

Fall 12-2021

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

Emily Cates

Follow this and additional works at: https://aquila.usm.edu/honors_theses



Part of the [Psychology Commons](#)

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

by

Emily Cates

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

December 2020

Approved by:

Mark Huff, Ph.D., Thesis Advisor, School of
Psychology

Sara Jordan, Ph.D., Director, School of Psychology

Ellen Weinauer, Ph.D., Dean
Honors College

ABSTRACT

Metamemory, or the ability to understand the capacities of one's own memory, is an important part of the learning process. One method for assessing metamemory is through the Judgment of Learning (JOL) task in which participants are asked to judge the likelihood of correctly remembering a target word in a cue-target word pair when only presented with a cue word at test. The associative direction of the cue-target pair has been shown to affect the accuracy of JOLs. Unlike forward pairs (e.g., credit-card), in which JOLs accurately predict recall, an illusion of competence has been reported for backward associates (e.g., card-credit), symmetrical associates (e.g., salt-pepper), and unrelated pairs (e.g., artery-bronze) in which JOLs overestimate later recall. The present study evaluates whether the illusion of competence pattern can be reduced or eliminated when participants use an item-specific or relational encoding strategy relative to reading (Experiment 1), and whether these encoding tasks are aided by warning participants about the illusion prior to study (Experiment 2). Across experiments, item-specific and relational encoding were found to reduce the illusion of competence for backward and unrelated pairs; however, warnings did not improve JOL estimations. Thus, the method of encoding, but not warnings, can facilitate JOL accuracy.

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

DEDICATION

This thesis is dedicated to my loving fiancé, close friends, and family. I am forever grateful for their love and support that has carried me through the roughest parts of my college years.

And to my nephew Jase, for whom I would move the world.

ACKNOWLEDGMENTS

I want to thank the following individuals for their support and input, without which this thesis would not have been possible. I would like to thank those at the Honors College for their support and assistance. Stacey Ready, Dr. Sabine Heinhorst, Amy Benoit-Warlick, Serena Buckley, Dr. Ellen Weinauer, and many others provided me with support and advice that made my thesis possible. I would like to thank my thesis advisor Dr. Huff for his immense support and guidance through the entire process. Dr. Huff both encouraged and challenged me to put my best work forward and to strive for excellence. I would also like to thank my graduate student Nick Maxwell who provided countless hours of support and advice. Nick taught me how to run studies and collect data and was always available to answer questions and provide feedback. Last, but not least, I would like to thank The University of Southern Mississippi for providing me with the opportunities to be successful.

TABLE OF CONTENTS

<u>ABSTRACT</u>	iv
<u>DEDICATION</u>	v
<u>ACKNOWLEDGMENTS</u>	vi
<u>LIST OF TABLES</u>	ix
<u>LIST OF ILLUSTRATIONS</u>	x
<u>LIST OF ABBREVIATIONS</u>	xi
<u>CHAPTER I: Introduction/Background</u>	1
<u>Item-Specific/Relational Processing on Memory Performance</u>	6
<u>CHAPTER II: Experiment 1: Item-Specific Versus Relational Encoding Instructions</u>	9
<u>Methods</u>	9
<u>Participants</u>	9
<u>Materials</u>	10
<u>Procedure</u>	11
<u>Results</u>	13
<u>Discussion</u>	19
<u>CHAPTER III: Experiment 2: Item-Specific Versus Relational Encoding with Warnings</u>	22
<u>Methods</u>	24
<u>Participants</u>	24

<u>Materials and Procedure</u>	25
<u>Results</u>	25
<u>Discussion</u>	29
<u>CHAPTER IV: General Discussion</u>	31
<u>CHAPTER V: Conclusion</u>	37
<u>CHAPTER VI: Tables and Figures</u>	38
<u>APPENDIX A: Supplemental Analyses</u>	48
<u>APPENDIX B: IRB Approval Letter</u>	53
<u>REFERENCES</u>	54

LIST OF TABLES

<u>Table 1: Mean Associative Strength Summary Statistics Forward, Backward, and Symmetrical Pairs</u>	38
<u>Table 2: Summary Statistics for Cue and Target Concreteness, Length, and Frequency Item Properties as a function of Pair Type</u>	39

LIST OF ILLUSTRATIONS

Figure 1	40
Figure 2	41
Figure 3	43
Figure 4	44
Figure 5	45
Figure 6	46

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
BAS	Backward Associative Strength
DRM	Deese/Roediger-McDermott [paradigm]
FAS	Forward Associative Strength
JOL	Judgement of Learning

CHAPTER I: **Introduction/Background**

Successfully monitoring the progress of one's learning is paramount for improving retention when studying 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 the 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 asked to estimate the likelihood that they can 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). JOLs, however, are more often elicited via a continuous 0 to 100 scale representing the percent likelihood that the target item will be successfully recalled at test (e.g., 100% = definitely would remember; 0% = definitely would not remember). The use of a 100-point scale is beneficial as it 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 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 fluent due to word repetitions (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 word elicits the target at test or vice versa) and by testing whether encoding tasks that emphasize the shared or distinctive characteristics of the word pairs through relational and item-specific encoding tasks, respectively, can improve the 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 letter-number pairs (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. Arbuckle and Cuddy reported that participants correctly predicted later recall for an average of 67% of trials and correctly postdicted their responses for 88% of trials, leading the authors to conclude that participants 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, Koriat and Bjork (2005) have shown that aspects of the associative relationship between cue-target study pairs, such as the direction and the strength of the relationship, can affect JOL accuracy. Specifically, the authors delineated between two types of associations thought to influence the relationship between JOLs and recall. First, a priori associations refer to associations in the forward direction (e.g., credit-card, stork-baby). The strength of these pair types is based on the likelihood that the cue word 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 such as card-credit, baby-stork, etc.). 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) evaluated JOL accuracy when participants studied 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 semantically related a posteriori pairs that shared no association based on the norms (Experiment 3). Across experiments, a posteriori pairs showed an *illusion of competence* pattern in which JOLs exceeded subsequent recall rates, indicating that participants overpredicted the likelihood that they would later recall the target word. This pattern was particularly robust for a posteriori 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). In their experiment, the correspondence between JOLs and subsequent recall was examined when participants studied and provided JOLs for strongly and weakly related forward associates, unrelated items, and identical cue-target word pairs. Overall, an illusion of competence emerged for identical word pairs in which JOLs exceeded subsequent recall rates. The authors ascribed this pattern to the identical pairs being easier to learn, and therefore, more fluent relative to both forward and unrelated pairs given identical pairs were repeated items. As a result, participants may not have encoded identical pairs as deeply because they thought they would be easier to recall given the cue word was perfectly predictive of the target.

More recently, Maxwell and Huff (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 differ from forward and backward pairs in that the associative strength between the cue and target word are equivalent in both directions (i.e., on-off would have the same associative strength as off-on). For forward and backward pairs, however, the association is stronger in one direction than the other (i.e., tuna-fish is strongly associated in the forward direction, but has a weaker association in the backward direction, fish-tuna). Across four experiments, Maxwell and Huff (in press) found a robust illusion of competence pattern for backward pairs and, additionally, the illusion of competence was extended to

symmetrical associates, suggesting that the bidirectional association found for symmetrical pairs is not sufficient for the cue word to regularly illicit the target word. Maxwell and Huff also suggested that participants may be using both the forward and backward associations when studying 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 can affect JOL accuracy, even when associative strength is matched across pair types.

An additional contribution of Maxwell and Huff's (in press) study was the use of calibration plots in which JOL ratings for pairs were rounded to each 10% interval and plotted against their corresponding recall accuracy (see Nelson & Dunlosky, 1991). Calibration plots are useful because they 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. Across experiments, Maxwell and Huff found that forward and symmetrical pairs were generally well-calibrated at JOL ratings below 80%, but for backward and unrelated pairs, an illusion of competence pattern emerged at JOL ratings greater than 30%. 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 but occurred at much lower JOL ratings for backward and unrelated pairs.

Given that the illusion of competence can be found diffusely across pair types depending upon the JOL rating, the goal of the present study was to examine methods that could potentially improve the accuracy of JOLs on subsequent 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 which now follows.

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 promote memory as “deep” tasks, while less successful tasks that focus on surface or perceptual features of study items are referred to 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 via item-specific processing, or through the processing of shared characteristics of study items via relational processing. Thus, item-specific processing entails having participants focus on the unique features of items at study (e.g., for the pair cat-turtle, cats are mammals and turtles are reptiles, cats have fur and turtles have shells, etc.). Relational processing entails having participants focus on the shared features (e.g., cats and turtles are animals, both can be

kept as pets, etc.). These types of processing qualitatively affect encoding strategies by changing how information encoded. Many studies have found differential memory benefits for item-specific and relational encoding tasks. For example, McCurdy, Sklenar, Frankenstein, and Leshikar (2020), showed that relational processing facilitated the generation effect for lower-constraint tasks (i.e. generating a target word in the presence of a cue) potentially because participants had to create a relationship between the two words. Relational processing could therefore be beneficial in studying unrelated word pairs since there is no existing relationship between the words. Separately, Huff and Bodner (2014) found that item-specific tasks were more likely to improve recall and recognition when study items were strongly related, but not when study items were weakly related. Similarly, relational tasks were more likely to improve recall and recognition when study items were weakly rather than strongly related (argued to be evidence for encoding variability of processing). Thus, although item-specific and relational processing tasks are generally classified as “deep” tasks according to the levels-of-processing framework, their relative memory benefits are affected by the association between study materials.

