JOLs above 90%. Significant effects of Pair Type, *F*(3, 219) = 80.95, *MSE* = 97661.16, *η*p2= .53, JOL Increment, *F*(10, 730) = 17.34, *MSE* = 24705.25, *η*p2 = .19, and a significant interaction, *F*(30, 2190) = 6.61, *MSE* = 5858.23, *η*p2 = .08, again confirmed these patterns.

Finally, we assessed the calibration between JOLs and recall for participants who completed the relational encoding task. Starting with participants in the no warning relational encoding group, JOL overestimations of unrelated pairs were again reduced relative to the read and item-specific groups, as overestimations emerged JOL ratings above 70%. Next, overestimations of backward pairs emerged when JOLs ratings were greater than 50%, while overestimations of symmetrical associates again occurred at JOLs greater than 80%. Finally, for forward associates, overestimations only occurred at JOLs greater than 90%. Finally, for participants who received the warning, JOL overestimations of unrelated pairs were reduced relative to the read and item-specific groups, as overestimations emerged JOL ratings above 60%. For backward associates, overestimation occurred for JOLs ratings greater than 60%, while overestimations of symmetrical and forward associate at higher JOLs (> 80). Once again, all patterns of overestimation were confirmed by significant effects of Pair Type, *F*(3, 198) = 67.82, *MSE* = 74599.51, *η*p2= .51, JOL Increment, *F*(10, 660) = 27.03, *MSE* = 31746.80, *η*p2 = .29, and a significant interaction, *F*(30, 1980) = 7.67, *MSE* = 6008.28, *η*p2 = .10.

[SUMMARY PARAGRAPH HERE] Similarly to Experiment 1, the calibration plots for Experiment 2 revealed qualitative information about specific JOL increments where encoding tasks start to reduce the illusion of competence. All encoding groups showed similar patterns for the forward and symmetrical pairs because these pairs are typically resistant to the illusion of competence. Unlike Experiment 1, however, the backward pairs also showed similar patterns across encoding groups. This shows that the item-specific encoding group did not benefit the backward pairs the way it did in the first experiment. For the unrelated pairs, the illusion of competence pattern emerged at higher JOL increments relative to the read group. Specifically, the no warning relational group was most effective at increasing the JOL increments in which the illusion of competence pattern was observed for unrelated pairs (>70%) followed by the no warning item-specific and warning relational groups (>60%), both compared to the read group (>30%).

**Discussion**

In Experiment 2, a warning manipulation was used in an attempt to further reduce the illusion of competence. We expected that providing a warning would further improve JOL calibration by encourage participants to reduce their high JOL ratings, particularly on backward pairs that were most susceptible to the illusion of competence. Despite providing participants with an explicit warning of the illusion of competence, providing them with a data figure plotting the illusion pictorially, and directing their attention towards the deceptive nature of backward pairs in particular, the warning proved unsuccessful.

There was a minimal illusion of competence for forward pairs and a very small illusion of competence for symmetrical pairs and encoding tasks did not differ for these pair types. The illusion of competence patterns observed in Experiment 1 replicated for backward and unrelated word pairs across encoding groups, with backward pairs having the greatest illusion of competence. Unlike Experiment 1, however, the item-specific encoding group did not reduce the illusion of competence for backward pairs. For unrelated pairs, there was an illusion of competence found in the read group, but both the item-specific and relational encoding groups reduced the illusion of competence. So, while some benefits were found in using the item-specific and relational encoding strategies, these benefits were restricted to unrelated pairs. Thus, the findings for Experiment 2 suggest that relative to silent reading, engaging in item-specific and relational encoding can eliminate the illusion of competence when study pairs are unrelated.

Finally, using calibration plots, we further assessed the correspondence between JOLs and recall. Overall, these plots showed similar patterns as Experiment 1. Across encoding groups, overestimations only occurred for forward and symmetrical associates at high JOL increments. For backward associates, the illusion of competence occurred; however, the magnitude of these overestimations was moderated by the encoding task. Overestimations were reduced when participants engaged in item-specific encoding at study. Finally, for unrelated, pairs, overestimations were reduced when participants engaged in relational encoding at study.

