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

**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 1 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, symmetrical, and unrelated 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). All participants were run 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, or Relational. 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. 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 completed a second study/test block with the same encoding instructions as the first . Once participants had completed the second block, they were debriefed on the study. Participants generally completed the experiment in under 1 hour.

**Results**

Before conducting the 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 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.

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. 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 manipulation was also found, *F*(2, 85) = 5.40, *MSE* = 814.98, *ηp2* = .05, in which JOL ratings/recall rates were significantly higher for relational (61.44) and item-specific encoding strategies (60.12) relative to the read-only group (53.33). All comparisons differed significantly, *t*s ≥ XX, *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 each encoding group, *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 between encoding manipulations). Below, two observations of particular interest are reported. 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 group 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). 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 groups (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). 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.

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.

**Experiment 2**

The goal of Experiment 2 was to further reduce the illusion of competence found for backward word pairs. Warnings have been shown in previous studies to reduce memory errors in eyewitness testimonies. In a study done by Chambers and Zaragoza (2001), they found that eyewitness suggestibility errors were reduced when participants were given a warning that the experimenter was trying to mislead them. Thus, warnings could potentially help reduce meta-memory error and improve accuracy. We applied this idea to Experiment 2 by including a pre-study warning before the second study block that warned participants about the illusion of competence found for backward word pairs and advised them to be cautious when providing their JOL ratings.

Koriat and Bjork (2005) looked at the effect that foresight bias had on JOL ratings, and our warning will serve the same purpose in that participants will be thinking about their JOL ratings before they give them. An important different, however, is that we will be combining the warning with the Item-Specific and Relational encoding strategies to see if the warnings are more effective under these conditions.

**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](http://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 were identical to those used in Experiment 1. 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 on their screen 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. 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. Participants were shown a graph similar to Figure XX to show the gap between JOL ratings and correct recall 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 which included data imputations to minimize the total number of JOL trials excluded. 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 XX, 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 XX), 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, which indicated that 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, which indicated that JOL/recall rates were lower in the read group than the relational group (47.94 vs. 53.88), *t*(131) = 2.48, *SEM* = 2.38, *d* = 0.43, but equivalent relative to the item-specific group (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 across groups, 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.

**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

** **

**Mean % JOL/Recall**

** **

**Mean % JOL/Recall**

** **

**Pair Type**

**Mean % JOL/Recall**

**Pair Type**

*Figure XX.* 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**

**Pair Type**

**Pair Type**

****

**Mean % JOL/Recall**

**Pair Type**

*Figure XX.* Mean JOL and recall rates as a function of pair type

1. 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?). [↑](#footnote-ref-1)