



Item-specific and relational encoding are effective at reducing the illusion of competence

Nicholas P. Maxwell¹ · Emily E. Cates¹ · Mark J. Huff¹

Received: 6 February 2023 / Accepted: 2 October 2023 / Published online: 6 November 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Metamemory, or the ability to understand the capacities of one's own memory, is important for learning. To investigate questions surrounding metamemory, researchers commonly have participants make judgments of learning (JOLs) at encoding, in which participants rate their likelihood of recalling the target in a cue–target word pair when shown only the cue at test. However, the associative direction of cue–target pairs can affect the calibration of JOLs. Unlike forward associates (e.g., credit–card), in which JOLs often 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 cue–target pairs (e.g., artery–bronze) such that JOLs overestimate later recall. The present study evaluates whether the illusion of competence can be reduced when participants apply deep item-specific or relational encoding tasks relative to silent reading. Across two experiments, we show that both item-specific and relational encoding strategies reduce the illusion of competence for backward associates and unrelated pairs while improving the calibration between JOLs and recall. Our findings suggest that these encoding strategies are effective at reducing the illusion of competence, with increased calibration primarily reflecting improved recall. Thus, item-specific and relational encoding strategies primarily affect retrieval processes rather than metacognitive processes that participants engage in at encoding.

Introduction

Successfully monitoring the progress of one's learning at study is paramount for improving retention, as effective memory monitoring allows individuals to adjust their encoding strategies to maximize later retrieval (Nelson & Narens, 1990). To investigate the learning process, researchers commonly have participants make judgments of learning (JOLs) while studying material for a later test. In these tasks, participants study sets of items (often cue–target word pairs) and must estimate their likelihood of correctly retrieving the target word if prompted by the cue on a later test. Using JOLs, researchers can assess various components of learning, including how participants decide which items have been sufficiently learned and how participants adjust their encoding strategies when encountering different types of study materials. While JOLs can be elicited using various

measurement scales (e.g., Likert scales, binary ratings, etc.; see Hanczakowski et al., 2013), they are commonly made using a continuous 0–100 scale representing the percent likelihood of the target successfully being 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 memory (measured via JOLs) and actual memory (measured via test performance). Thus, JOL accuracy can be easily assessed in terms of calibration by directly comparing JOL ratings with the percentage of targets that participants correctly recall at test.¹

Several factors have been shown to affect JOL accuracy, including perceptions of fluency by studying identical cue–target word pairs (Castel et al., 2007; Mueller et al., 2016), changes in font size (Rhodes & Castel, 2008), increasing the time participants spend studying word pairs (Koriat & Ma'ayan, 2005), and changing JOL timings (i.e., delayed

✉ Mark J. Huff
mark.huff@usm.edu

¹ School of Psychology, The University of Southern Mississippi, 118 College Dr. #5025, Hattiesburg, MS 39406, USA

¹ JOL accuracy can also be assessed in terms of *resolution* or the relative accuracy between JOLs and recall (see Rhodes, 2016 for a comparison of calibration and resolution). However, in the present study, we focus on calibration, given that the illusion of competence has often been framed as miscalibration between JOLs and recall (e.g., Koriat & Bjork, 2005, 2006).

vs. immediate JOLs; Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). A highly impactful factor on JOL calibration is the direction of the association between the cue–target word pair (i.e., probability of a cue word eliciting a specific target as response; see Nelson et al., 2000). For example, associations in the forward direction between cue and the target often yield highly accurate JOL predictions, whereas associations in the backward direction or unrelated pairs often produce JOL overpredictions of later recall (e.g., *the illusion of competence*; Koriat & Bjork, 2005; Maxwell & Huff, 2021). In the present study, we further assess the relationship between associative direction and calibration by testing whether encoding tasks which emphasize shared or distinctive characteristics of cue–target pairs (i.e., item-specific and relational encoding, respectively) can improve the predictive accuracy of JOLs. Because few studies have had participants engage in these encoding tasks concurrently with JOLs, the degree to which item-specific/relational tasks may improve calibration by differentially influencing the magnitude of JOLs or cued-recall rates is presently unclear. Below, we begin by detailing the illusion of competence, including previous research which has sought to reduce this effect. We then discuss evidence suggesting that item-specific/relational encoding tasks may be effective at reducing or eliminating this metacognitive illusion.

The illusion of competence

Interest in the correspondence between JOLs and retrieval is not new. In an early example, Arbuckle and Cuddy (1969) had participants study letter–number pairs (e.g., A–73) and predict whether they would or would not remember the pairs on a later test. Overall, the authors reported that participants correctly predicted later recall for an average of 67% of trials, leading them to conclude that participants generally had insight into how difficult each pair would be to remember and that participants used their perceptions of difficulty to adjust their predictions accordingly. Subsequent research has consistently demonstrated a relatedness effect on JOLs. Related cue–target pairs (e.g., mouse–cheese) typically receive higher JOLs and are recalled to a greater extent versus unrelated pairs (e.g., mouse–onion; e.g., Koriat & Bjork, 2005; Maxwell & Huff, 2021; see Rhodes, 2016, for review). Thus, JOLs are sensitive to cue–target relations, and participants use their perceptions of cue–target relatedness to inform the magnitude of their JOLs (i.e., cue utilization; see Koriat, 1997, for review).

While perceived relatedness influences the magnitude of JOLs, certain situations occur in which relatedness is a

poor predictor of later recall. For example, Koriat and Bjork (2005) differentiated between two types of cue–target associations which differentially affect JOL calibration. First, a priori associations refer to cue–target pairs which are strong forward associates based on free association norms (e.g., credit–card, stork–baby, etc.; Nelson et al., 2004; De Deyne et al., 2019) and reflect the probability that a cue word will elicit a specific target as a response (e.g., mouse–cheese vs. mouse–?). Separately, a posteriori associations refer to any *perceived* relatedness between cue–target pairs that is apparent when words are presented together, rather than separately. While a posteriori associations have traditionally included weak forward associates (e.g., article–newspaper; see Koriat & Bjork, 2005), they may also include strong associates presented in the reverse order (i.e., backward associates like card–credit, baby–stork, etc.). Backward associates reflect a specific type of a posteriori pair, as their relatedness is only apparent when items are presented together. Thus, a posteriori pairs could potentially be weak forward associates, strong backward associates, or contain both types of cue–target relations.