Given the interactive benefits of item-specific and relational encoding with different associative materials, 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 encode cue-target pairs using either item-specific or relational tasks

relative to a standard read-only control task across forward, backward, symmetrical, and unrelated pair types. In Experiment 2, it was then examined whether combining item-specific and relational tasks with an explicit warning about the illusion of competence could further reduce JOL miscalibration. Finally, in both experiments, analyses used by Maxwell and Huff (in press) were followed 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.

CHAPTER II: 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 was tested whether item-specific/relational encoding tasks 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 item-specific/relational encoding tasks 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 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 particularly beneficial for backward and unrelated pairs where the cue is less effective at prompting 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 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 conducted using *G*Power* (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that this sample had sufficient power (.80) to detect a small-to-medium effect size (Cohen's $d = 0.27$) or larger.

Materials

Stimuli included 180 associative word pairs initially used by Maxwell and Huff (in press). Pairs were taken from the University of South Florida Free Association Norms (Nelson et al., 2004). These consisted of 40 forward pairs (e.g., credit-card), 40 backward pairs (e.g., card-credit), 40 symmetrical pairs (e.g., salt-pepper), 40 unrelated pairs (e.g., art-lion), and 20 weakly related, non-tested buffer pairs used to control for primacy and recency effects. Pairs were divided evenly into two study blocks, each containing 20 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 derived from the English Lexicon Project (Balota et al., 2007; Maxwell & Huff, in press; Table 2). Further, the two study blocks were 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 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 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 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, study items that were missing JOL ratings or had ratings that were outside of the 0-100 range were removed. The screening processing removed fewer than 0.5% of items. When scoring recall responses, test items that were skipped were scored as incorrect and a liberal criterion for scoring correct items was adopted such that misspellings or pluralizations were scored as correct. All analyses were collapsed across block (analyses split by block are available in the Supplemental Materials; osf.io/cgse6/), and it should be noted that the data patterns were stable as a function of block. Partial-eta squared (η_p^2) 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, the strength of the evidence supporting the null hypothesis was further analyzed 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) \times 3 (Encoding Group: Item-Specific vs. Relational vs Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA evaluated differences between mean JOL ratings and recall rates across the four pair types and the three encoding groups. An effect of measure was found, $F(1, 85) = 18.79$, $MSE = 694.46$, $\eta_p^2 = .07$, 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, $F(2, 85) = 5.40$, $MSE = 814.98$, $\eta_p^2 = .05$, 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, $ts \geq 2.96$, $ds \geq 0.78$, except for the relational and item-specific groups, which were equivalent, $t < 1$, $p_{BIC} = .87$. Finally, an effect of pair type was found, $F(3, 255) = 766.58$, $MSE = 107.66$, $\eta_p^2 = 0.58$, 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, $ts \geq 2.69$, $ds \geq 0.17$.

A significant two-way interaction between measure and pair type confirmed that the illusion of competence replicated across encoding groups, $F(2, 85) = 5.21$, $MSE = 107.66$, $\eta_p^2 = .02$. Critically, however, a significant three-way interaction was found, $F(6, 255) = 15.56$, $MSE = 87.42$, $\eta_p^2 = .04$, 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, reliable illusion of competence patterns were detected across each of the three encoding groups, though at different rates. First, starting

with the read group, a robust illusion of competence was detected 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), $t < 1$, $SEM = 3.46$, $p = .67$, $p_{BIC} = .83$, 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), $t < 1$, $SEM = 4.52$, $p = .35$, $p_{BIC} = .78$), 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 or eliminate the illusion of competence pattern, but these reductions depended upon the 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 pairs.

Next, the correspondence between JOLs provided at study and correct recall for each of the pair types was assessed 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) \times 4 (Pair Type) \times 11 (JOL Increment) mixed ANOVA. Overall, this analysis yielded effects of Encoding Group $F(1, 85) = 3.36$, $MSE = 2394802$, $\eta_p^2 = .07$, Pair Type, $F(3, 255) = 113.64$, $MSE = 176766.84$, $\eta_p^2 = .57$, and JOL Increment, $F(10, 850) = 31.28$, $MSE = 51954.33$, $\eta_p^2 = .27$. Additionally, significant 2-way interactions emerged between Pair Type and JOL Increment, $F(30, 2550) = 6.99$, $MSE = 6409.98$, $\eta_p^2 = .12$, and Pair Type and Encoding Group, $F(6, 255) = 5.77$, $MSE = 8972.79$, $\eta_p^2 = .57$. However, the 3-way interaction was non-significant, $F(60, 2550) = 1.05$, $MSE = 958.27$, $p = .38$, $p_{BIC} = .76$. Although this interaction failed to reach statistical significance, a series of planned analyses based on the predictions made was conducted in which calibration plots were analyzed separately for each of the three encoding groups.

Starting with the read group, for unrelated pairs, JOLs were found to overestimate later recall at JOL increments of 30% or greater. 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) \times 11 (JOL increment) mixed ANOVA, these patterns were confirmed by effects of Pair Type, $F(3, 81) = 34.53$, $MSE = 60024.55$, $\eta_p^2 = .56$, JOL Increment, $F(10, 270) = 9.98$, $MSE = 12827.46$, $\eta_p^2 = .27$, and a significant interaction, $F(30, 810) = 2.32$, $MSE = 2006.78$, $\eta_p^2 = .08$.

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, $F(3, 84) = 59.58$, $MSE = 89671.24$, $\eta_p^2 = .68$, JOL Increment, $F(10, 280) = 11.34$, $MSE = 21592.66$, $\eta_p^2 = .29$, and a significant interaction, $F(30, 840) = 3.39$, $MSE = 3064.65$, $\eta_p^2 = .11$.

Finally, for the relational group, JOL overestimations of unrelated pairs were reduced relative to the read and item-specific groups, as overestimations emerged at 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 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, $F(3, 90) = 30.15$, $MSE = 43368.74$, $\eta_p^2 = .50$, JOL Increment, $F(10, 300) = 12.90$, $MSE = 22868.15$, $\eta_p^2 = .30$, and a significant interaction, $F(30, 900) = 3.83$, $MSE = 3305.73$, $\eta_p^2 = .10$.

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 in which overestimations were only found at JOLs greater than 90%. However, for unrelated and backward pairs, the illusion of competence pattern emerged at higher JOL increments in the item-specific and relational groups 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 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, an illusion of competence pattern was found such that JOLs generally exceeded that of later recall, though this pattern was moderated by pair direction and encoding group. Consistent with the predictions made, participants who engaged in the item-specific and relational encoding strategies showed a reduction in the illusion of competence through improved correct recall rates relative to the read group. Starting with backward pairs, a robust illusion of competence was found in the read group, however the illusion of competence was diminished following item-specific and relational encoding with the former task being the most effective. These results were consistent with the predictions made that item-specific encoding would be most beneficial in reducing the illusion of competence for related pairs (cf. Maxwell & Huff, in press). For forward pairs, no illusion of competence was found in any of the encoding groups—patterns that replicate previous findings (e.g., Koriat & Bjork, 2005; Maxwell & Huff). For symmetrical pairs, an illusion of competence was found in the read group, but this pattern was eliminated in the item-specific and relational groups. Finally, for the unrelated pairs, there was an illusion of competence found in both the read and item-specific groups, but the illusion of competence was eliminated for the relational group. As such, this pattern of findings was consistent with the prediction that relational

encoding would benefit recall across pair types, especially for backward and unrelated pairs.

Calibration plots were then computed to further explore the correspondence between JOLs and recall across pair types and encoding groups. Across encoding groups, participants were generally well-calibrated for forward and symmetrical pair types. For the read group, participants overpredicted unrelated pairs at all JOL increments and overpredicted backward pairs at JOL ratings greater than 50%. This pattern indicates that the read group was unable to accurately predict later recall for pairs that did not readily converge upon the target. For the item-specific group, participants similarly overpredicted unrelated pairs at almost all JOL increments, but, unlike the read group, only overpredicted backward pairs at JOL increments of 80% and greater. In the relational group however, JOLs only overpredicted later recall on increments greater than 50% and, like the item-specific group, only overpredicted recall at JOLs at increments greater than 80% on backward pairs. Collectively then, these patterns indicate that there were significant improvements in JOL calibration for both item-specific and relational groups relative to reading with the relational group showing a particular improvement on unrelated pairs given lower JOL ratings.

The improved calibration for item-specific and relational tasks was likely due to both tasks increasing correct recall (vs. adjusting JOL ratings) relative to reading, given both tasks are classified as deep processing tasks. Indeed, overall JOL rates across the three encoding groups were stable ($F(2, 85) < 1$, $MSE = 147.50$, $p = .59$, $p_{BIC} = .98$), though recall rates were greater in the item-specific ($M = 57.62$) and relational groups ($M = 58.67$), relative to the read group ($M = 45.68$; $t_s \geq 3.18$, $d_s \geq 0.57$). Because JOL rates

remained relatively stable, an important question is whether JOL calibration can be improved further if participants can successfully adjust their JOL ratings in response to deceptive word pairs (i.e., backward and unrelated pairs) that produce illusion of competence patterns. This possibility was examined in Experiment 2 by testing whether participants are able to titrate their JOLs in response to an explicit warning while also using item-specific and relational encoding tasks to boost correct recall.