**General Discussion**

The purpose of the present study was to improve the calibration of JOL ratings and recall rates for backward and unrelated word pairs through the use of item-specific and relational encoding and through the use of warnings, thus reducing the illusion of competence found for these pairs. Experiment 1 focused on the use of item-specific and relational processing. The goal of the item-specific processing tasks was to have participants focus on the distinct characteristics of the words in the cue-target pairs. We specifically hypothesized that item-specific processing would be most beneficial for the backward pairs because focusing on the distinct properties of the words made participants create a new, forward association for the words instead of relying on the weak association present. The goal of the relational processing tasks was to have participants focus on the shared characteristics of the words in the cue-target pairs. We hypothesized that relational processing would be most beneficial for the unrelated pairs because focusing on the shared characteristics of the words would help participants create an association that they could later use to better remember the pair. Across both experiments, item-specific and relational processing each affected the calibration between JOLs and overall recall. Specifically, the item-specific encoding strategy was able to greatly reduce the illusion of competence found in the backward pairs in the first experiment and the relational encoding strategy was able to reduce the illusion of competence found for the unrelated word pairs in both experiments. Experiment 2 focused on further reducing the illusion of competence through the use of warnings. The goal of the warnings was to make participants aware of the illusion of competence found for backward pairs and to encourage them adjust their JOLs accordingly for those pairs. Our results showed that despite these warnings, participants continued to show a discrepancy between their JOL ratings and their actual recall rates. The purpose of Experiment 1 was to use item-specific and relational encoding tasks to improve recall rates. Forward pairs were found to be resistant against the illusion of competence, and symmetrical pairs had a smaller illusion of competence that was mediated through the use of both encoding tasks. The illusion of competence was highest for the backward and unrelated pairs which showed that the associative direction of the pair types affected recall rates. Item-specific processing was found to greatly reduce the illusion of competence for backward pairs and relational processing was found to eliminate the illusion of competence for the unrelated word pairs.

The purpose of Experiment 2 was to try and build upon the benefits found for item-specific and relational encoding at increasing JOL calibration in Experiment 1 by providing participants with a warning about the misleading nature of some of the word pairs. Participants were warned that backward pairs specifically were notorious for being misleading because the association present was weak and that participants tended to overestimate their JOLs for these pairs even though they weren’t typically well remembered. Participants were shown examples of all of the pair types and a figure depicting the illusion of competence typically found for the word pairs. The goal of this warning was to get participants to reduce their JOL ratings. Surprisingly, no main effect of warning was found, meaning that warnings were not effective in reducing the illusion of competence. The illusion of competence found for forward and symmetrical pairs was very small and there was no difference across encoding tasks. The illusion of competence found for backward and unrelated pairs replicated, with backward pairs having the highest illusion of competence. Unlike Experiment 1, the item-specific encoding group did not reduce the illusion of competence for backward pairs. The unrelated pairs, however, saw a reduction in the illusion of competence for both the item-specific and relational encoding groups.

While encoding manipulations were effective in Experiment 2, a surprise finding was that the inclusion of warnings did not reduce the illusion of competence despite extra measures being taken to ensure that the warning was effective (e.g. showing participants a graph depicting the illusion of competence) and past research showing that participants have been able to successfully adjusted their memory responses due to the inclusion of experimenter-provided instructions. One example of warnings being effective is in a study done by Chambers and Zaragoza (2001) in which participants were able to reduce memory errors when given a warning that the experimenter was intentionally misleading them. Another study done by Karanian et al.(2020) also showed that participants were able to adjust their memory when provided with a warning about potential misinformation. However, a study conducted by Ecker, Lewandowsky, and Tang (2010) showed that giving a general warning about misinformation was not as effective as a specific warning. One explanation for warnings not being effective in our study is that participants were unable to fully understand the magnitude of the warning because they had no prior understanding of JOLs and/or the illusion of competence found for backward pairs and they felt that the warning did not apply to them. One approach to correct this would to have participants rate how likely they believe that they will fall for the illusion of competence in order to evaluate whether the participants feel that they are “invincible” when providing their ratings. Finally, though we explained the four types of associates and their corresponding illusion of competence patterns and provided participants with examples and graphs, our inclusion of four pair types may have confused participants. As such, the warning may have been more effective if only comparing forward and backward pairs.

Calibration plots were created to assess how well coordinated participants’ JOLs were with their correct recall rates. The calibration plots are important because they provide us with a way to see specifically when item-specific and relational processing was beneficial in reducing the illusion of competence which provided us with more qualitative information about the specific JOL interval in which we see item-specific and relational benefits. For example, in Experiment 1 we were able to see that participants in the item-specific group were only overconfident for backward pairs above the higher JOL increments compared to those in the read group who were overconfident for backward pairs above all JOL increments over half. These patterns show us when participants begin to become overconfident, and this information could even be used to improve future studies by warning participants about using really high JOL ratings.