To test the correspondence between JOLs and recall for both a priori and a posteriori cue–target pairs, Koriat and Bjork (2005) evaluated JOL calibration when participants studied unrelated pairs, weak forward associates, and strong forward associates/a priori pairs (Experiment 1), a priori and a posteriori pairs (e.g., strong forward and backward associates; Experiment 2), and unrelated pairs, forward associates, and semantically related a posteriori pairs that shared no forward or backward associations (e.g., bed–night; Experiment 3). Across experiments, all a posteriori pair types produced an illusion of competence pattern in which JOLs exceeded later recall rates, indicating that participants overpredicted the likelihood that they would later recall the target word. This pattern was particularly robust for backward associates, as the cue word, when presented in isolation, does not ostensibly converge upon the studied target word. Thus, though participants predict that backward associates are highly likely to be recalled at test, recall of this pair type is typically much lower than predicted.

The illusion of competence pattern reported on backward associates is robust. For example, Maxwell and Huff (2021) demonstrated that this effect occurs on backward associates even after employing various manipulations designed to improve JOL accuracy, including changes to experiment pacing (self-paced vs. experimenter paced) and JOL timing (concurrent vs. immediate vs. delayed). Furthermore, this illusion is not limited to backward associates, and extends to other pair types which similarly appear related at encoding

yet lack sufficient cues at retrieval. For example, Castel et al. (2007) found that the illusion of competence extended to identical cue–target pairs (e.g., mouse–mouse). Similarly, Maxwell and Huff (2021) reported that the illusion extends to symmetrical cue–target pairs (e.g., on–off), even after controlling for lexical characteristics which affect recall (i.e., frequency, concreteness, length, etc.). Unlike forward and backward associates, symmetrical associates contain cue–target associations which are equivalent in both directions (i.e., on–off would have equivalent associative strength as off–on). Because symmetrical associates contain both forward and backward associations, they contain equal levels of a priori and a posteriori association. Taken together, these findings suggest that the illusion of competence is persistent and emerges for any pair type that lacks an a priori forward association.

Because the illusion of competence represents a misalignment between memory and metamemory systems, researchers have sought to reduce or eliminate this pattern of overestimation by improving the predictive capacity of JOLs. For example, Koriat and Bjork (2006) found that memory overestimations could be reduced by informing participants of conditions at test (i.e., warning them about the illusion of competence), such that participants lowered their JOLs, making them more reflective of later recall. However, an alternative approach, tested in the present study, would be to have participants complete tasks at encoding which facilitate the generation of retrieval cues which both inform JOLs and aid recall. According to Koriat's (1997) cue-utilization theory, JOLs are based on several types of cues, including intrinsic cues (i.e., inherent properties of the stimuli including frequency, imageability, and cue–target relations) and extrinsic cues (i.e., external factors including study pacing and encoding manipulations such as deep vs. shallow encoding; e.g., Slamecka & Graf, 1978; see Koriat, 1997; Rhodes, 2016). Both cue types can also interact. For instance, if encoding tasks draw attention toward or away from intrinsic cues such as pair relatedness, JOL calibration may be differentially affected. Thus, the qualitative features of how a study task affects processing of cue–target pairs may impact both JOLs and recall, producing changes in calibration.

Item-specific/relational processing on memory performance

Memory researchers have long known that certain study tasks are more successful at improving retention relative to others. For example, the levels-of-processing (LOP)

framework classifies tasks that promote memory via elaborative processing as “deep” tasks, while less beneficial tasks that focus on surface or perceptual features of study items constitute “shallow” tasks (Craik, 2002; Craik & Lockhart, 1972). Several deep tasks have been identified and include generation (Slamecka & Graf, 1978), production (MacLeod et al., 2010), and survival processing (Nairne et al., 2007), to name a few. Deep tasks can be further subdivided based on the task's propensity to encourage the processing of item-specific or relational features of study pairs (i.e., the item-specific/relational framework; Einstein & Hunt, 1980; Hunt & Einstein, 1981). Based on this framework, encoding tasks differ in the likelihood that they encourage participants to process unique features of study items (i.e., item-specific processing) or shared characteristics of study items (i.e., relational processing). Thus, both item-specific and relational processing qualitatively affect encoding by changing how information is encoded in memory.

Previous research has found differential memory benefits for item-specific and relational encoding tasks. For example, McCurdy et al. (2020) showed that relational processing facilitated the generation effect on memory 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, as participants would be directed to create their own connections between items. Separately, Huff and Bodner (2014) found that item-specific tasks were more successful at improving recall and recognition when studying strong versus weakly related items. Relational tasks, however, were more beneficial for weakly versus strongly related items. Thus, tasks which affect processing of extrinsic mnemonic cues can interact with the intrinsic mnemonic cues of the study materials (e.g., Mulligan, 2011). Collectively then, despite a general classification of item-specific and relational tasks as promoting “deep” processing in the LOP framework, their relative memory benefits are affected by the association between study materials (see Huff & Bodner, 2014, for discussion).

While it is evident that deep encoding tasks benefit memory, few studies have incorporated these tasks alongside JOLs. In a recent exception, Tekin and Roediger (2020; Experiment 3) compared a deep encoding task (pleasantness ratings) to a shallow encoding task (vowel-counting) on recognition memory. Critically, half of their participants also provided JOLs at encoding, allowing for a comparison of JOLs within the context of deep and shallow encoding

tasks. While the focus of their study was on JOL reactivity (i.e., changes in memory due to providing JOLs) rather than changes in accuracy, deep encoding enhanced correct recognition without affecting the magnitude of JOLs. Tekin and Roediger's findings suggest that deep encoding tasks may be effective at reducing the illusion of competence by increasing memory performance relative to JOLs (i.e., boosting memory performance to be more aligned with predicted memory). However, based on Koriat's (1997) cue-utilization theory, deep encoding tasks would also be likely to influence the magnitude of JOLs, particularly those which emphasize intrinsic cues which participants use as the basis for their JOLs (e.g., relational encoding tasks). In the present study, we specifically test this possibility by comparing changes in JOL calibration between participants completing deep item-specific and relational encoding tasks with participants completing standard JOLs (i.e., JOLs made in the absence of an additional deep encoding task).

Finally, because the illusion of competence is contingent on cue–target associations, item-specific and relational encoding tasks may be particularly effective at influencing the correspondence between JOLs and recall, given the different effects of these tasks that are reported based on pair relatedness (e.g., Huff & Bodner, 2014). As such, we test this possibility within the context of cued-recall testing (vs. recognition), while also investigating potential interactive effects between JOL accuracy and associative direction. In doing so, the present study provides a novel approach toward reducing the illusion of competence, as to date, no study has directly investigated the effects of item-specific and relational strategies on JOL accuracy.