CHAPTER III: Experiment 2: Item-Specific Versus Relational Encoding with Warnings

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 an explicit warning regarding the illusion of competence. There are several demonstrations indicating that participants can adjust their memory responses when presented with experimenter-provided warnings. For example, in the false memory literature, participants are often able to reduce their suggestibility when warned about possible exposure to misleading details (e.g., Chambers & Zaragoza, 2001; Eakin, Schreiber, & Sergent-Marshall, 2003; Karanian, Rabb, Wulff, Torrance, Thomas, and Race, 2020; see Blank & Launay, 2014, for a meta-analysis). Moreover, in the highly potent Deese/Roediger-McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) 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).

Unlike the false memory literature however, few studies that have examined the effects of 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 their study, all participants completed an initial study-test block in which JOLs were provided for forward, backward, and unrelated pairs. Prior to completing a 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 word pairs and report the likelihood that another participant would recall the target word when presented with the cue. 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 backward pairs because they are perceived as having a stronger association than is actually present. Thus, the theory-based group received an experience-based warning regarding the illusion of competence and the specific pair types that were most susceptible to overestimations. In contrast, the mnemonic-based group completed filler tasks and not informed of the illusion of competence. On the second study/test block, theory-based participants showed a reduction in the illusion of competence relative to the mnemonic based group, indicating that participants could adjust their JOLs in response to experimenter-provided feedback.

Given the warning benefits reported by Koriat and Bjork (2006), the purpose of Experiment 2 was to examine whether JOL accuracy could be further improved if participants were warned about the deceptive nature of word pairs prior to study relative to a no-warning group. Like Experiment 1, 2 blocks containing separate lists of cue-target pairs were studied and immediately tested. Modeling Koriat and Bjork's procedure, after block 1, participants in the warning group were explicitly informed about the illusion of competence with deceptive pairs. To enhance warning effectiveness, a data figure taken from Maxwell and Huff, (in press; see Figure 3) which depicted the illusion of competence pattern was also shown to participants. Immediately following the warning, participants then studied the second block of word pairs followed by a cued-recall test. It

was reasoned that warnings would be most effective if 1) participants were initially exposed to the different pair types and thus completed a 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 empirical data depicting the illusion of competence pattern.

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

Methods

Participants

A total of 216 participants were recruited for Experiment 2. Of these participants, 129 (17 in lab; 112 online¹) 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,

¹Due to COVID-19 restrictions, Experiment 2 participants were sampled from in-lab and online sources. Participant source did not interact with any of the results, $F_s < 1$, though most 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.

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

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 also shown a data figure (from Maxwell & 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, which similarly removed fewer than 0.5% of trials. In the following analyses, because the warning manipulation was only applied to the second block, JOL and recall analyses only used participant data on the second block in both the warning and no-warning groups. For completeness, analyses for both blocks are included in the Supplemental Materials

(available at: <https://osf.io/cgse6/>), and the data patterns largely follow those found in block 2.

In the following analyses, the effects of warning on JOLs and recall rates were examined first. No main effect of warning was found, $F < 1$, $p_{\text{BIC}} = .92$, and warning did not interact with any other factor, largest $F = 2.03$, $p = .16$, $p_{\text{BIC}} = .83$. Means across warning and no warning groups in Experiment 2 (see Figure 4) are reported, but for concision, the analyses below collapse across warning groups.

Mean JOL ratings and recall rates as a function of encoding task and pair type are reported in Figure 5. A 3 (Encoding Group \times 2 (Measure) \times 4 (Pair Type) mixed ANOVA yielded an effect of measure, $F(1, 198) = 58.71$, $MSE = 654.06$, $\eta_p^2 = .23$, in which JOL ratings were greater than recall rates (55.82 vs. 46.14). An effect of encoding group was also found, $F(2, 198) = 3.60$, $MSE = 1361.38$, $\eta_p^2 = .04$, 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 between the read and item-specific groups (47.94 vs. 51.39), $t(139) = 1.61$, $SEM = 2.14$, $p = .11$, $p_{\text{BIC}} = .76$. There was no difference between the relational and item-specific groups (53.88 vs. 51.39), $t(138) = 1.16$, $SEM = 2.14$, $p = .25$, $p_{\text{BIC}} = .85$. An effect of pair type was also found, $F(3, 594) = 1253.93$, $MSE = 168.01$, $\eta_p^2 = .86$, 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, $ts > 3.60$, $ds > 0.18$.

A measure \times pair type interaction was also found, $F(3, 639) = 134.27$, $MSE = 112.44$, $\eta_p^2 = .39$, 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 encoding group \times pair type interaction, $F(6, 639) = 298.36$, $MSE = 186.55$, $\eta_p^2 = .09$. Importantly, and consistent with Experiment 1, the three-way interaction was also reliable, $F(6, 639) = 298.36$, $MSE = 112.44$, $\eta_p^2 = .02$. 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 as 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 encoding eliminated the illusion of competence, but only for unrelated pairs.

A series of calibration plots was again constructed to assess the correspondence between the JOLs provided at study and correct recall for each of the four pair types (Figure 6). Consistent with Experiment 1, calibration plots were initially analyzed using a 3 (Encoding Group) \times 4 (Pair Type) \times 11 (JOL increment) mixed ANOVA. As in Experiment 1, the 3-way interaction failed to reach statistical significance, $F(60, 6150) = 0.97$, $MSE = 823.39$, $p = .55$, $p_{BIC} = .85$. Given the encoding group predictions made, however, calibration plots were analyzed separately for each of the encoding groups.

Starting with the read group, overestimations of unrelated pairs were observed for JOL rates above 20%. Next, overestimation of backward pairs occurred at JOLs greater than 40%. For symmetrical associates, overestimations occurred for JOLs greater than

70%. Finally, overestimation of forward associates occurred at JOL ratings above 80%. Using a 4 (Pair Type) \times 11 (JOL increment) repeated measures ANOVA, yielded effects of Pair Type, $F(3, 207) = 163.37$, $MSE = 219312.96$, $\eta_p^2 = .70$, JOL Increment, $F(10, 690) = 22.76$, $MSE = 26835.39$, $\eta_p^2 = .25$, and an interaction, $F(30, 2070) = 5.79$, $MSE = 4698.59$, $\eta_p^2 = .08$.

Next, for the item-specific encoding group, overestimations of unrelated pairs emerged at JOL ratings above 30%. For backward pairs, overestimations occurred at JOL ratings greater than 50%. 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%. Effects of Pair Type, $F(3, 213) = 156.17$, $MSE = 215634.69$, $\eta_p^2 = .69$, JOL Increment, $F(10, 710) = 30.77$, $MSE = 37590.21$, $\eta_p^2 = .30$, and an interaction, $F(30, 2130) = 8.21$, $MSE = 7013.33$, $\eta_p^2 = .10$, again confirmed these patterns.

Finally, the calibration between JOLs and recall for participants who completed the relational encoding task were assessed. JOL overestimations of unrelated pairs emerged for JOL ratings above 40%. Next, overestimations of backward pairs emerged when JOLs ratings were greater than 40%, while overestimations of symmetrical associates again occurred at JOLs greater than 80%. Finally, for forward associates, overestimations only occurred at JOLs greater than 90%. All patterns of overestimation were again confirmed by effects of Pair Type, $F(3, 195) = 86.72$, $MSE = 142608.89$, $\eta_p^2 = .57$, JOL Increment, $F(10, 650) = 22.36$, $MSE = 303656.40$, $\eta_p^2 = .26$, and an interaction, $F(30, 1950) = 7.08$, $MSE = 885.80$, $\eta_p^2 = .10$.

Similar 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. However, unlike Experiment 1, backward pairs also showed similar patterns across encoding groups. Thus, the item-specific encoding group did not benefit backward associates to the same degree as in Experiment 1. For unrelated pairs, the illusion of competence pattern emerged at higher JOL increments in the item-specific and relational groups.

Discussion

In Experiment 2, warning instructions were used as a means of further enhancing JOL calibration with later recall. It was expected that providing a warning would encourage participants to titrate their JOL ratings in response to the different pair types. The warning manipulation was modeled after Koriat and Bjork (2006) by providing participants with an initial block of cue-target study trials prior to providing them with a warning about the illusion of competence and emphasizing the deceptive nature of backward and unrelated pairs with a graphical depiction of JOLs and recall data. Despite these efforts, however, warnings were ineffective at reducing the illusion of competence when participants completed item-specific, relational, and read tasks.

Although warnings were ineffective at improving JOL calibration, Experiment 2 again showed that item-specific and relational encoding tasks can improve JOL calibration. Specifically, item-specific and relational encoding eliminated the illusion of competence patterns for unrelated pairs and greatly improved calibration in the

calibration plots relative to the read group. These calibration benefits were not found on backward pairs—a pattern inconsistent with Experiment 1. This discrepancy is further discussed in the General Discussion but note that item-specific and relational tasks did provide some improvement in JOL calibration on unrelated pairs and the improved calibration for relational encoding was consistent with Experiment 1.

Calibration plots again provided a more precise assessment of the specific JOL increments in which illusions of competence emerged. Overall, the illusion of competence replicated for all backward and unrelated pair types. Furthermore, consistent with findings in Experiment 1, relational encoding improved the correspondence between JOLs and recall for unrelated pairs. Thus, even though the item-specific/relational framework was effective at increasing calibration, the illusion of competence pattern persisted.