One limitation of our study was that Experiments 1 and 2 did not show the same results for the reduction of the illusion of competence. One potential explanation for why our results did not replicate in Experiment 2 is because warnings may have affected encoding. Participants may have been distracted by trying to adjust their JOL ratings and were thus not as focused on implementing the encoding strategy they had been told to use. One way this limitation could be addressed is by reminding participants of the encoding strategy they are supposed to be using in order to bring their attention back to the encoding manipulation. Another limitation of our study was that encoding duration could have played a role in reducing the illusion of competence. Participants were allowed to self-pace their study of the word pairs, and those in the encoding manipulation groups took longer to study the word pairs than those in the control group. So, while the encoding manipulations themselves helped participants to better remember the word pairs, participants could have also benefitted from simply studying the word pairs for a longer amount of time. This limitation could be prevented by setting a standard study time for participants in all groups.

There is still much work that can be done to try and fully eliminate the illusion of competence found for backward word pairs. In order to improve the accuracy of JOLs, future studies might could limit the number of ratings that a participant could give above 90 in order to make them more conservative in their JOL ratings. One future direction that this study could go would be to have participants take a pre-test at the beginning of the study in which they use all three of the encoding strategies. The participants could then be told, regardless of their actual score, that a particular strategy was their strongest strategy and that they should use that strategy throughout the study. This manipulation could be used to assess if participants’ opinions toward a particular strategy affect their performance. One way future studies could improve on the type of warnings given to participants would be to show participants their own performance data instead of general data. This type of warning could help to show participants the importance of the warning because they would be able to see how the illusion of competence affects them directly. Another type of relational encoding task that could be used is to have participants use the cue and target words in a sentence which could help facilitate future recall. Future studies could also implement a training program in which participants are trained on using the different types of encoding strategies or on how to identify the different types of word pairs. These training sessions have the potential to help participants improve their JOL accuracy.

The present study showed that the illusion of competence can be reduced using the item-specific/relational framework. Across both experiments, we showed that illusion of competence for backward associates can be reduced via item-specific encoding and that overestimation of unrelated pairs is reduced when participants use a relational encoding strategy. While our study did not find warnings to be effective in further reducing the illusion of competence, warnings have been shown to be effective in previous studies, so more research is needed to evaluate ways to improve warnings. These findings show that the type of encoding strategy used to study an item can have memorial benefits and that different encoding strategies can have different levels of impact depending on the context of the items studied.

**References**

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

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. doi.org/10.3758/BF03193014

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition,* *3*(2), 77-88. doi:10.1016/j.jarmac.2014.03.005

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. doi.org/10.3758/BRM.41.4.977

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. doi.org/10.3758/BF03194036

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition,* *29*(8), 1120-1129. doi:10.3758/bf03206381

Craik, F. I. M. (2002). Levels of processing: Past, present … and future? *Memory, 10*(5-6). 305-318. doi.org/10.1080/09658210244000135

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. doi.org/10.1016/S0022-5371(72)80001-X

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. doi:10.3758/s13428-018-1115-7

Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *29*(5), 813-825. doi:10.1037/0278-7393.29.5.813

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition,* *38*(8), 1087-1100. doi:10.3758/mc.38.8.1087

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. doi.org/10.1037/0278-7393.6.5.588

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. doi.org/10.3758/BF03193146

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York, NY: Psychology Press. doi:10.4324/9780203782934

Gallo, D. A., Roediger, H. L., & Mcdermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review,* *8*(3), 579-586. doi:10.3758/bf03196194

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

Gretz, M. R., & Huff, M. J. (2020). Multiple species of distinctiveness in memory? Comparing encoding versus statistical distinctiveness on recognition. *Memory,* *28*(8), 984-997. doi:10.1080/09658211.2020.1803916

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. doi.org/10.1016/j.jml.2013.05.003

Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging, 18*(4), 755-769. doi.org/10.1037/0882-7974.18.4.755

Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *39*(4), 1246-1256. doi:10.1037/a0031338

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. doi.org/10.1016/j.jml.2014.02.004

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. doi.org/10.1016/S0022-5371(81)90138-9