The present study

Given the interactive benefits of item-specific and relational encoding on associative materials (e.g., Huff & Bodner, 2014), the present study tested whether these encoding strategies could facilitate the correspondence between JOLs and recall, particularly on backward associates and unrelated pairs in which the illusion of competence is strongest (Koriat & Bjork, 2005; Maxwell & Huff, 2021). In doing so, we assessed changes in JOL accuracy for groups of participants who studied forward, backward, and symmetrical paired associates and unrelated cue–target pairs and made JOLs while engaging in item-specific encoding, relational encoding, or silent reading. Specifically, we assessed whether the use of item-specific/relational strategies would benefit JOL calibration (i.e., absolute accuracy or the degree of match

between predicted recall as measured via JOLs and actual recall at test) relative to reading. While JOL accuracy can also be described in terms of resolution (i.e., relative accuracy or the degree to which JOLs accurately discriminate between what is and is not remembered; see Rhodes, 2016), the present study focuses specifically on changes in calibration, as previous illusion of competence studies often frame this effect as arising due to a miscalibration between JOLs and recall (e.g., Castel et al., 2007; Koriat & Bjork, 2005; Maxwell & Huff, 2021).

Finally, because of our interest in the effects of item-specific/relational encoding on JOL calibration, each experiment additionally included a set of calibration plots modeled after Maxwell and Huff (2021) which assessed changes in calibration across each item type as a function of encoding strategy by visualizing changes in mean recall at various JOL increments (see also Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). Commonly, these plots use JOL increments of 10, allowing for a comparison across 11 total levels (i.e., 0–100 in multiples of 10). By including calibration plots alongside traditional mean analyses, researchers can easily assess whether JOLs are over or underconfident and, importantly, detect whether metacognitive illusions like the illusion of competence uniformly affect recall at all JOL levels or whether calibration is greater for high JOL ratings versus low ratings. Our use of calibration plots, therefore, provides a useful supplement to standard analyses of mean JOLs and recall rates, as these plots can detect qualitative changes in calibration that might otherwise be overlooked.

Experiment 1: effects of item-specific/relational encoding on JOL accuracy

The goal of Experiment 1 was to test whether having participants complete item-specific or relational tasks at encoding would reduce the illusion of competence observed on backward, symmetrical, and unrelated cue–target pairs by influencing participants JOLs, increasing correct recall, or both. Overall, we expected that relative to silent reading, having participants engage in item-specific/relational encoding tasks would reduce the illusion of competence by improving correct recall relative to the read control task, but that the relative reduction in the illusion of competence would depend upon the associative pair type. Specifically, because relational encoding encourages participants to generate associations between cue–target pairs, we expected that relational encoding would be especially beneficial for unrelated pairs where the cue is ineffective at prompting target

retrieval. Separately, because item-specific (vs. relational) processing has been shown to be more beneficial to memory when pairs are strongly related (Huff & Bodner, 2014), we expected that the item-specific task would be most beneficial for improving calibration on related pairs and, as a result, would be most effective at reducing and/or eliminating the illusion of competence for backward and symmetrical associates. For forward associates, which typically do not show an illusion of competence pattern (Maxwell & Huff, 2021), we predicted that the item-specific task could increase recall rates higher than the initial JOL ratings resulting in a situation in which JOLs *underpredict* subsequent recall.

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. Sample sizes for each group were based on Maxwell and Huff (2021), and a sensitivity analysis conducted using G*Power 3.1 (Faul et al., 2009) indicated that this sample was sufficient to detect small-to-medium main effects and interactions (Cohen's $d = 0.28$) or larger.

Materials

The stimuli used were 180 cue–target pairs taken from Maxwell and Huff (2021), which were generated using the University of South Florida Free Association Norms (Nelson et al., 2004). These pairs consisted of 40 forward associates (e.g., credit–card), 40 backward associates (e.g., card–credit), 40 symmetrical associates (e.g., salt–pepper), 40 unrelated pairs (e.g., art–lion), and 20 weakly related, non-tested buffer pairs to control for primacy and recency effects. Pairs were divided evenly into two study blocks, each containing 20 of each 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., forward associate strength (FAS) and backward associative strength (BAS) values) using the Nelson et al. (2004) free association norms. Additionally, these pairs were designed to control for lexical and semantic properties that could potentially influence recall rates, including word length, SUBTLEX frequency (Brysbaert & New, 2009), and concreteness values derived from the English Lexicon Project (Balota et al., 2007; Maxwell & Huff, 2021). Further, both study blocks were matched on 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 associates from one counterbalance became backward associates on another and vice versa. Alternating pair direction allowed for greater control of item differences, particularly on forward and backward associates, as the same items were used in the forward and backward directions across counterbalances. Pair order was similarly flipped and counterbalanced across unrelated pairs and symmetrical associates. Semantic and lexical characteristics for each pair type are reported in the Appendix (Tables 1, 2).

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. Test order was newly randomized for each participant.

Procedure

The experimental procedure followed the general procedure used by Maxwell and Huff (2021). All participants completed the study individually on computers using *E-Prime 3* software (Psychology Software Tools, Pittsburgh, PA, 2016). Participants were randomly assigned to one of the three encoding groups. 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 instructed to use the full range of the scale to help reduce anchoring.

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 given the example of the word pair “Cat–Turtle” and were instructed to 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. Item-specific and relational groups similarly completed their encoding tasks silently, and participants in both groups were instructed to use their respective strategies prior to making their JOLs. Participants only saw one type of task instruction. After receiving the encoding instructions, participants completed a ten-word practice set using their assigned encoding task. Following completion of the practice trials, participants were required to describe their study strategy to the experimenter in their own words before starting the experiment. Participants were then given their first block of word pairs to study at their own pace and provided their JOL ratings while the word pair was displayed. Finally, after studying half of the pairs, participants were presented with a quick reminder to use their respective encoding strategy.