CHAPTER IV: General Discussion

Overall, the present study sought to improve the predictive efficacy of JOL ratings on subsequent recall of forward, symmetrical, backward, and unrelated cue-target word pairs. Previous research has consistently shown that JOLs tend to be over predictive on unrelated and deceptive backward pairs resulting in an illusion of competence pattern (Koriat & Bjork, 2005; Maxwell & Huff, in press). This pattern was attempted to be attenuated through the use of deep item-specific and relational encoding tasks relative to a read-control group. In Experiment 1, forward pairs did not show an illusion of competence pattern and symmetrical pairs showed a small illusion of competence that was eliminated in the item-specific and relational groups relative to reading. As expected, the illusion of competence was highest for backward and unrelated pairs and item-specific and relational tasks were found to reduce, but not eliminate, the illusion. Specifically, for backward pairs, both item-specific and relational tasks were found to reduce the illusion of competence, though the item-specific task produced the greater reduction. In contrast however, the relational group produced a greater reduction for unrelated pairs than the item-specific group. Collectively then, both item-specific and relational encoding tasks can improve JOL accuracy over a standard read task, though their relative effectiveness depends upon the associative direction of the pair type.

Encoding groups were also compared using a series of calibration plots which plotted study pairs at different JOL intervals against their subsequent recall rates. Calibration plots provide a more fine-grained assessment of the correspondence between JOLs and recall rates by revealing specific intervals in which JOLs do and do not align with subsequent recall. Although the omnibus analysis did not find an interaction

between the illusion of competence and encoding task, some differences were found in the calibration plots when encoding groups were analyzed separately. Specifically, the calibration plots revealed that across encoding groups, participants were well-calibrated for forward and symmetrical pairs but less calibrated for backward and unrelated pairs due to the presence of an illusion of competence. When examining the read group, participants showed overpredictions at all JOL increments for unrelated pairs and overpredictions at all JOL increments greater than 50% for backward pairs. In the item-specific group, overpredictions were found almost all JOL increments for the unrelated pairs, but only for JOL increments above 80% for backward pairs. Finally, the relational group showed overpredictions at all JOL increments over 50% for unrelated word pairs and above 60% for backward pairs. The JOL accuracy benefits following item-specific and relational encoding appear to be due to improvements in overall recall rates of word pairs rather than adjustments in JOL ratings in response to different pair types.

In Experiment 2, the JOL accuracy benefits following item-specific and relational encoding were further examined by employing an explicit warning about the misleading nature of some of the word pairs. Specifically, participants were instructed that backward pairs were misleading because the cue word, when presented in isolation at test, was not predictive of the studied target. Participants completed an initial study/test block containing all pair types so that they would have an opportunity to experience encoding and retrieving the different pair types and were then provided with information regarding the illusion of competence. Additionally, participants were provided with a data figure depicting the illusion of competence typically found for backward and unrelated pairs and were told that they would study a second block of cue/target pairs and to try to avoid

producing the illusion (cf. Koriat & Bjork, 2006). Despite this explicit warning however, the illusion of competence pattern was unchanged relative to the no warning group. Consistent with the findings in Experiment 1, the illusion of competence was absent or small for forward and symmetrical pairs, but robust for backward and unrelated pairs. Both item-specific and relational encoding groups were found to improve JOL accuracy for unrelated pairs, but unlike Experiment 1, these benefits did not extend to backward pairs as the illusion of competence was similar in magnitude to the read group. Calibration plots largely echoed these patterns where item-specific and relational tasks showed overpredictions at higher JOL increments on unrelated pairs relative to the read group.

While the encoding manipulations remained at least partially effective in Experiment 2, the surprise finding was that warnings were ineffective at reducing the illusion of competence despite great efforts to educate participants about deceptive word pairs prior to study. The warning instructions were modeled after Koriat and Bjork's (2006) warning procedure which found that warnings improved JOL accuracies, and note that there are several examples of warnings effectively reducing associative false memory illusions (Huff, Meade, & Hutchison, 2011; McCabe & Smith, 2002), and susceptibility to misinformation (Blank & Launay, 2014). Despite these memory and metamemory warning benefits, two possibilities are suggested as to why the warnings failed to improve JOL calibration. First, although the warning provided discussed different pair types including the deceptive nature of backward pairs, the warning still may not have been specific enough to produce a reduction in the illusion of competence. For example, previous research has shown that broad/general warnings about misinformation are less

effective than specific warnings (Ecker, Lewandowsky, and Tang, 2010; Huff & Umanath, 2018) and that explicit warnings may be sensitive to different types of misleading items (Umanath, Ries, & Huff, 2019). Within the context of the present study, the warning could be made more specific by only warning participants about the illusion of competence for one type of deceptive word pair (e.g. backward pairs) and including several examples of backward pairs to facilitate identification at study. Second, though participants were provided with a graph depicting the general patterns of the illusion of competence, this graph only provided a general data pattern from another study and did not display a participant's individual performance on the task. Participants may have been more responsive if they were provided with their own JOL/recall data when providing the warning, which may have improved the effectiveness of the warning. Given phenomena such as the better-than-average effect (Cross, 1977; Zell, Stickhouser, Sedikides, & Alicke, 2020), it is also reasonable to expect that participants may be more dismissive of general behavioral patterns that are unfavorable relative to information regarding individual patterns.

Although Experiments 1 and 2 similarly implemented item-specific and relational encoding tasks, it should be noted that the encoding effects on the illusion of competence were not always consistent. Specifically, for backward pairs, Experiment 1 showed the greatest reduction for the illusion of competence in the item-specific group, but Experiment 2 did not show this reduction and the illusion of competence was still observed for backward pairs. Furthermore, backward pairs also saw a reduction for the illusion of competence in the relational group in Experiment 1, but this reduction did not replicate in Experiment 2. Finally, another difference in Experiment 2 was that both the

item-specific and relational groups showed an elimination for the illusion of competence for unrelated pairs, whereas Experiment 1 only showed an elimination in the relational group. It is suggested that these discrepancies are may be attributed in part to the effects of the warning on encoding. Although warnings were found to be ineffective, it is possible that participants may have been cognizant of the deceptive nature of some of the word pairs and may have been trying to monitor for these pairs, which negatively affected their use of item-specific and relational encoding processes. Consistent with this possibility, a cross-experimental comparison of recall rates in item-specific and relational encoding groups in Experiment 1 and item-specific and relational encoding groups in Experiment 2 indicated that recall rates were lower in Experiment 2 where warnings were provided relative to Experiment 1 (41.05 vs. 54.19), $t(210) = 5.92$, $SEM = 2.25$, $d = 0.27$, indicating that the encoding tasks may not have been completed as effectively. Second, in an attempt to improve the success of warnings by giving participants an opportunity to experience the different pair types, the Experiment 2 warning manipulation was only conducted on second block pairs rather than both blocks, which could have reduced the effectiveness of item-specific and relational encoding tasks. Consistent with this possibility, a cross-block comparison found that overall recall rates in the item-specific and relational groups in block 1 exceeded that of recall rates found in the item-specific and relational groups in block 2 (43.28 vs. 41.05), $t(207) = 2.49$, $SEM = 0.73$, $d = 0.04$, further suggesting that differences between experiments were in part due to a block effect. Thus, it is possible warnings may have reduced the effectiveness of item-specific and relational encoding at improving JOL calibration, rather than improving it—a

interesting pattern that suggests there may be limits to JOL accuracy benefits when deep encoding tasks are combined with warning instructions.

Finally, the analyses included calibration plots to provide a more precise assessment of specific JOL increments in which participant JOLs become miscalibrated with recall. In both experiments, these plots revealed participants were generally well-calibrated for forward and symmetrical pairs across encoding groups and only showed overpredictions at the highest JOL increments. Furthermore, in Experiment 1, item-specific and relational processing each improved the correspondence between JOLs and recall relative to the read group for backward and unrelated pairs, particularly at lower JOL increments. In Experiment 2, item-specific and relational encoding were also beneficial towards decreasing overpredictions of unrelated pairs and backward associates at lower JOL increments, though at a lesser magnitude relative to Experiment 1. As such, these plots showed important qualitative differences between the item-specific and relational encoding groups relative to silent reading which allowed for increased precision on the specific JOL ratings that show reductions in the illusion of competence.

CHAPTER V: Conclusion

The present study showed that the illusion of competence can be reduced using the item-specific/relational framework. In Experiment 1, it is shown that the 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. The relational encoding task again proved beneficial in reducing the illusion of competence found for unrelated word pairs in Experiment 2. While the study found that warnings were ineffective 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. The calibration plots used in this study provided qualitative information about the specific JOL increments at which the item-specific and relational tasks begin to reduce the illusion of competence. These plots serve as a way to visually depict where participants were becoming overconfident with their JOL ratings in comparison to their recall rates. The general patterns found by the calibration plots were that item-specific and relational encoding tasks improved calibration at higher JOL increments, particularly for backward and unrelated pairs. 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.

CHAPTER VI: Tables and Figures

Table 1: Mean Associative Strength Summary Statistics Forward, Backward, and Symmetrical Pairs

Mean Associative Strength Summary Statistics Forward, Backward, and Symmetrical Pairs.

Condition	Variable	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
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

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

Summary Statistics for Cue and Target Concreteness, Length, and Frequency Item Properties as a Function of Pair Type.

Pair Type	Position	Variable	<i>M</i>	<i>SD</i>
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	1.28
		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).

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.

