[TITLE HERE]

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 the and individual’s knowledge of learning process. A common method used to gauge metamemory knowledge is the Judgment of Learning (JOL) task. In the 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 the target word when only provided with the cue word on a subsequent memory test. These estimates can be elicited using several types of measurement scales (e.g., 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 of the target item being 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 an easy comparison between predicted target recall (via JOLs) and the proportion of targets that are correctly recalled at test.

Although JOL ratings can be accurate and predictive (i.e., well-calibrated) with subsequent recall, may factors can affect the efficacy of JOLs on later recall. These include perceived ease in identical cue-target word pairs (Castel, McCabe, & Roediger, 2007), increased in time spent studying word pairs (Koriat & Ma’ayan, 2005), and the direction and magnitude of associative relationships 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). Additionally, we also examine 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 relationship between JOLs and correct recall relative to a control group.

Interest in the relationship between memory predictions and accuracy is not a novel area of exploration. In an early example, Arbuckle and Cuddy (1969) asked participants to study letter pairs and report whether they would or would not remember the pair, followed by a memory test of the pair with a postdiction that they were initial correct or incorrect regarding their initial prediction. Overall, this study found that participants, on average, were over 60% right in their predictions of their recall. The authors concluded 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) supports the notion that both the associative strength and direction of cue-target word pairs affects correspondence between JOL ratings at study and subsequent correct recall. Specifically, the authors delineated between types 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). 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 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). Second, *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 that in which the pair order has been flipped (i.e., backward pairs, card-credit). 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 each experiment, an *illusion of competence* was found for a posteriori pairs in which participants’ JOLs exceeded subsequent recall rates. In particular, this effect was particularly robust the backward pairs presented in their second experiment, as the target words within this pair type do not readily converge upon the cue. Thus, although participants predict that backward pairs as 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 found by Castel et al. (2007) who reported that the illusion of competence also occur when participants study 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. One explanation for this finding is that participants viewed the identical pairs to be easier to learn relative to forward associates and unrelated pairs due to their high perceived semantic similarity. As a result, participants may not have encoded the identical pairs as deeply because they thought they would be easier to recall (Castel et al., 2007).

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). 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., salt-pepper would have the same associative strength as pepper-salt), whereas forward and backward cue-target pairs have a stronger strength depending on the direction of the pair (i.e. tuna-fish (F) has a stronger associative strength than fish-tuna (B)). Furthermore, these pairs differ from identical pairs in that they have equal levels of forward and backward associative strength without needing to repeat the same word. Across four experiments, a strong illusion of competence pattern was found for backward pairs and, additionally, the illusion of competence was shown to extend to symmetrical associates, indicating that… [EXPAND]. Furthermore, Maxwell and Huff found that the illusion of competence was not contingent upon the timing of when JOLs were provided at study, as similar illusion of competence patterns were found when JOLs were provided under experimenter-paced encoding durations and when JOLs were provided after a delay—two manipulations that have been shown to enhance the calibration between JOLs and later recall (Hertzog, Dixon, Hultsch, & MacDonald, 2003; Rhodes & Tauber, 2011). Finally, Maxwell and Huff employed the use of calibration plots in which JOL ratings were plotted against their corresponding recall accuracy Nelson & Dunlosky, 1991), allowing the authors to pinpoint the JOL rating at which the illusion of competence emerged for each pair type. [SENTENCE HERE]

[NEW PARAGRAPH HERE]

**Item-Specific/Relational Framework**

Memory researchers have long known that certain study tasks are more successful at improving retention than others. Several tasks have been identified, 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 tasks propensity to encourage the processing of item-specific or relational features. According to the item-specific/relational processing framework (Hunt & Einstein, 1980; Einstein & Hunt, 1981). According to this framework, encoding tasks differ in the likelihood that they can encourage the processing unique features of study items through item-specific processing, or they can encourage the processing of shared characteristics of study items through relational processing.