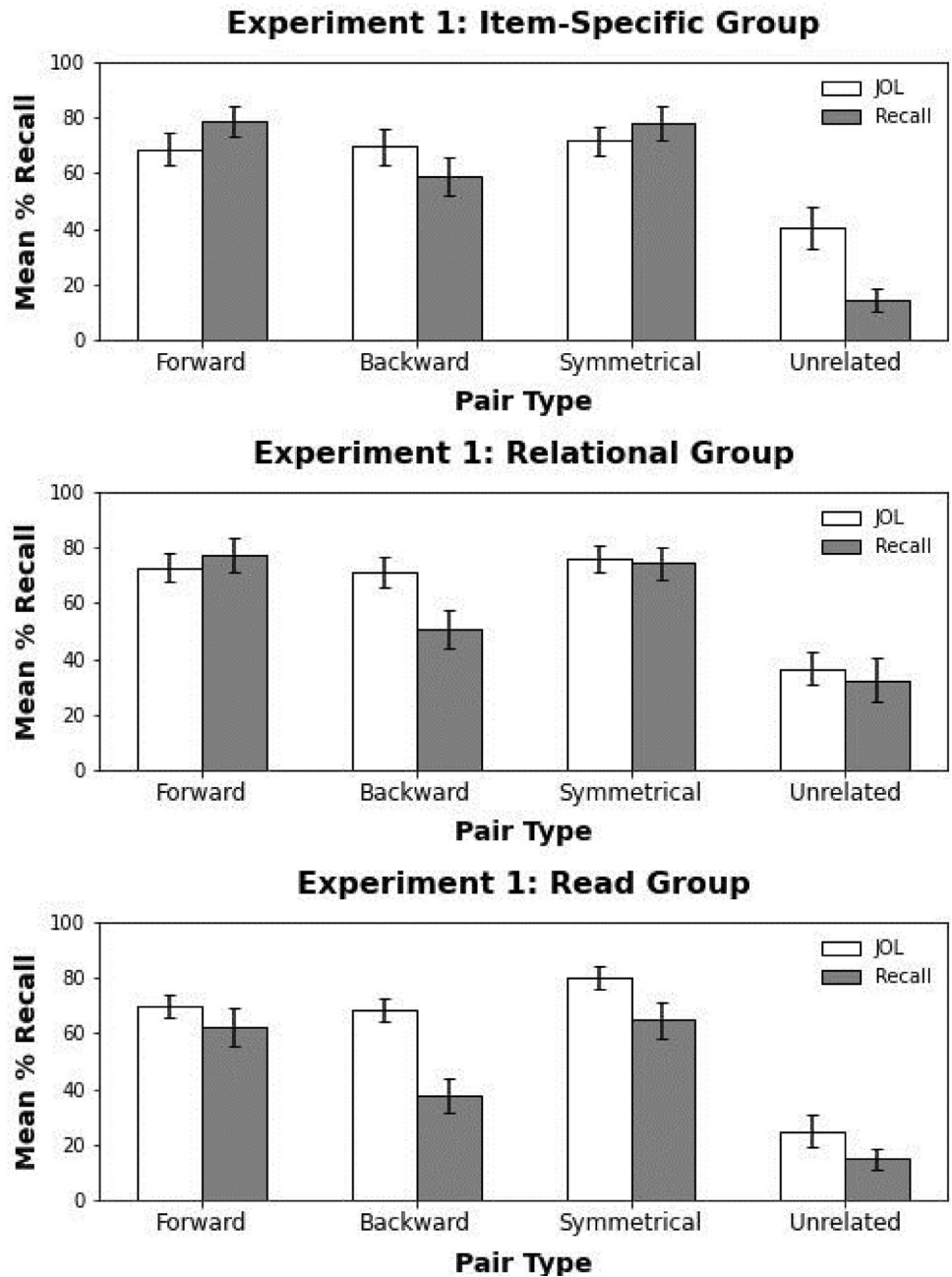
Following the first study list, participants completed a 2-min arithmetic filler task which was directly followed by a cued-recall test. This test presented the first word from each pair, and participants were instructed to recall the target word from memory. Participants were encouraged not to leave 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 h.

Results

Prior to conducting the analyses, study items that were missing JOL ratings or had ratings that were outside of the 0–100 range were removed. These responses were rare and fewer than 0.5% of items were removed. 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; <https://osf.io/svzg8/>), and we note that the data patterns were similar between blocks. 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 $p < 0.05$ significance level was used unless noted otherwise. For all reported non-significant comparisons, we further analyzed the strength of the evidence supporting the null hypothesis using a Bayesian estimate (Masson, 2011; Wagenmakers, 2007). In this analysis, a model that assumes an effect is compared to a model that assumes a null effect. This process 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 using standard null hypothesis significance testing.

Mean JOL and recall rates as a function of pair type are reported in Fig. 1. For completeness, all comparisons are reported in Appendix Table 3. Additionally, while we primarily focus on changes in calibration, we report analyses assessing changes in the Appendix. 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 compared differences between mean JOL ratings and recall rates across pair types and encoding groups. An effect of Measure was found, $F(1, 85) = 18.79$, $MSE = 694.46$, $\eta_p^2 = 0.18$, such that collapsed across encoding groups and pair types, mean JOL ratings exceeded later recall rates (62.66 vs. 54.19). Next, an effect of Encoding Group was detected, $F(2, 85) = 5.40$, $MSE = 814.98$, $\eta_p^2 = 0.11$, in which JOL ratings/recall rates were significantly higher for the relational (61.44) and

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



item-specific (60.12) groups relative to the read-only group (53.33). All comparisons differed significantly, $t_s \geq 2.96$, $d_s \geq 0.78$, except for the relational and item-specific groups, which were equivalent, $t < 1$, $p_{BIC} = 0.87$. Finally, an effect of Pair Type was found, $F(3, 255) = 766.58$, $MSE = 107.66$, $\eta_p^2 = 0.90$, in which JOL ratings/recall rates were higher for symmetrical associates (74.22), followed by forward associates (72.29), backward associates (59.01), and unrelated

pairs (27.55). Comparisons across all pair types differed statistically, $t_s \geq 2.69$, $d_s \geq 0.17$.

A significant two-way interaction between Measure and Pair Type was found, $F(3, 255) = 56.94$, $MSE = 87.42$, $\eta_p^2 = 0.40$, which indicated that the illusion of competence pattern depended upon the pair type studied. Follow-up comparisons indicated that mean JOLs approximated later recall for forward associates (70.62 vs. 73.95, respectively;

$t(87) = 1.46$, $SEM = 2.33$, $p = 0.15$, $p_{BIC} = 0.77$) and symmetrical associates (75.82 vs. 72.63; $t(87) = 1.27$, $SEM = 2.56$, $p = 0.21$, $p_{BIC} = 0.81$). However, an illusion of competence pattern occurred for backward associates, as mean JOLs exceeded recall (69.96 vs. 49.24; $t(87) = 8.32$, $SEM = 2.52$, $d = 1.17$), a pattern which similarly extended to unrelated pairs (34.21 vs. 20.89; $t(87) = 5.22$, $SEM = 2.59$, $d = 0.72$).