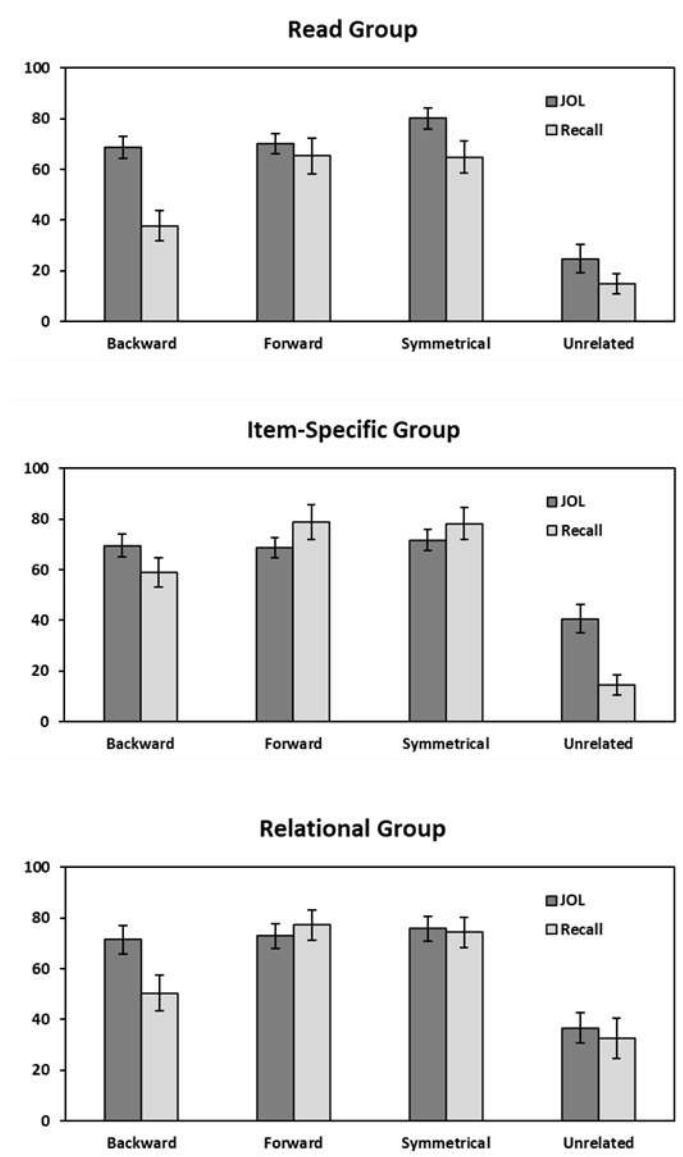


Figure 2. Calibration plots as a function of pair type in Experiment 1 for participants in the Read group (top panel), Item-Specific group (middle panel), and Relational group (bottom panel). Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued-recall. Overconfidence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

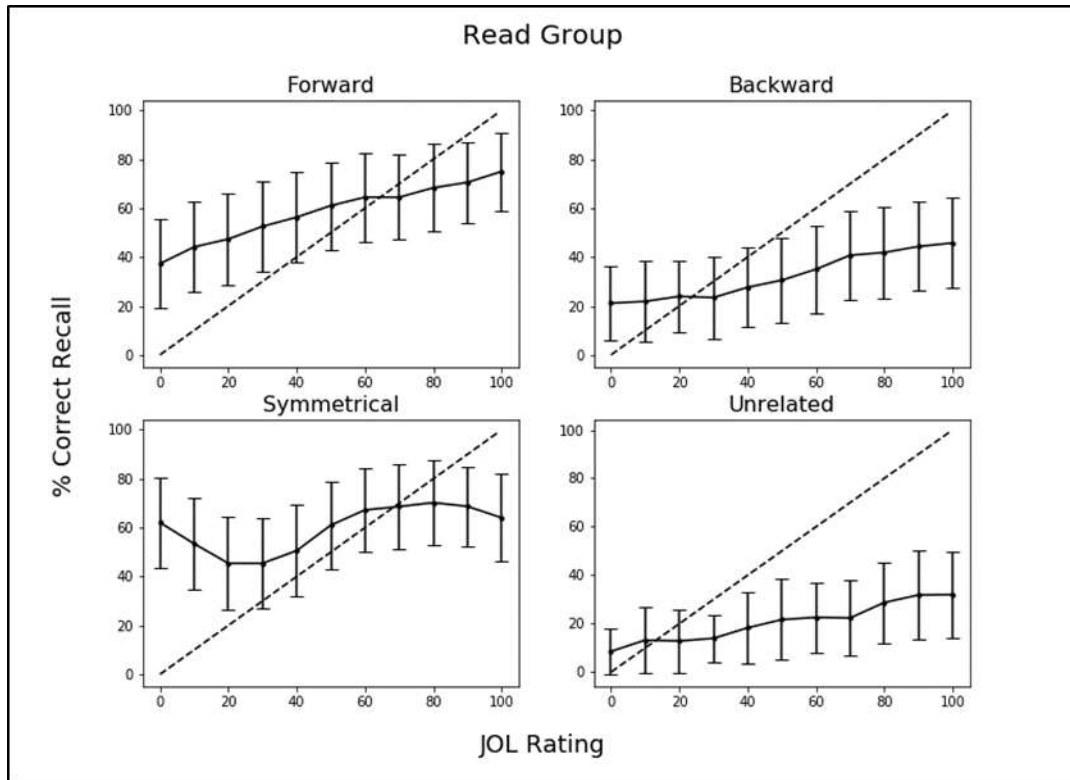


Figure 2 (Continued)

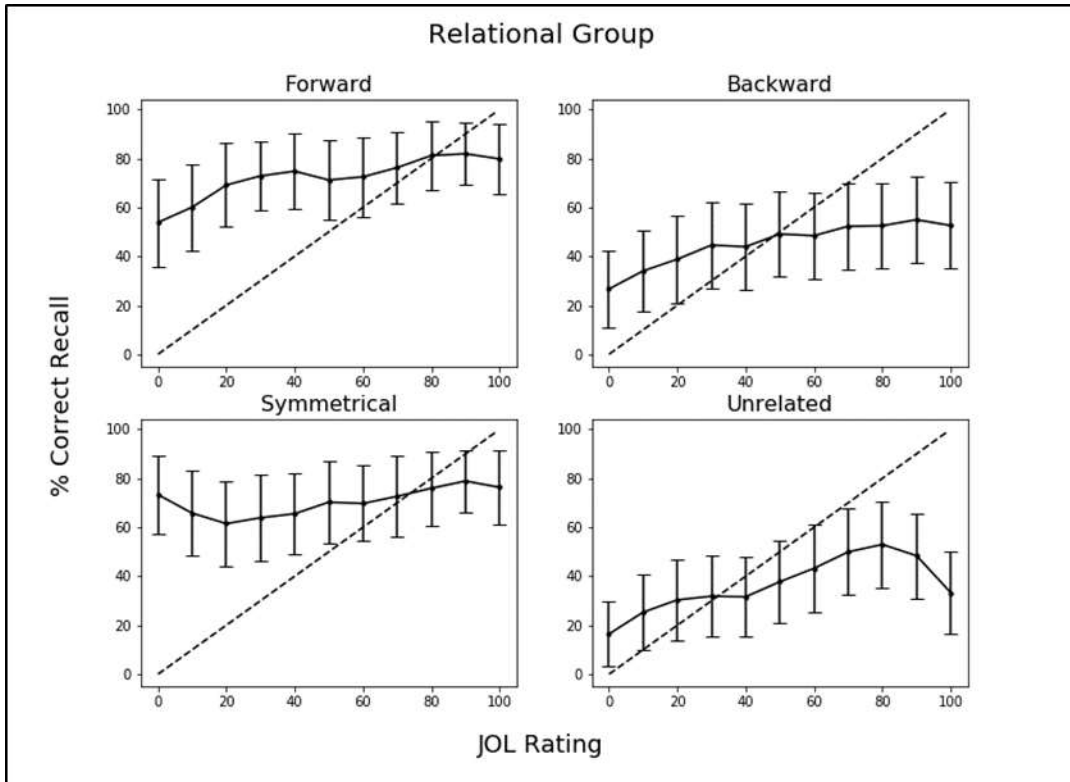
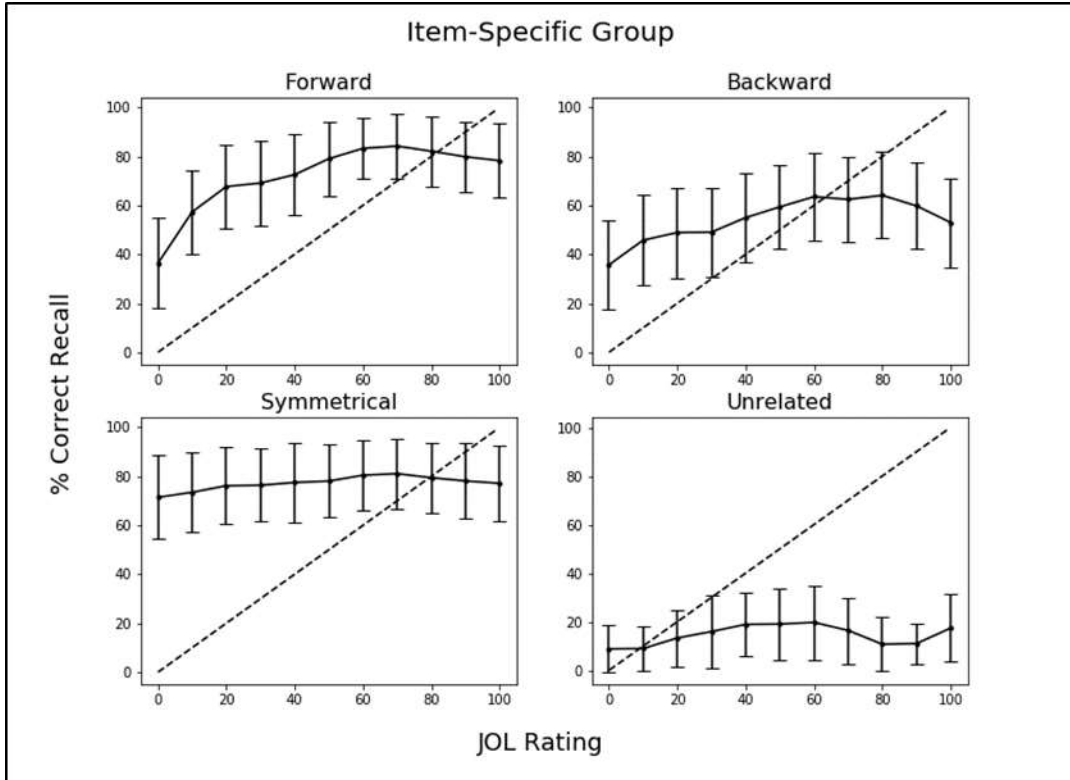


Figure 3. Sample data illustrating the illusion of competence for backward, symmetrical, and unrelated study pairs. This graph was provided to participants in the Experiment 2 warning group. Data pattern is modeled after Maxwell and Huff (in press).