Karanian, J. M., Rabb, N., Wulff, A. N., Torrance, M. G., Thomas, A. K., & Race, E. (2020). Protecting memory from misinformation: Warnings modulate cortical reinstatement during memory retrieval. *Proceedings of the National Academy of Sciences,* *117*(37), 22771-22779. doi:10.1073/pnas.2008595117

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. [doi:10.1037/0278-7393.31.2.187](https://doi.org/10.1037/0278-7393.31.2.187)

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *32*(5), 1133-1145. doi:10.1037/0278-7393.32.5.1133

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.doi:10.1016/j.jml.2005.01.001

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. doi:10.1037/a0018785

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods,* *43*(3), 679-690. doi:10.3758/s13428-010-0049-5

Mccabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition,* *30*(7), 1065-1077. doi:10.3758/bf03194324

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. doi:10.1080/09658211.2020.1749283

Maxwell, N. P., & Huff, M. J. (in press). The deceptive nature of associative word pairs: Effects of associative direction on judgments of learning. *Psychological Research*, 1-19. doi.org/10.1007/s00426-020-01342-z

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. doi:10.1037/0278-7393.33.2.263

Nelson, D. L., Mcevoy, C., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition,* *28*(6), 887-899. doi:10.3758/bf03209337

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. doi:10.3758/bf03195588

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*(4), 267-270. [doi.org/10.1111/j.1467-9280.1991.tb00147.x](https://psycnet.apa.org/doi/10.1111/j.1467-9280.1991.tb00147.x)

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation,* *26*, 125-173. doi:10.1016/s0079-7421(08)60053-5

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

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

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory,* *4*(6), 592-604. doi:10.1037/0278-7393.4.6.592

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. doi:10.1037/a0038388

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review,* *14*(5), 779-804. doi:10.3758/bf03194105





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





*Figure 2.* Calibration plots as a function of pair type in Experiment 1 for participants in the Read group (top panel), Item-Specific group (middle panel), and Relational group (bottom panel). 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.* Sample data illustrating the illusion of competence for backward, symmetrical, and unrelated study pairs. This graph was provided to participants in the Experiment 2 warning group. Data pattern is modeled after Maxwell and Huff (in press).

** **

**Mean % JOL/Recall**

** **

**Mean % JOL/Recall**

** **

**Mean % JOL/Recall**

**Pair Type**

**Pair Type**

*Figure 4.* Mean JOL and recall rates as a function of pair type in the Read (top panels), Item-Specific (middle panels), and Relational (bottom panels) Warning and No Warning groups in Experiment 2. Bars represent 95% confidence intervals.

****

**Mean % JOL/Recall**

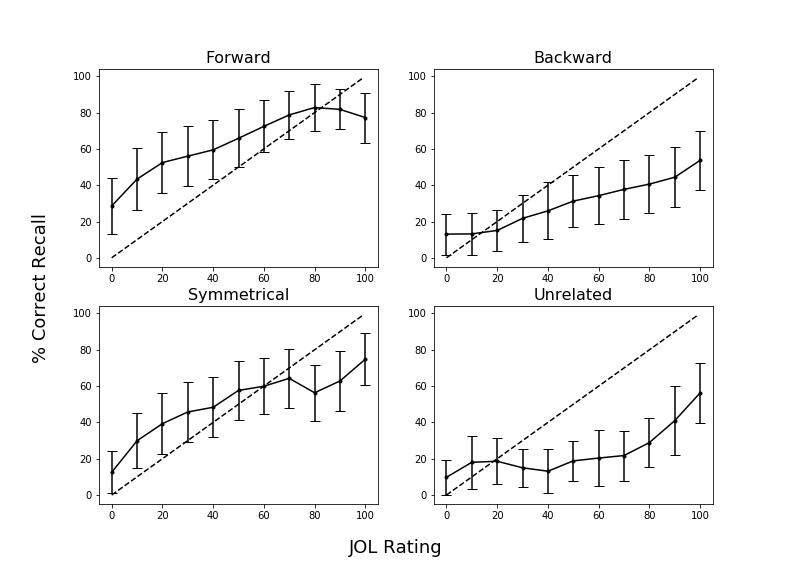
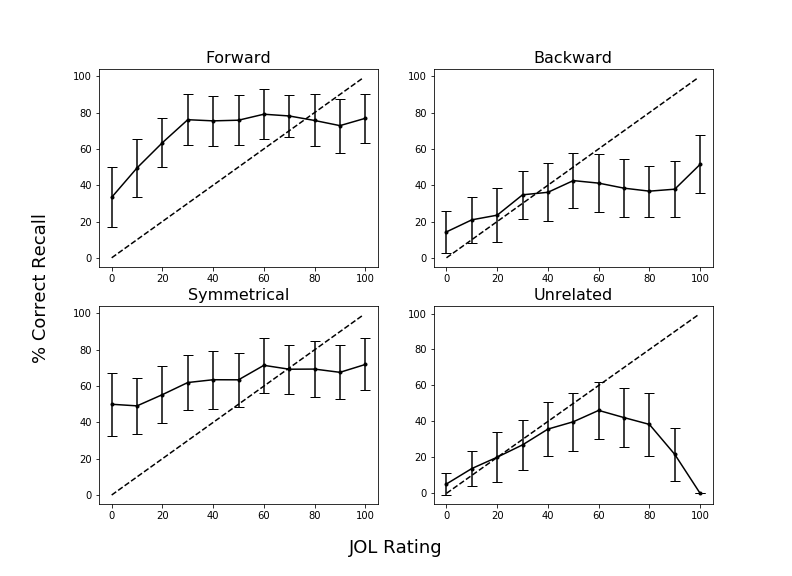
**Mean % JOL/Recall**