Critically, a significant three-way interaction was also found, $F(6, 255) = 15.56$, $MSE = 87.42$, $\eta_p^2 = 0.27$, in which the magnitude of the illusion of competence differed as a function of encoding group. Starting with backward associates, reliable illusion of competence patterns were detected across encoding groups, though at different rates. In the read-control group, a robust illusion of competence was detected in which JOLs greatly exceeded later recall accuracy (68.62 vs. 37.78), $t(27) = 9.44$, $SEM = 3.41$, $d = 2.19$. For the item-specific group, JOLs also exceeded recall (69.55 vs. 59.01), $t(28) = 2.16$, $SEM = 5.12$, $d = 0.58$, though at a lesser magnitude relative to the read condition. A similar pattern was observed in the relational group, where JOLs exceeded recall, but again at a lower rate than the read group (71.55 vs. 50.49), $t(30) = 5.41$, $SEM = 4.05$, $d = 1.18$.

For forward associates, an illusion of competence pattern was not found for any of the encoding groups, with JOLs matching later recall for both the read group (70.04 vs. 65.23), $t(27) = 1.32$, $SEM = 3.42$, $p = 0.19$, $p_{BIC} = 0.69$, and the relational group (72.96 vs. 77.22), $t(30) = 1.15$, $SEM = 3.86$, $p = 0.26$, $p_{BIC} = 0.74$. For the item-specific group, however, JOLs were lower than later recall rates (68.67 vs. 78.84), $t(28) = 2.42$, $SEM = 4.41$, $d = 0.65$ —a situation in which JOLs underestimated later recall.

For symmetrical associates, the illusion of competence was moderated by the encoding task. For the read group, JOLs exceeded later recall accuracy (80.22 vs. 64.85), $t(27) = 3.59$, $SEM = 4.48$, $d = 1.06$; however, for both the item-specific and relational groups, the illusion of competence did not emerge as JOLs were equivalent to subsequent recall rates (71.62 vs. 78.24), $t(28) = 1.41$, $SEM = 4.90$, $p = 0.17$, $p_{BIC} = 0.66$, and (75.77 vs. 74.41), $t < 1$, $SEM = 3.46$, $p = 0.67$, $p_{BIC} = 0.83$, 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 (40.64 vs. 14.35), $t(28) = 5.71$, $SEM = 4.81$, $d = 1.56$, as JOLs exceeded later recall. However, the illusion of competence was not found in the relational group (36.59 vs. 32.52), $t < 1$, $SEM = 4.52$, $p = 0.35$, $p_{BIC} = 0.78$, indicating that relational encoding provides a unique benefit on unrelated pairs by improving the correspondence between JOLs and subsequent recall.

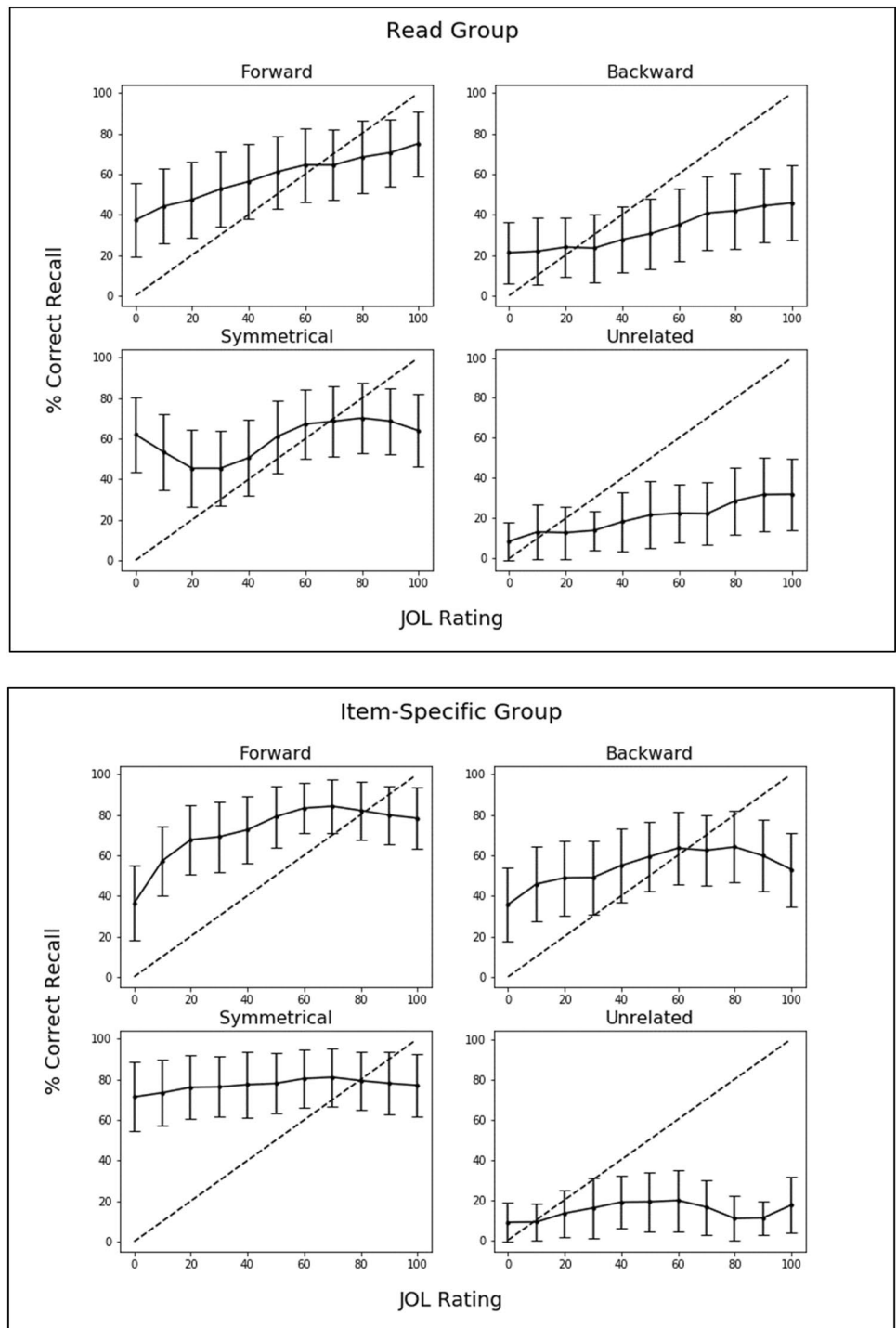
Taken together, item-specific and relational processing tasks were each found to reduce and/or eliminate the illusion of competence pattern, but these reductions depended upon the pair type studied. Both item-specific and relational tasks were successful at eliminating the illusion of competence for symmetrical associates and reducing it for backward associates relative to reading. However, relational encoding eliminated the illusion of competence on unrelated pairs, while the illusion of competence persisted on unrelated pairs for item-specific encoding. The different task effects on reducing the illusion of competence appeared to reflect an increase in correct recall rather than an adjustment of JOL ratings.