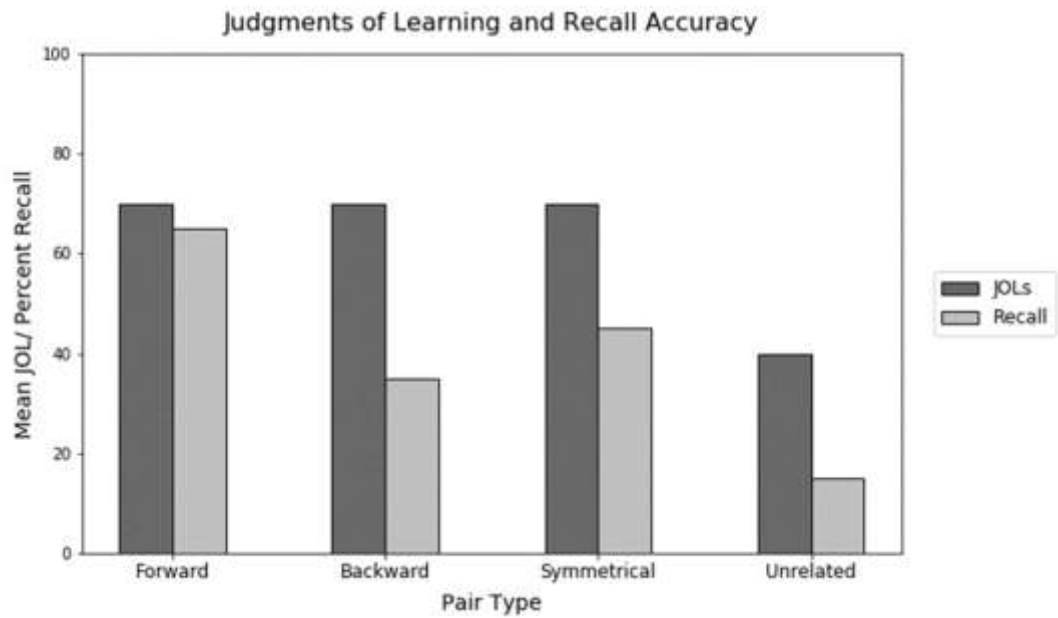


Figure 4. Mean JOL and recall rates as a function of pair type in the Read (top panels), Item-Specific (middle panels), and Relational (bottom panels) Warning and No Warning groups in Experiment 2. Bars represent 95% confidence intervals.

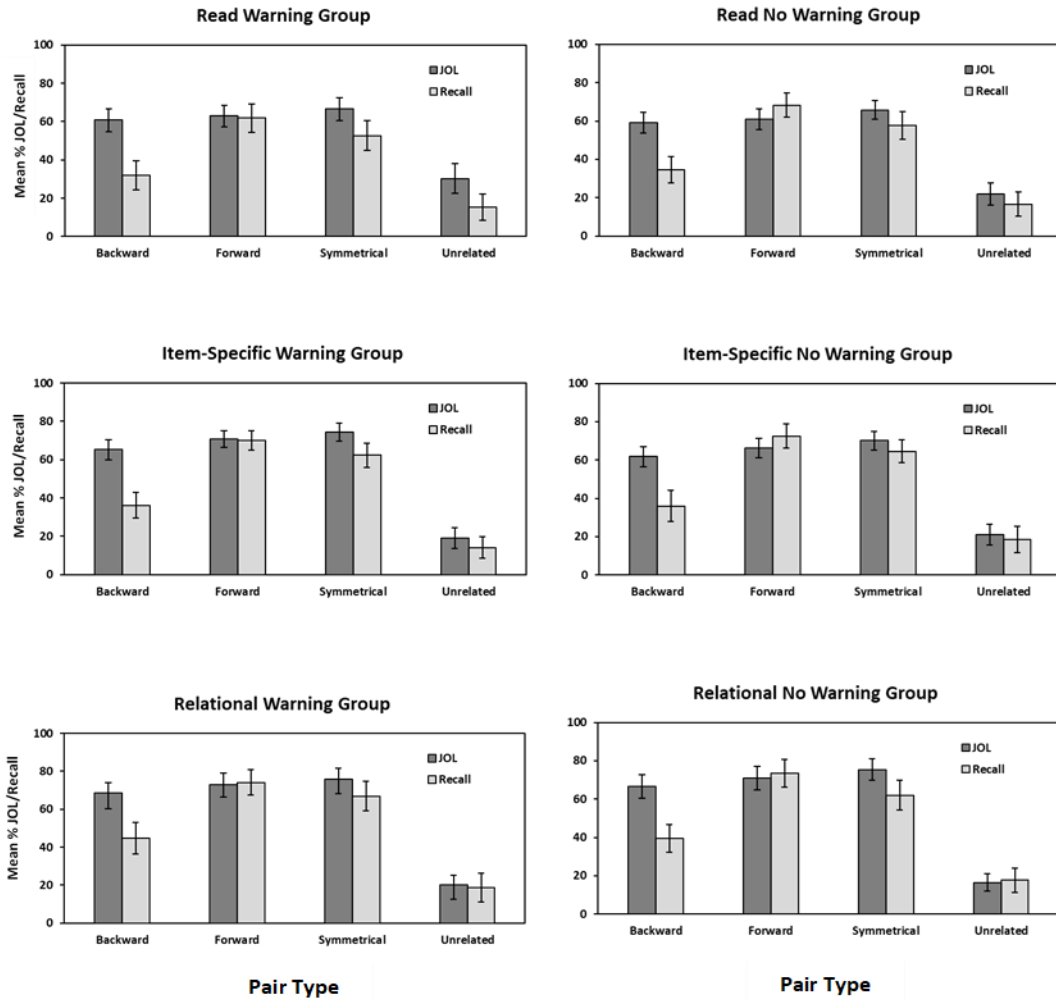


Figure 5. Mean JOL and recall rates as a function of pair type collapsed across warning for the read, item-specific, and relational groups in Experiment 2.

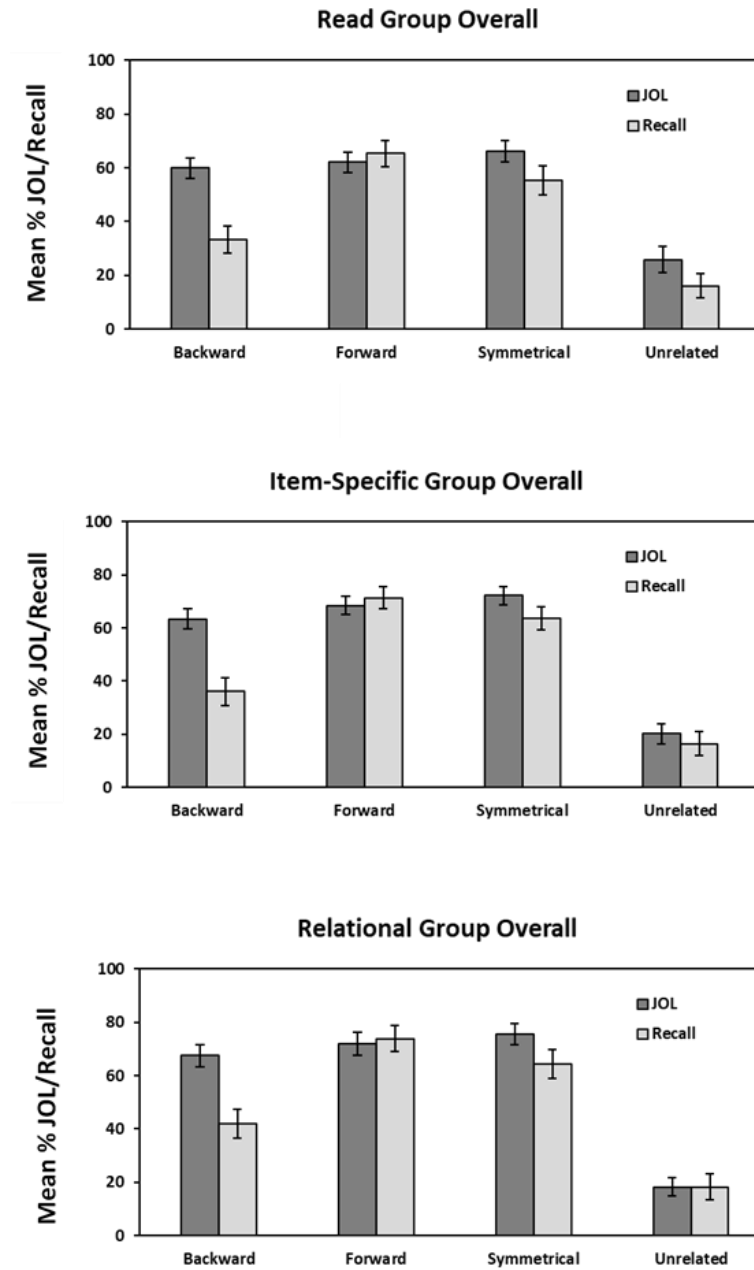


Figure 6. Calibration plots as a function of pair type in Experiment 2 for participants in the Read group (top panel), Item-Specific group (middle panel), and Relational group (bottom panel) collapsed across warning. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued-recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence intervals.

