**Methods**

**Participants**

Ninety-five University of Southern Mississippi undergraduates participated in this study for partial course credit. All participants were native English speakers with normal or corrected-to-normal vision. [Add Experiment 2 participant demographics.]

**Materials**

The stimuli used were one-hundred-eighty associative word pairs originally used by Maxwell and Huff (under review). These word pairs were taken from the University of South Florida Free Association Norms (Nelson et al., 2004). These pairs 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 buffer pairs that were not tested to control for primacy and recency effects. The word pairs were divided evenly into two study blocks, each containing 20 forward, backward, unrelated, and symmetrical pairs and 10 buffer pairs, for a total of 90 pairs in each list. All participants saw both lists presented separately in study-test blocks, and the order of the lists 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. 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). 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 randomly anew for each participant.

**Procedure**

The design for this study was modeled after Maxwell and Huff (under review). For Experiment 1, all participants were run individually on computers using the *E-Prime* 3 software (Psychology Software Tools, Pittsburgh, PA). Participants were randomly assigned to one of three different encoding groups: Item-specific, Relational, or a Read-only control. For each study group, participants were told that they would study a series of word pairs and that their memory for these pairs would be tested later. The cue word was always presented on the left and the target word was always presented 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 which served as the control condition. For the Relational group, participants were instructed to study the word pairs by thinking about how the words in each pair were related. The example given was if a participant saw the pair “Cat-Turtle”, they may 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. The example given was if a participant saw the pair “Cat-Turtle”, they 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 to give their best guess as to what the target word was if they were unable to retrieve the target word, but participants were able to skip to the next cue by pressing enter if they could not remember. After the first cued-recall test was finished, participants then went through a second study/test block with the same encoding instructions as the first block. Once participants had completed the second block, they were debriefed on the study. Participants generally completed the experiment in under 1 hour.

Due to COVID-19 restrictions, the participants in Experiment 2 were either ran individually on an in-lab computer using the *E-Prime* 3 software (Psychology Software Tools, Pittsburgh, PA), collected online through the Prolific platform (citing?), or collected online through the university’s SONA program (citing?). The procedure for Experiment 2 was identical to Experiment 1 with exception of an added warning portion during the second block. Participants were randomly assigned to either the Warning or No Warning group. For those in the No Warning group, no procedures were changed from Experiment 1. For the Warning group, participants were given a message before the second study block about the illusion of competence found for backward, symmetrical, and unrelated pairs and were shown data from Experiment 1 supporting this warning. [insert quotes of what the warning said]. This warning served to hopefully improve participants’ accuracy in their JOL ratings.

**Results**

Before conducting the analyses, all data were screened for missing responses and outliers (i.e., JOLs outside of the 0-100 range). Skipped recall responses were scored as incorrect, but correct items that were misspelled were counted 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. Figure 1 plots mean JOL ratings and cued-recall rates for each word pair type. A sensitivity analysis using *G**\*Power* (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that this sample had sufficient power (.80) to detect a small effect size (Cohen’s *0* = 0.27) or larger. For all analyses, a *p* < .05 signiﬁcance level was used unless noted otherwise. [Possibly need to run another *G\*Power* analysis and get another p value?]

A 2 (Measure: JOL vs. Recall) × 3 (Encoding Manipulation: Item-Specific vs. Relational vs Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed measures 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. First, a significant effect of measure was found, *F*(1, 85) = 18.79, *MSE* = 694.46, *η*p2 = .07, which indicated that across both the pair types and encoding manipulations, JOL ratings exceeded later recall rates (62.66 vs. 54.19, *t*(87) = 4.18, *SEM* = 2.06, *d* = XX). Next, an effect of encoding manipulation was detected, *F*(2, 85) = 5.40, *MSE* = 814.98, *ηp2* = .05, in which JOL ratings/recall rates were significantly higher when participants studied using either relational (61.44) or item-specific encoding strategies (60.12) relative to the read only condition (53.33). All comparisons differed significantly, *t*s ≥ XX, *d*s ≥ XX, with the exception of the comparison between relational and item-specific encoding, which was non-significant, *t*(56) = 0.47, *p* = .64. Finally, this model revealed a significant effect of pair type, *F*(3, 255) = 766.58, *MSE* = 107.66, *η*p2 = 0.58, in which JOL ratings/recall rates were greatest 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 Direction confirmed that the illusion of competence replicated across each encoding manipulation, *F*(2, 85) = 5.21, *MSE* = 107.66, *ηp2* = 02. Critically, a significant three-way interaction was detected, *F*(6, 255) = 15.56, *MSE* = 87.42, *η*p2 = .04, in which the illusion of competence differed based on the encoding strategy used (See Figure 1 for comparison between encoding manipulations). Below, two observations of particular interest are reported. For completeness, all comparisons for each encoding manipulation are reported in Table XX.