To test this possibility, we conducted separate Encoding Group \times Pair Type ANOVAs on recall and JOLs. For recall, effect of encoding group emerged, $F(2, 85) = 6.49$, $MSE = 919.34$, $\eta_p^2 = 0.13$, with recall rates greatest in the item-specific ($M = 57.62$) and relational groups ($M = 58.67$), relative to the read group ($M = 45.68$; $ts \geq 3.13$, $ds \geq 0.57$), with the item-specific and relational groups being equivalent, $t < 1$, $p_{BIC} = 0.88$. Mean JOL rates, however, were stable across the three encoding groups, $F(2, 85) < 1$, $MSE = 147.50$, $p = 0.59$, $p_{BIC} = 0.98$, though encoding group interacted with pair type, $F(6, 255) = 7.27$, $MSE = 107.67$, $\eta_p^2 = 0.15$. For unrelated pairs, JOLs increased when participants used item-specific ($M = 40.64$) and relational encoding ($M = 36.59$) strategies relative to the read group ($M = 24.85$; $ts \geq 2.82$, $ds \geq 0.73$). Differences between the item-specific and relational encoding groups were equivalent, $t < 1$, $p_{BIC} = 0.84$. For related pairs, however, JOLs did not differ as a function of encoding strategy, except for symmetrical associates in which mean JOLs were higher for participants in the read group ($M = 80.20$) compared to the item-specific encoding group ($M = 71.65$; $t(55) = 2.49$, $SEM = 3.52$, $p = 0.02$, $d = 0.66$). All other JOL comparisons involving related pairs were non-significant, $ts \leq 1.34$, $p_{BICs} \geq 0.76$.

Calibration plots

We further assessed the absolute accuracy between JOLs and recall for each pair types using a series of calibration plots (cf. Maxwell & Huff, 2021). To generate these plots, JOLs were first rounded to the nearest 10% increment, which resulted in 11 JOL bins ranging from 0 to 100%. For example, the 0% JOL increment contains the proportion of correct recall for items given a judgment of 0%, the 10% increment contains the proportion of correct recall for items given a judgment of 10%, etc. Mean correct recall for each JOL bin was then plotted. By plotting mean recall as function of JOL bin, these calibration plots allowed us to qualitatively assess

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

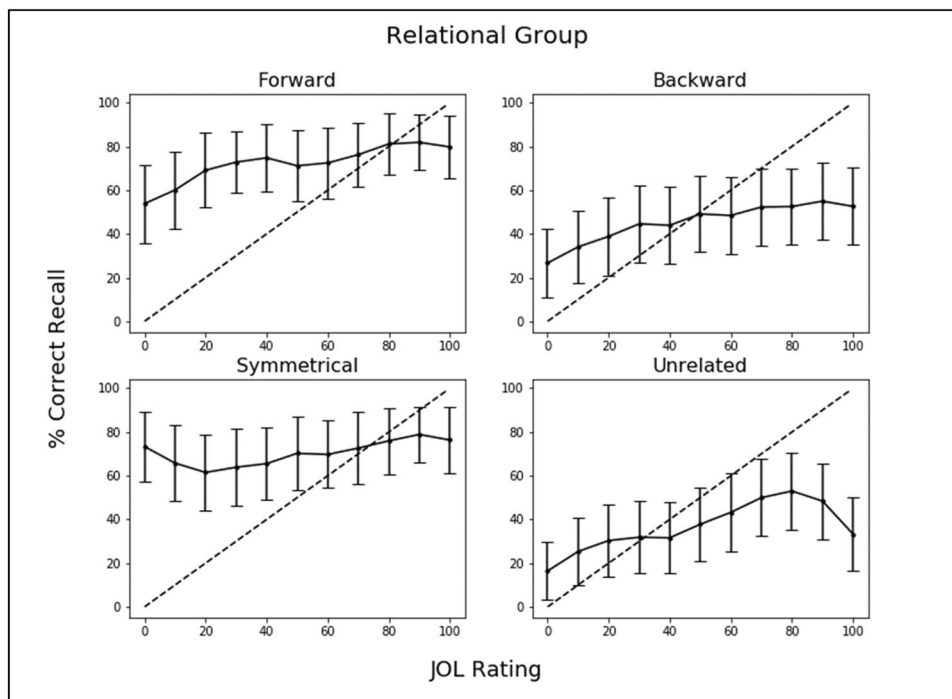


whether the illusion of competence uniformly affected recall across JOL levels (e.g., Maxwell & Huff, 2021) and whether changes in the illusion of competence occurred as a function

of encoding type (i.e., whether overestimations emerged at different JOL increments based on encoding instructions).

Figure 2 displays calibration plots for each encoding group as a function of pair direction. Plots are structured

Fig. 2 (continued)



such that they include a calibration line denoting a perfect correspondence between JOL ratings and mean correct recall (e.g., a 40% JOL and a 40% correct recall rate would be perfectly calibrated). Overestimations are reflected by data points falling below the calibration line. Underestimations are represented by data points falling above the calibration line. These plots revealed 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 associates, where illusions of competence are generally not found, all encoding groups showed similar calibration patterns. For unrelated pairs and backward associates, the illusion of competence pattern emerged at higher JOL increments in the item-specific/relational encoding groups relative to the read group. Item-specific encoding was most effective at increasing the JOL increment in which the illusion of competence pattern was detected for backward pairs (> 80%), while relational encoding was most effective at increasing the JOL increment for unrelated pairs (> 50%), again demonstrating qualitative differences in how item-specific and relational encoding at improving JOL accuracy.