****

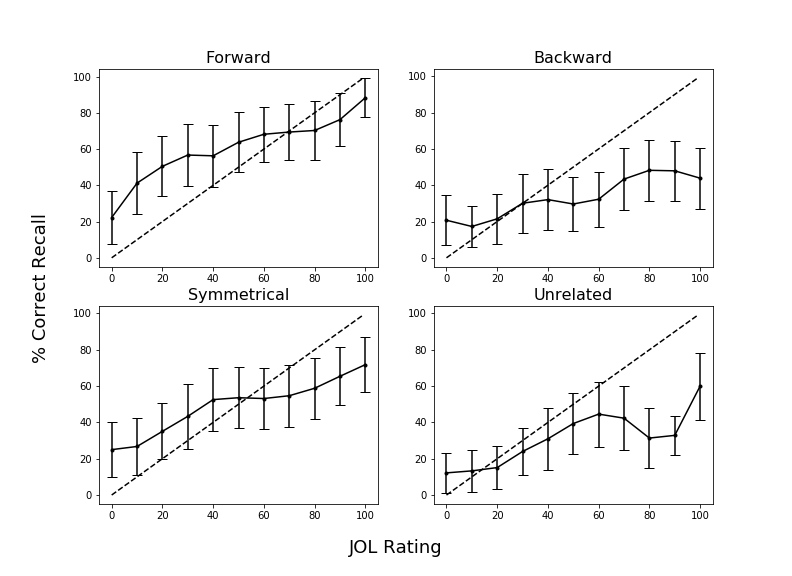
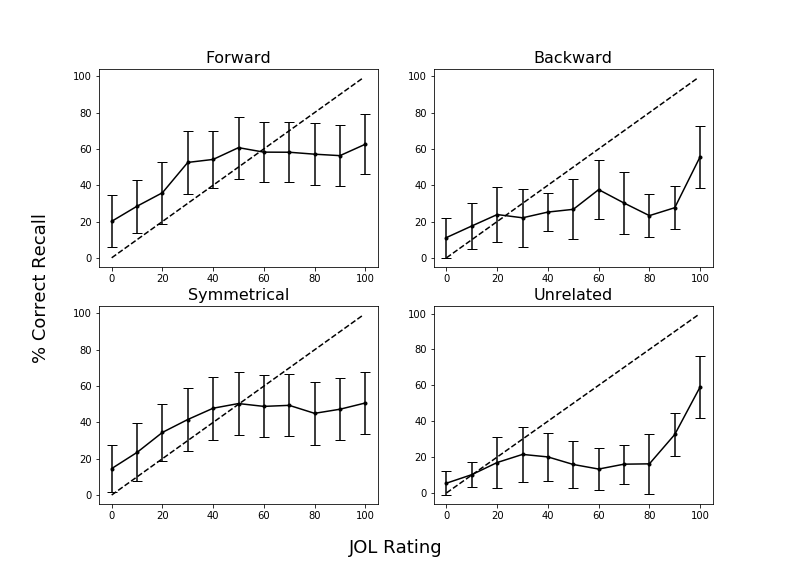
**Mean % JOL/Recall**