Given the benefits of item-specific and relational processing on memory, the present study tested whether these encoding strategies can be used to reduce the illusion of competence found for backward and symmetrical related pairs and unrelated pairs. Specifically, Experiment 1 compares JOLs and cued-recall performance for each of the aforementioned encoding strategies to a standard read-only JOL control group who receives no explicit encoding instructions. Next, Experiment 2 tests whether combining these encoding manipulations with an explicit warning about the deceptive nature of backward, symmetrical, and unrelated study pairs further reduces the illusion of competence. Finally, across both experiments, we follow the analyses used by Maxwell & Huff (in press) by plotting participants JOL ratings against their recall rates using a series of calibration plots to gauge whether participants over/under predict subsequent recall.

**Experiment 1: Item-Specific vs Relational Encoding**

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 create an association instead of relying on the weak cues between pairs low in FAS, it was expected that this encoding manipulation would be particularly beneficial for improving recall of unrelated pairs. Finally, because item-specific 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 reducing the illusion of competence for backward and symmetrical pairs.

**Experiment 1: Item-Specific vs. Relational Encoding Instructions**

**Methods**

**Participants**

Ninety-five University of Southern Mississippi undergraduates participated in for partial course credit. All were native English speakers with normal or corrected-to-normal vision. [Add Experiment 1 participant demographics.]

**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 associate pairs (e.g., credit-card), 40 backward associate pairs (e.g., card-credit), 40 symmetrical associate 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.

Associative 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 ability, 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 across 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 that of 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 different 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 when providing their ratings 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 or how cats and turtles 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 or how cats are mammals, but turtles are reptiles. Participants only saw one type of study instruction. After the 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 they 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 had completed the second study/test block, they were debriefed and awarded credit for their participation. 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. Partial-eta squared (*η*p2) and Cohen’s *d* eﬀect sizes were included for signiﬁcant Analyses of Variance (ANOVAs) and *t*-tests, respectively. A sensitivity analysis using *G**\*Power* (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that the sample had sufficient power (.80) to detect a small effect size (Cohen’s *d* = 0.27) or larger. For all analyses, a *p* < .05 signiﬁcance level was used unless noted otherwise. [PBIC EXPLAINATION HERE?]

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 conducted to test for differences between mean JOL ratings and recall rates across the four pair types and at each of the three encoding manipulations. 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* = XX). An effect of encoding group was also found, *F*(2, 85) = 5.40, *MSE* = 814.98, *ηp2* = .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 ≥ XX, with the exception of the comparison between the relational and item-specific groups, which was non-significant, *t* < 1. 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 pair types differed statistically, *t*s ≥ 2.68, *d*s ≥ XX.

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, *ηp2* = 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). Below, two observations of particular interest are reported. First, the illusion of competence replicated for backward pairs in 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* = XX). For the item-specific group, JOLs also exceeded recall (69.57 vs 58.97, *t*(28) = 2.16, *SEM* = 5.12, *d* = XX), 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 lower rate (71.54 vs 50.49, *t*(30) = 5.41, *SEM* = 4.05, *d* = XX).

For the unrelated pairs, the illusion of competence was similar in the read group (24.78 vs 14.77, for JOLs and recall rates, respectively, *t*(27) = 3.23, *SEM* = 3.26, *d* = XX) in the item-specific encoding group (40.65 vs 14.35, *t*(28) = 5.71, *SEM* = 4.81, *d* = XX) and Read groups. However, the illusion of competence was eliminated in the relational group (36.62 vs. 32.51, *t* < 1), indicating that relational encoding provides a unique benefit on unrelated pairs by improving the correspondence between JOLs and subsequent recall.

We next assessed the correspondence between the JOLs provided at study and correct recall for each of the pair types using a series of calibration plots. 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 Figures 2-4, split by each encoding manipulation. 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., 30% JOL and 30% correct recall). Using these plots, overestimations (i.e., data points falling below the calibration line) were found to emerge at different JOL ratings across each pair type. Furthermore, these patterns were moderated by the encoding strategy employed at study.