These patterns were confirmed using a 3 (Encoding Group: Item-Specific vs. Relational vs. Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) \times 11 (JOL Increment) mixed ANOVA. Overall, this

analysis yielded significant main effects of Pair Type, $F(3, 252) = 90.75$, $MSE = 1485.79$, $\eta_p^2 = 0.51$, and JOL Increment, $F(10, 840) = 24.97$, $MSE = 1805.17$, $\eta_p^2 = 0.23$. Importantly, a significant interaction was detected between Pair Type and JOL Increment, $F(30, 2520) = 6.93$, $MSE = 919.81$, $\eta_p^2 = 0.08$, confirming the presence of an illusion of competence pattern. However, the interactions between JOL Increment and Encoding Group were non-significant, $F_s \leq 1.27$, $p \geq 0.19$, $p_{BICs} > 0.99$.

Discussion

Experiment 1 tested whether item-specific and relational encoding manipulations could reduce the illusion of competence by improving the correspondence between JOLs and recall. Relative to the read-only control group, both item-specific and relational encoding strategies were effective at reducing the illusion of competence for backward and symmetrical associates. Furthermore, and consistent with our predictions, relational encoding was effective at reducing the illusion of competence for unrelated pairs. These patterns were confirmed via calibration plots, which provided qualitative information regarding the JOL bin in which overestimations emerge. For backward associates, item-specific encoding increased this bin relative to relational encoding

and the read-only control group, while unrelated pairs showed a similar increase when relational encoding was utilized. Thus, our findings in Experiment 1 suggest that item-specific and relational encoding manipulations are each effective at reducing the illusion of competence by improving the calibration between JOLs and recall.

While Experiment 1 demonstrated that item-specific/relational tasks can effectively reduce the illusion of competence for backward and symmetrical associates and unrelated pairs, we acknowledge that because participants applied their encoding strategies silently, we are unable to verify separately that item-specific and relational encoding tasks were consistently applied. Several experiments have reported reliable item-specific/relational processing differences using these same task instructions (e.g., Huff & Bodner, 2013, 2014); however, we sought to confirm participants' use of item-specific and relational processing through the use of a "think aloud" procedure. In this procedure, participants are not only instructed and required to practice their given study strategy, but also required to state aloud the item-specific or relational characteristics for each word pair with an experimenter present to provide feedback. We therefore applied this procedure in Experiment 2 to ensure the application of item-specific and relational encoding processes while also providing a replication of Experiment 1.

Experiment 2: think-aloud encoding

Experiment 2 sought to replicate item-specific and relational findings from Experiment 1 by verifying that participants were consistently engaging in their instructed item-specific or relational encoding strategy. In doing so, participants in the item-specific and relational encoding groups applied their strategies using a vocal, think-aloud procedure in which they verbalized their encoding processes for each study pair. Overall, we expected that findings from Experiment 1 would replicate. Specifically, both item-specific and relational encoding strategies were expected to reduce the illusion of competence, with item-specific encoding being most effective on backward associates and relational encoding most effective on unrelated pairs.

Methods

Participants

One-hundred and two University of Southern Mississippi undergraduate students completed Experiment 2 for partial course credit. Participants were randomly assigned to one of

three encoding groups: item-specific encoding ($n=34$), relational ($n=32$), or the read-only group ($n=36$). The sample size was modeled after Experiment 1 with the constraint that each group contain at least 30 participants. A sensitivity analysis conducted via G*Power 3.1 (Faul et al., 2009) confirmed that our sample had sufficient power (0.80) to detect small-to-medium main effects/interactions (Cohen's $d=0.26$).