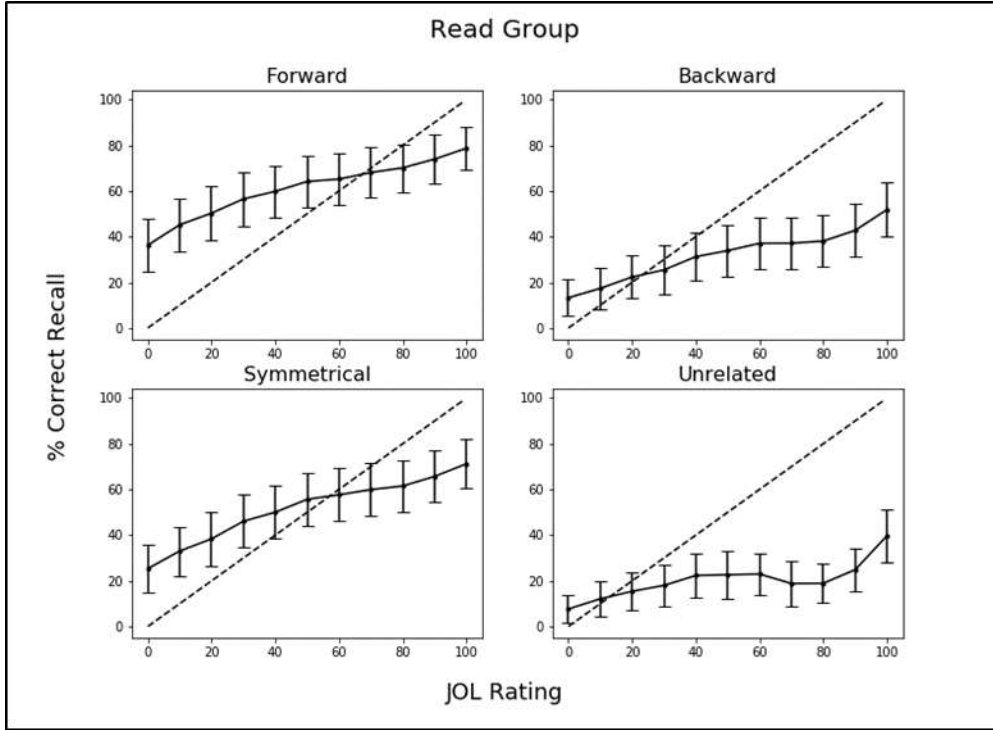
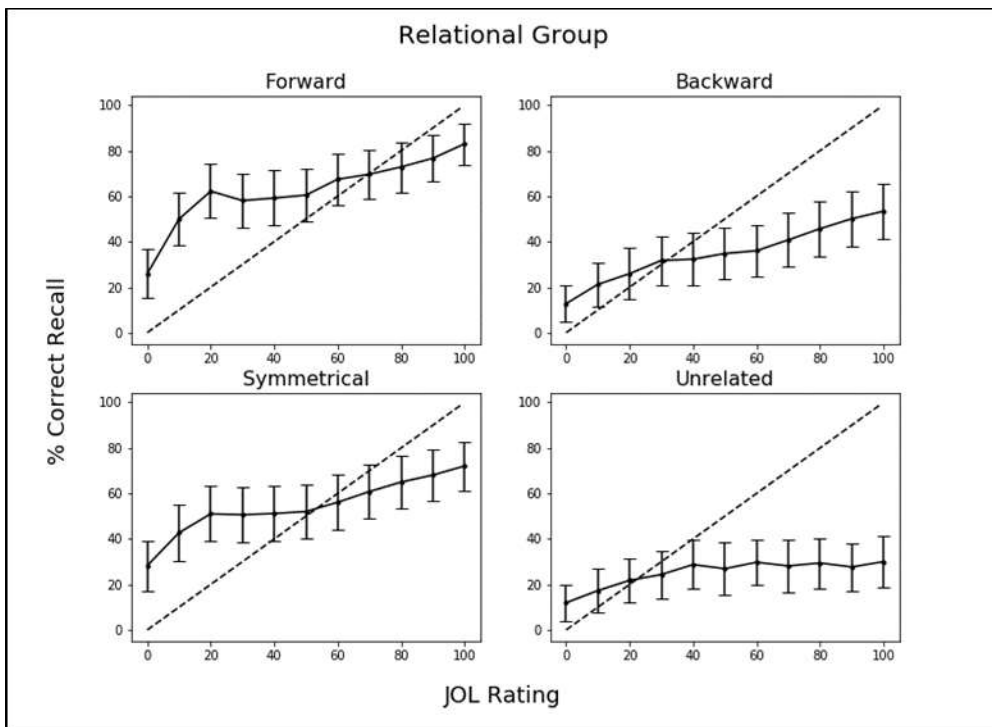
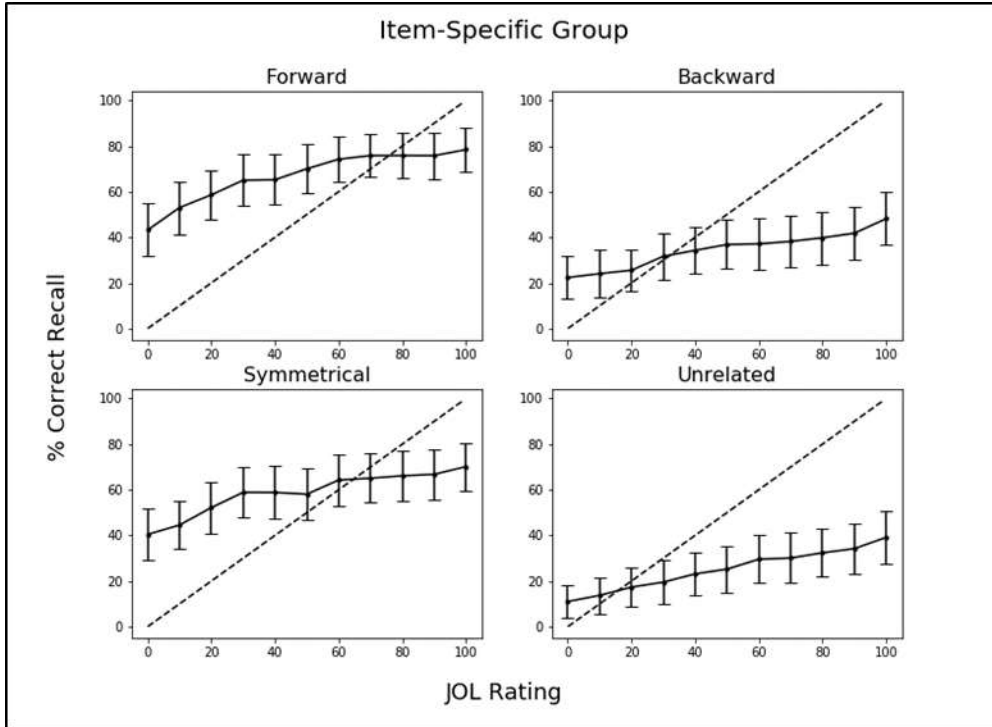


Figure 6 (continued)



APPENDIX A: SUPPLEMENTAL ANALYSES

Experiment 1 Block Effects

An assessment was made to determine whether mean JOL/recall responses in Experiment 1 differed as a function Block. Using a 4 (Pair Type: Forward vs Backward vs Symmetrical vs Unrelated) \times 3 (Encoding Group: Item-Specific vs. Relational vs Read) \times 2 (Block: Block 1 vs Block 2) \times 2 (Measure: JOL vs Recall) mixed ANOVA yielded an effect of Pair Type on JOLs/recall, $F(3, 255) = 766.76$, $MSE = 215.27$, $\eta_p^2 = .52$. Overall, a significant main effect of block was detected, $F(1, 85) = 10.82$, $MSE = 166.35$, $\eta_p^2 = .01$, in which collapsed across Pair Type, Measure, and Encoding Group, mean JOLs/recall rates were highest in block 1 (59.56) relative to block 2 (57.29).

These effects were qualified by two significant three-way interactions. First, an interaction occurred between Block, Encoding Group, and Measure, $F(2, 85) = 12.70$, $MSE = 158.03$, $\eta_p^2 = .01$, in which, collapsed across pair types, the magnitude of the illusion of competence differed across Encoding Groups as a function of Block. Starting with participants in the silent reading group, mean JOLs exceeded recall in both block 1 (63.40 vs 44.88) and block 2 (58.56 vs 46.48; $t_s \geq 4.19$, $d_s \geq 0.99$). Next, for participants in the item-specific group, mean JOLs were equivalent to recall for participants in block 1 (62.23 vs 61.17; $t(28) < 1$, $SEM = 4.68$, $p = .81$, $p_{BIC} = .84$), however they significantly differed in block 2 (62.97 vs 54.07; $t(28) = 2.34$, $SEM = 3.98$, $d = 0.65$). Finally, for participants in the relational group, JOLs and recall were equivalent in the first block (63.15 vs 65.28; $t(30) < 1$, $SEM = 3.81$, $p = .69$, $p_{BIC} = .83$), while JOLs exceeded recall in the second block (61.70 vs 55.63, $t(30) = 2.82$, $SEM = 3.58$, $d = 0.60$).