Starting with the participants in the silent reading group, for unrelated pairs, JOL overestimations were observed at nearly all JOL ratings (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%). 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 participants in the item-specific encoding 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 encoding 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.

**Discussion**

[WORDS HERE]

**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 reduce (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). 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 (2005)… [EXPAND ON THIS]

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. 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 the 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 the following exceptions. Specifically, for participants assigned to the warning groups, participants were given a message on their screen before the second study block about the illusion of competence found for backward, symmetrical, and unrelated pairs. They were then presented data modeled after Maxwell and Huff (in press) supporting this warning, which showed (see Figure 5 for the graph participants viewed; the exact warning instructions have been made available at [OSF LINK])Participants were told that previous studies showed that people tended to give higher JOL ratings for backward pairs than they were able to recall and that they should be cautious when providing JOLs for backward pairs. This warning served to hopefully improve participants’ accuracy in their JOL ratings.

**Results**

Data were initial screened for missing responses and outliers as in Experiment 1. In the following analyses, because the warning manipulation only occurred on the second block in the warning group, analyses in both warning and no warning groups only included JOLs and recall responses on the second block.

Figure 6 reports JOL and recall rates as a function of pair type, encoding group, and warning group in Experiment 2. In the analyses we first examined the effects of the illusion of competence 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 7), but for concision, do not include warning as a factor in the analyses below.

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 group, *t*(70) = 1.69, *SEM* = 2.20, *p* = .10, *p*BIC = .68) and relational group, *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. The first set of these plots were used to assess calibration for participants in the no warning group (Figures 8-10). Starting with participants who completed the silent reading task without receiving a 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%). These patterns were confirmed by effects of Pair Type, *F*(3, 105) = 29.20, *MSE* = 42847.95, *η*p2= .46, JOL Increment, *F*(10, 350) = 5.79, *MSE* = 7530.79, *η*p2 = .14, and a significant interaction, *F*(30, 1050) = 2.89, *MSE* = 2463.93, *η*p2 = .08.

Next, for participants in the no warning item-specific encoding group, 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%. Significant effects of Pair Type, *F*(3, 111) = 32.71, *MSE* = 53181.60, *η*p2= .47, JOL Increment, *F*(10, 370) = 6.34, *MSE* = 9233.50, *η*p2 = .15, and a significant interaction, *F*(30, 1110) = 2.69, *MSE* = 2.69, *η*p2 = .07, again confirmed these patterns.

Finally, for 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%. Once again, all patterns of overestimation were confirmed by significant effects of Pair Type, *F*(3, 99) = 37.00, *MSE* = 41768.89, *η*p2= .53, JOL Increment, *F*(10, 330) = 13.69, *MSE* = 17261.65, *η*p2 = .29, and a significant interaction, *F*(30, 990) = 5.07, *MSE* = 4045.62, *η*p2 = .13.

Next, we created a set of calibration plots for participants who receive the warning prior to beginning Block 2 (Figures 11-13). For participants completing the silent reading task, 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%. These patterns were confirmed by effects of Pair Type, *F*(3, 96) = 31.55, *MSE* = 36606.43, *η*p2= .50 JOL Increment, *F*(10, 320) = 4.77, *MSE* = 8029.19, *η*p2 = .13, and a significant interaction, *F*(30, 960) = 2.05, *MSE* = 2066.35, *η*p2 = .06.

Next, when participants engaged in item-specific encoding, 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, 99) = 57.107, *MSE* = 44116.87, *η*p2= .63, JOL Increment, *F*(10, 330) = 15.33, *MSE* = 20404.38, *η*p2 = .32, and a significant interaction, *F*(30, 990) = 5.07, *MSE* = 4242.61, *η*p2 = .13, again confirmed these patterns.