**Pair Type**

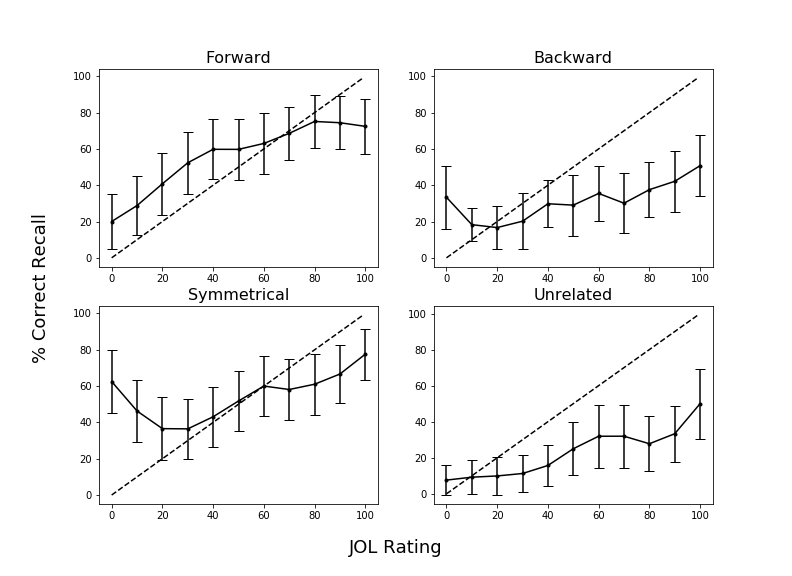
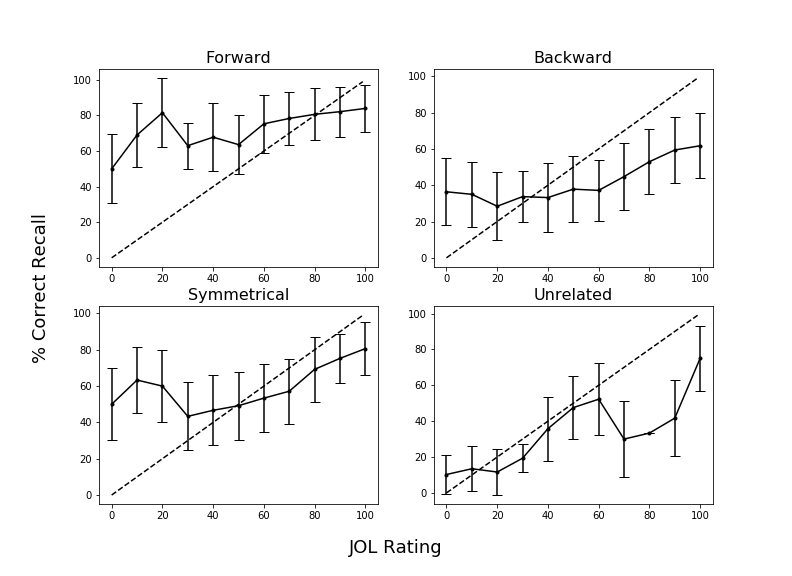
*Figure 5.* Mean JOL and recall rates as a function of pair type collapsed across warning for the read, item-specific, and relational groups in Experiment 2.



*Figure 6.* Calibration plots as a function of pair type in Experiment 2 for participants in the No-Warning Read group (top panel) and the Warning Read group (bottom panel) for Block 2 recall. 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 7.* Calibration plots as a function of pair type in Experiment 2 for participants in the No-Warning Relational group (top panel) and the Warning Relational group (bottom panel) for Block 2 recall. 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 8.* Calibration plots as a function of pair type in Experiment 2 for participants in the No-Warning Item-Specific group (top panel) and the Warning Item-Specific group (bottom panel) for Block 2 recall. 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.

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 2

*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).

1. Due to COVID-19 restrictions, the participants in Experiment 2 were sampled from both in-lab and online sources. The participant source was not found to interact with any of the results, *F*s < 1, though the vast majority of participants were recruited online. In-lab participants were tested using *E*-*Prime 3*, and online participants were tested using *Collector*, an open-source program for data collection on Psychology experiments (Garcia & Kornell, 2015). All procedural details and instructions were identical in both modalities, the only difference was the presence vs. absence of an experimenter. [↑](#footnote-ref-1)