Next, an interaction was detected between Block, Pair Type, and Measure, $F(3, 255) = 6.13$, $MSE = 120.71$, $\eta_p^2 = .01$, in which, collapsed across encoding group, the illusion of

competence differed as function of block for each pair type. Starting with forward pairs, mean JOLs and recall were equivalent in block 1 (70.69 vs 71.69; $t(87) < 1$, $SEM = 3.11$, $p = .74$, $p_{BIC} = .90$), however, JOLs exceeded mean recall in block 2 (70.55 vs 76.21, $t(87) = 2.61$, $SEM = 2.20$, $d = 0.35$). Next, for backward pairs, JOLs exceed later recall for both block 1 (71.69 vs 55.46) and block 2 (68.24 vs 42.99). All comparisons for backward pairs were significant ($ts \geq 4.90$, $ds \geq 0.75$) and the illusion of competence replicated across both blocks. Third, for symmetrical pairs, JOLs in block 1 did not differ from recall (75.94 vs 75.40, $t(87) < 1$, $SEM = 2.86$, $p = .84$, $p_{BIC} = .90$). However, in block 2, JOLs exceed later recall (75.70 vs 69.86, $t(87) = 2.19$, $SEM = 2.70$, $d = 0.34$) indicating a small illusion of competence. Finally, for unrelated pairs, JOLs exceeded later recall in both block 1 (33.37 vs 22.11) and block 2 (35.04 vs 19.67.) Both comparisons differed statistically ($ts \geq 3.99$, $ds \geq 0.57$); thus, the illusion of competence replicated across blocks for unrelated pairs.

No other three-way interactions with Block were detected, and the four-way interaction was non-significant, $F(6, 255) = 1.76$, $MSE = 120.71$, $p_{BIC} = .99$. For completeness, mean JOLs and recall rates split by block are reported in Table 1.

Experiment 2 Warning Effects

Next, using a 4 (Pair Type: Forward vs Backward vs Symmetrical vs Unrelated) \times 3 (Encoding Group: Item-Specific vs. Relational vs Read) \times 2 (Warning: Warning vs No-Warning) \times 2 (Measure: JOL vs Recall) it was tested whether mean JOL/recall rates differed as a function of the warning manipulation in Experiment 2. Consistent with the full analyses, only data for block 2 was included in this analysis. Overall, the effect of warning was non-significant, $F(1, 211) < 1$, $MSE = 1721.42$, $p_{BIC} = .92$, indicating that informing participants about the deceptive

nature of a backward associates, symmetrical associates, and unrelated pairs did not influence their JOLs or recall rates.

Experiment 2 Block Effects

Finally, an assessment was made to determine whether mean JOL/recall rates differed as a function of Block in Experiment 2 via a 4 (Pair Type: Forward vs Backward vs Symmetrical vs Unrelated) \times 3 (Encoding Group: Item-Specific vs. Relational vs Read) \times 2 (Block: Block 1 vs Block 2) \times 2 (Measure: JOL vs Recall). This analysis yielded a significant effect of Block, $F(1, 214) = 20.77$, $MSE = 250.17$, $\eta_p^2 = .01$, such that collapsed across Encoding Group, Pair Type, and Measure, mean JOLs/recall rates were higher in block 1 (49.83) relative to block 2 (47.43). These effects were then qualified by a significant three-way interaction between Encoding Group, Measure, and Block, $F(6, 642) = 3.89$, $MSE = 69.76$, $\eta_p^2 = .001$. As such, when collapsed across pair types, the magnitude of the illusion of competence differed across Encoding Groups as a function of Block.

Starting with participants in the silent reading group, mean JOLs exceeded recall in both block 1 (54.28 vs 37.26) and block 2 (52.14 vs 36.08; $ts \geq 6.78$, $ds \geq 0.86$). Next, for participants in the item-specific group, mean JOLs exceed later recall in both block 1 (58.41 vs 43.73) and block 2 (55.12 vs 40.84; $ts \geq 5.27$, $ds \geq 0.85$). Finally, this pattern occurred again for participants in the relational group, as JOLs and recall differed both block 1 (60.68 vs 45.60) and block 2 (57.33 vs 43.84; $ts \geq 4.89$, $ds \geq 0.71$).

All other three-way interactions with Block were non-significant, and the four-way interaction was non-significant, $F(6, 642) = .17$, $MSE = 54.18$, $p_{BIC} = .99$. Table 2 reports mean JOL and Recall rates for each pair type split by both block and warning group for each of the three encoding manipulations.

Table 1

Mean JOLs and Recall in Experiment 1 Split by Block

Block	Measure	Encoding Task	Forward	Backward	Symmetrical	Unrelate
One	JOL	Item-Specific	67.47	69.20	70.51	41.87
		Relational	72.57	72.23	76.04	31.74
		Read	71.96	73.30	81.44	26.33
	Recall	Item-Specific	76.68	67.36	84.28	16.26
		Relational	77.74	57.58	76.82	34.63
		Read	59.67	40.79	76.82	34.63
Two	JOL	Item-Specific	69.87	69.91	72.34	39.42
		Relational	73.36	70.86	75.49	41.45
		Read	68.15	63.62	78.99	23.37
	Recall	Item-Specific	81.00	50.61	72.19	12.44
		Relational	76.70	43.34	72.01	30.42
		Read	70.71	34.77	65.05	15.26

Table 2

Mean JOLs and Recall for the Warning and No Warning Groups in Experiment 2 Split by Block

Warning	Block	Measure	Encoding Task	Forward	Backward
No Warning	One	JOL	Item-Specific	71.55	68.97
			Relational	72.86	70.42
			Read	64.55	60.94
		Recall	Item-Specific	64.95	31.90
			Relational	59.91	32.99
			Read	58.42	33.31
	Two	JOL	Item-Specific	65.31	61.03
			Relational	69.17	64.90
			Read	59.45	57.89
		Recall	Item-Specific	62.72	28.88
			Relational	60.24	32.42
			Read	59.44	31.00
Warning	One	JOL	Item-Specific	72.25	68.33
			Relational	75.61	70.02
			Read	64.63	61.47
		Recall	Item-Specific	69.21	36.98
			Relational	75.16	45.81
			Read	53.82	26.87
	Two	JOL	Item-Specific	68.56	63.78
			Relational	73.01	67.02
			Read	61.33	58.11
		Recall	Item-Specific	63.51	31.70
			Relational	72.10	41.94
			Read	49.68	25.46

APPENDIX B: IRB APPROVAL LETTER

Office of
Research Integrity



118 COLLEGE DRIVE #5125 • HATTIESBURG, MS | 601.266.6576 | USM.EDU/ORI

Modification Institutional Review Board Approval

The University of Southern Mississippi's Office of Research Integrity has received the notice of your modification for your submission Exploring the effects of associative direction on judgments of learning and cued recall (IRB #: IRB-18-15).

Your modification has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services regulations (45 CFR Part 46), and University Policy to ensure:

- The risks to subjects are minimized and reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered involving risks to subjects must be reported immediately. Problems should be reported to ORI via the Incident template on Cayuse IRB.
- The period of approval is twelve months. An application for renewal must be submitted for projects exceeding twelve months.

PROTOCOL NUMBER: IRB-18-15

PROJECT TITLE: Exploring the effects of associative direction on judgments of learning and cued recall

SCHOOL/PROGRAM: School of Psychology, Psychology

RESEARCHER(S): Nicholas Maxwell ,Mark Huff

IRB COMMITTEE ACTION: Approved

7. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

PERIOD OF APPROVAL: March 30, 2020

A handwritten signature in cursive script that reads "Donald Sacco".

Donald Sacco, Ph.D.
Institutional Review Board Chairperson

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: 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: 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: 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: 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: 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: 10.1016/S0022-5371(72)80001-X
- Cross, K. P. (1977). Not can, but *will* college teaching be improved? *New Directions for Higher Education*, 17, 1-15. doi:10.1002/he.36919771703
- 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
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17-22. doi: 10.1037/h0046671
- 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: 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: 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>.
- 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: 10.1016/j.jml.2013.05.003
- 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: 10.1016/j.jml.2014.02.004

- Huff, M. J., Meade, M. L., & Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory*, 19, 317-330. doi: 10.1080/09658211.2011.568494
- Huff, M. J., & Umanath, S. (2018). Evaluating suggestibility to additive and contradictory misinformation following explicit error detection in younger and older adults. *Journal of Experimental Psychology: Applied*, 24, 180-195. doi: 10.1037/xap0000138
- 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: 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
- 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: 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. L., & 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: 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>
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803-814. doi: 10.1037/0278-7393.21.4.803
- 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

Umanath, S., Ries, F., & Huff, M. J. (2019). Comparing suggestibility to additive versus contradictory misinformation in younger and older adults following divided attention and/or explicit error detection. *Applied Cognitive Psychology, 33*, 793-805. doi: 10.1002/acp.3522

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

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin, 146*(2), 118–149. doi: 10.1037/bul0000218