Finally, for participants completing the relational encoding task, 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, 84) = 28.88, *MSE* = 33597.49, *η*p2= .51, JOL Increment, *F*(10, 280) = 13.29, *MSE* = 15793.53, *η*p2 = .32, and a significant interaction, *F*(30, 840) = 3.78, *MSE* = 3143.54, *η*p2 = .12.

**Discussion**

[WORDS HERE SUMMARIZING EX2 AND IMPLICATIONS]

**General Discussion**

[WORDS HERE SUMMARING FINDINGS FROM EACH EXPERIMENT]

[IMPLICATIONS AND IMPORTANT THINGS]

[PARAGRAPH ON LIMITATIONS AND FUTURE DIRECTIONS]

**Conclusion**

[FINAL SUMMARY AND TAKE HOME MESSAGE]

**References**

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

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

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

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.

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), 175–91. [doi:10.3758/BF03193146](https://doi.org/10.3758/BF03193146)

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.

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)

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*, 267–270.

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

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*, 131-48. 10.1037/a0021705.

[EX 1 BAR CHART HERE]



*Figure 2.* Calibration plots as a function of pair type in Experiment 1 for participants in the read group. 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.* Calibration plots as a function of pair type in Experiment 1 for participants in the item-specific encoding group. 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 4.* Calibration plots as a function of pair type in Experiment 1 for participants in the relational encoding group. 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 5.* 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 6.* 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**