First, the illusion of competence replicated for backward pairs within each of the three encoding manipulations. A robust illusion of competence was detected in the Read condition in which JOLs exceeded later recall accuracy (68.58 vs. 37.78, *t*(27) = 9.44, *SEM* = 3.41, *d* = XX). For the Item-Specific encoding, the difference between JOLs and Recall was significant (69.57 vs 58.97, *t*(28) = 2.16, *SEM* = 5.12, *d* = XX), though to a lesser magnitude relative to the Read condition. A similar pattern was observed when participants were asked to study using the Relational encoding strategy (71.54 vs 50.49, *t*(30) = 5.41, *SEM* = 4.05, *d* = XX). [ADD SOMETHING HERE ABOUT WHAT THAT MEANS/IMPLICATIONS] These findings indicate that, compared to just reading the word pairs, the Item-Specific and Relational study strategies are either helping participants better remember the word pairs, thus increasing recall rates, or are helping participants to better calibrate their JOL ratings.

For the unrelated pairs, the illusion of competence occurred in the Item-Specific encoding (40.65 vs 14.35, *t*(28) = 5.71, *SEM* = 4.81, *d* = XX) and Read conditions (24.78 vs 14.77, *t*(27) = 3.23, *SEM* = 3.26, *d* = XX). However, the use of Relational encoding removed the illusion of competence for unrelated item pairs, as JOLs and recall were well calibrated for this pair type within this encoding condition (36.62vs 32.51, *t*(30) = 0.95, *SEM* = 4.52, *p* = 0.35). [ADD SOMETHING HERE ABOUT WHAT THAT MEANS/IMPLICATIONS] These findings indicate that the Relational study strategy helped participants to create an association between the two unrelated words, and were thus better able to recall the target word at test.

[NICK STILL NEEDS TO GET THIS DATA]

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 Figure 2. Each plot includes a calibration line which reflects perfect correspondence between JOL ratings and correct recall (e.g., 30% JOL and 30% correct recall). Overestimations (i.e., data points that fall below the calibration line) were found to emerge at different JOL ratings for each pair type. For unrelated pairs, JOL overestimations occurred across nearly all JOL ratings (JOLs > 20%), however overestimations emerged later for associative pairs. For backward pairs, overestimations occurred at JOLs greater than 60%, for symmetrical pairs, overestimations occurred at JOLs greater than 80%, and for forward pairs, overestimations were only found at the highest JOL ratings (90-100%). These patterns were confirmed by effects of Pair Type, *F*(3, 81) = 71.70, *MSE* = 1471.60, *η*p2= .73, JOL Increment, *F*(10, 270) = 6.35, *MSE* = 1204.60, *η*p2 = .19, and a significant interaction, *F*(30, 810) = 1.80, *MSE* = 879.71, *η*p2 = .06. Thus, evidence for illusions of competence were found across pair types, however overestimations only emerged at the highest JOL ratings for forward associates.

A 3 (Encoding Manipulation: Item-Specific vs. Relational vs Read) × 2 (Warning: Warn vs. No Warn) × 2 (Measure: JOL vs. Recall) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed measures ANOVA was conducted to test [I need some help wording this] for differences between mean JOL ratings and recall rates across the four pair types and at each of the three encoding manipulations.

* Main effect Measure: *F*(1, 121) = 94.60, *MSE* = 1001.56, *p <* .001, *ηp2* = .31
* Main effect Pair Type: *F*(3, 633) = 989.83, *MSE* = 190.69, *p <* .001, *ηp2* = .82
* Main effect Warning: *F*(1, 211) = .71, *MSE* = 1753.48, *p <* .40
* Main effect Study Group: *F*(2, 211) = 4.03, *MSE* = 1753.48, *p <* .02, *ηp2* = .04

**References**

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

*Figure 1*. Comparison of mean JOL ratings and recall rates across each pair type for each of the three study conditions. Error bars represent 95% confidence intervals.