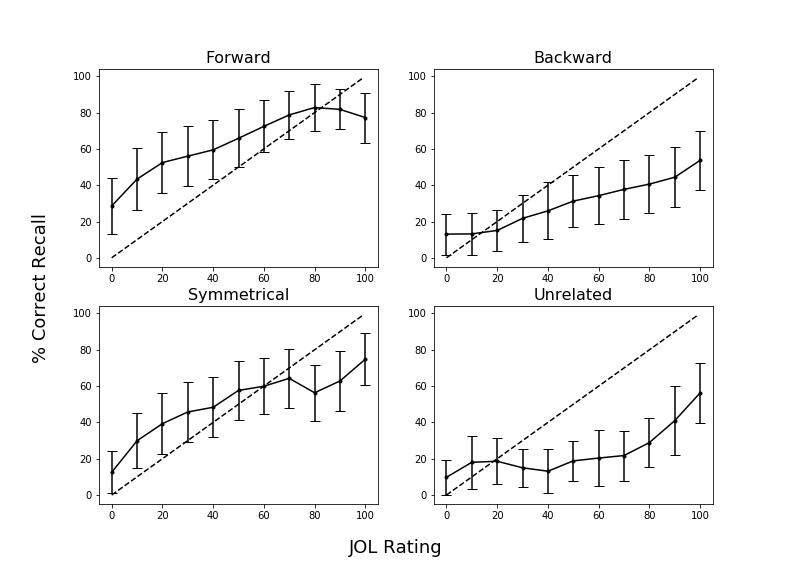
**Mean % JOL/Recall**

****

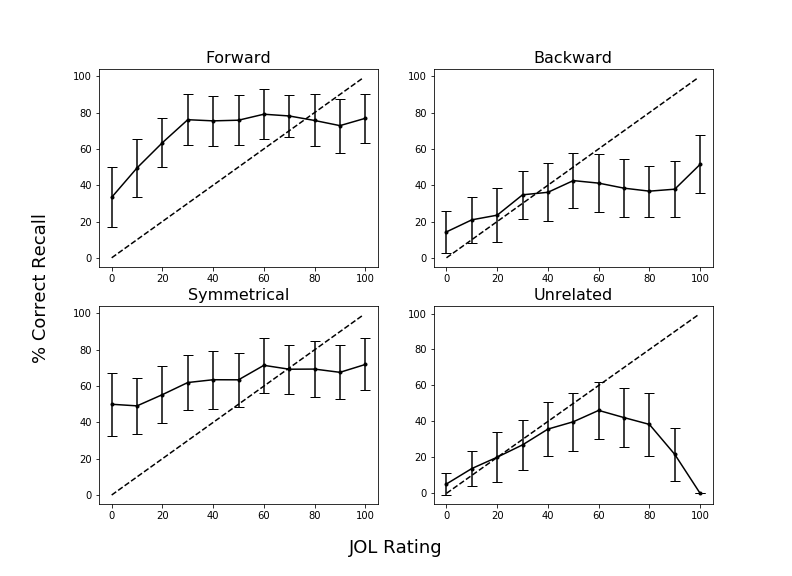
**Mean % JOL/Recall**

**Pair Type**

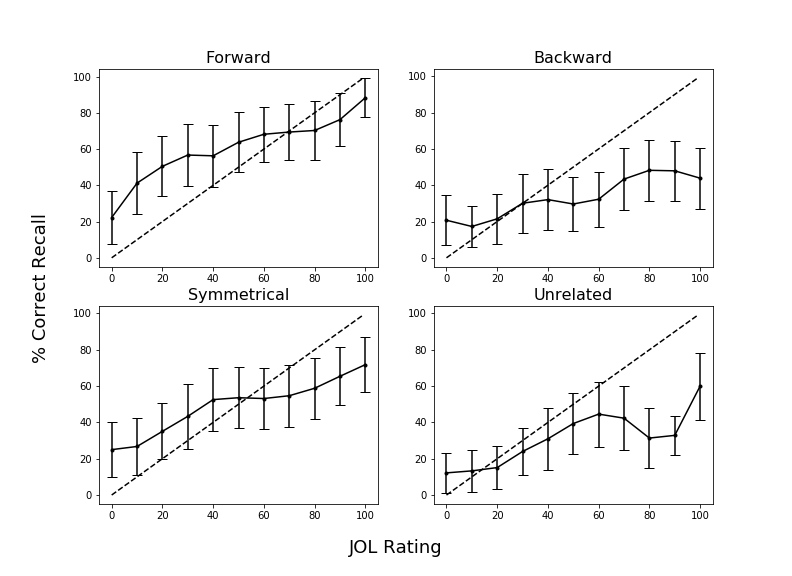
*Figure 7.* Mean JOL and recall rates as a function of pair type collapsed across warning for the read, item-specific, and relational groups in Experiment 2.



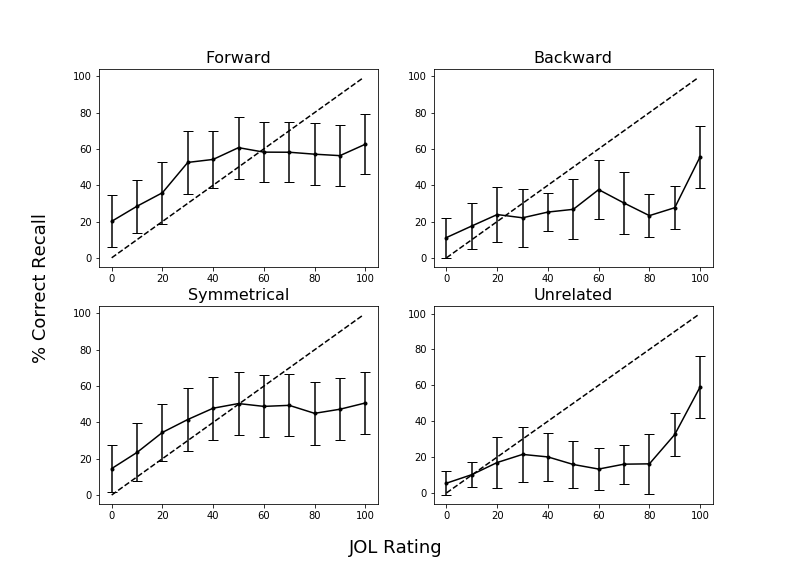
*Figure 8.* Calibration plots as a function of pair type in Experiment 2 for participants in the read group who did not receive a warning before block 2. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued-recall. Overconﬁdence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% conﬁdence interval.



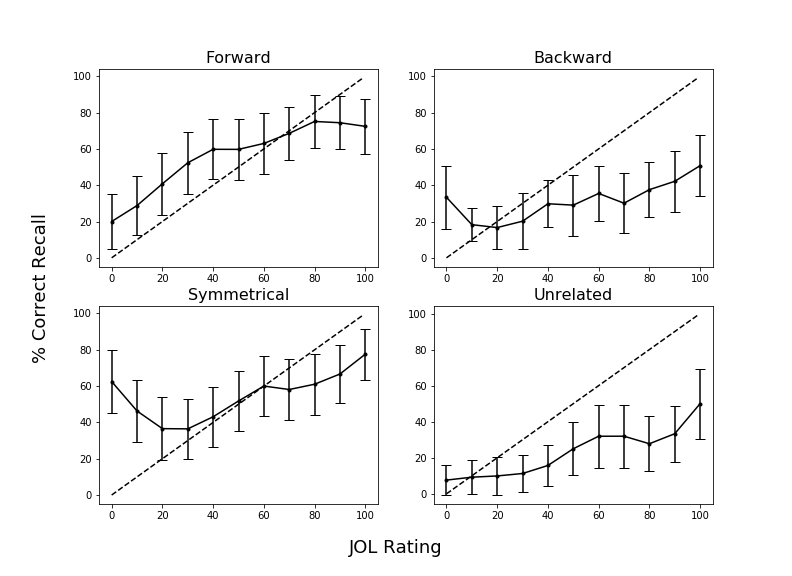
*Figure 9.* Calibration plots as a function of pair type in Experiment 2 for participants in the item-specific encoding group who did not receive a warning before block 2. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued-recall. Overconﬁdence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% conﬁdence interval.



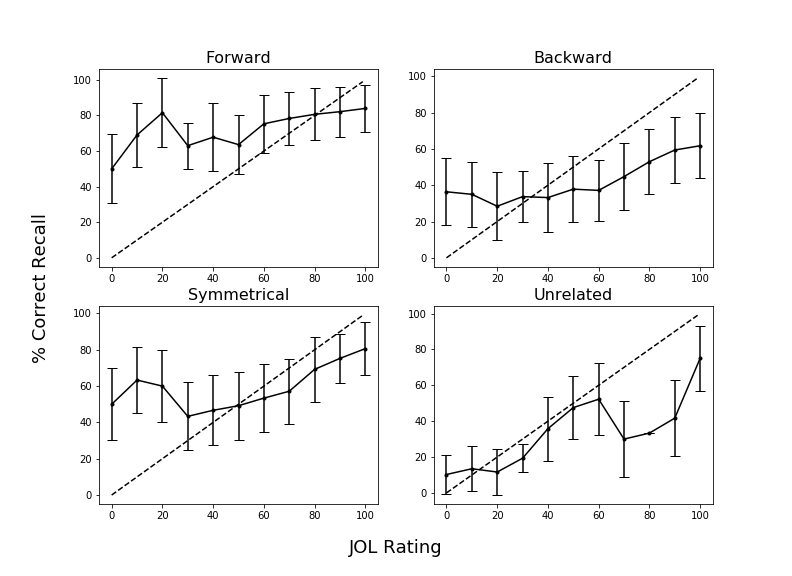
*Figure 10.* Calibration plots as a function of pair type in Experiment 2 for participants in the relational encoding group who did not receive a warning before block 2. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued-recall. Overconﬁdence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% conﬁdence interval.



*Figure 11.* Calibration plots as a function of pair type in Experiment 2 for participants in the read group who received a warning before block 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.



*Figure 12.* Calibration plots as a function of pair type in Experiment 2 for participants in the item-specific encoding group who received a warning before block 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.



*Figure 13.* Calibration plots as a function of pair type in Experiment 2 for participants in the relational encoding group who received a warning before block 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.

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* < 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)