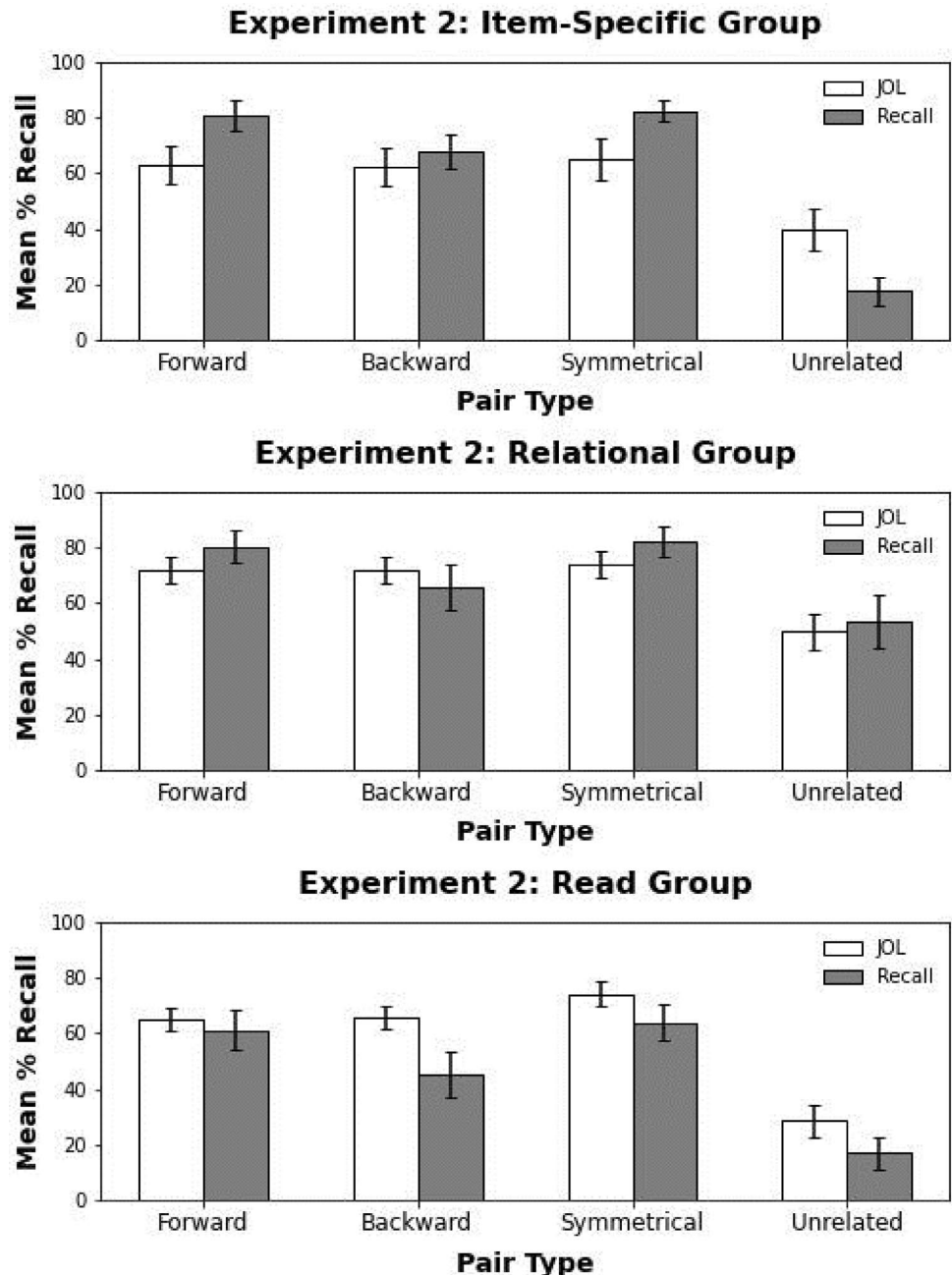
Materials and procedure

Experiment 2 used the same materials as Experiment 1 and followed the same general procedure with the following two exceptions. First, after receiving their respective encoding strategies, the item-specific and relational encoding groups received additional instruction to vocalize their thought processes aloud during encoding. For example, a participant encoding the pair "mouse–cheese" with a relational strategy might state that these concepts are related because mice eat cheese, while a participant encoding the same pair with an item-specific strategy might instead state that these items differ because mice are animals while cheese is a type of food. Second, given the additional time needed for participants to implement this think-aloud procedure, Experiment 2 only consisted of one study-test block. All other procedures, including the practice trials and filler task, were identical to Experiment 1.

Results

Figure 3 displays mean JOL and recall percentages as functions of pair type and encoding group. For completeness, all comparisons are reported in Table 4, and analyses assessing changes in resolution are reported in the Appendix. Like Experiment 1, differences between mean JOLs and recall were tested via 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. An effect of measure was not found, $F(1, 99) < 1$, $MSE = 981.87$, $p = 0.59$, $p_{BIC} = 0.90$, as overall, mean JOLs did not differ from mean recall (60.85 vs. 59.39, respectively). An effect of encoding group was found, $F(2, 99) = 23.48$, $MSE = 749.85$, $\eta_p^2 = 0.32$, as mean JOLs/recall percentages were highest for the participants in the relational encoding group (69.73), followed by participants in the item-specific (59.83) and read groups (52.12; $t_s \geq 3.05$, $d_s \geq 0.73$). Additionally, an effect of Pair Type was detected, $F(3, 297) = 359.85$, $MSE = 181.82$, $\eta_p^2 = 0.78$, in which JOLs/recall rates were highest for symmetrical pairs (73.60), followed by forward (70.15), backward (62.86), and

Fig. 3 Mean JOL and recall rates as a function of pair type in the Item-Specific group (top panel), Relational group (middle panel), and the Read group (bottom panel) in Experiment 2. Bars represent 95% confidence intervals



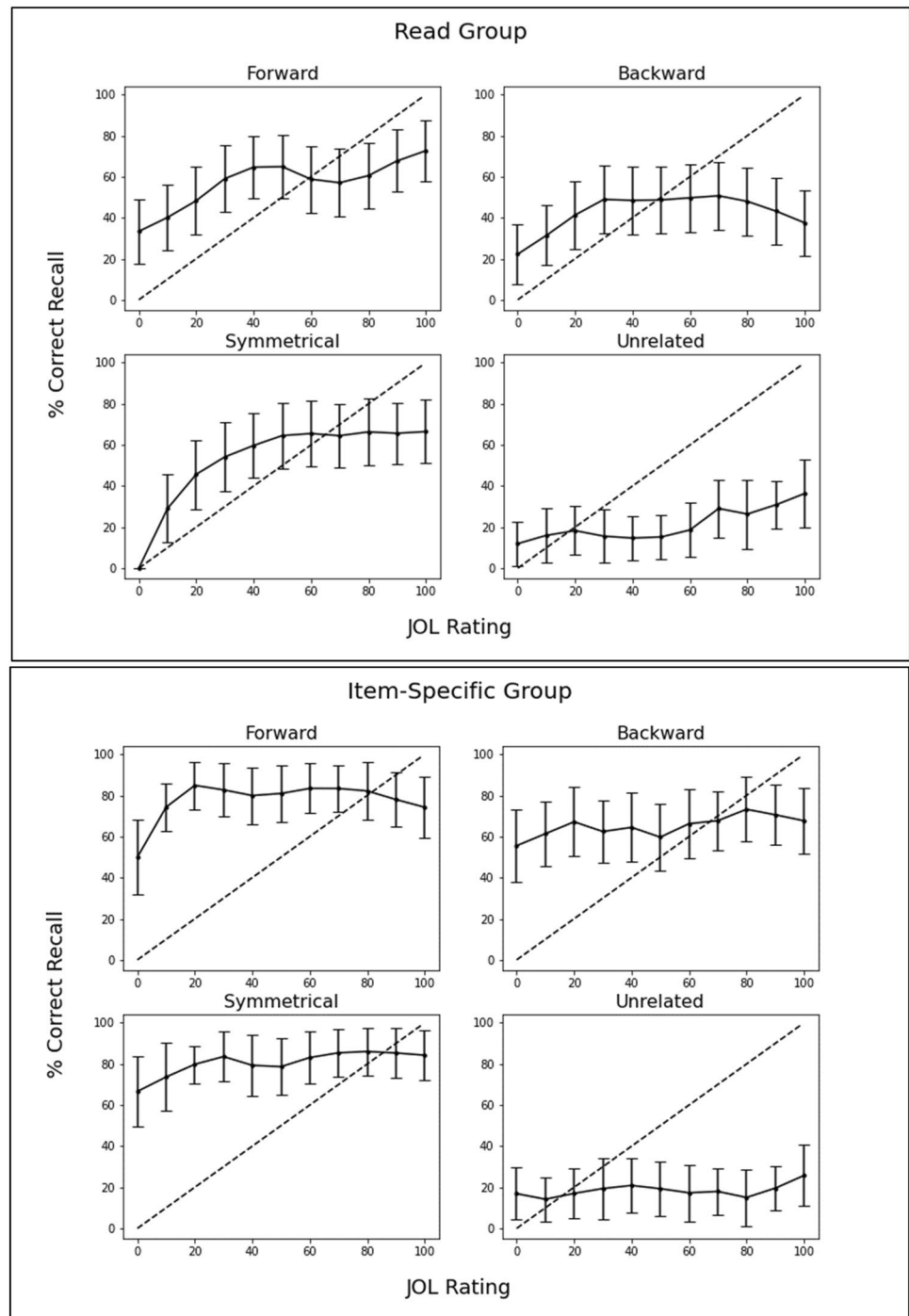
unrelated pairs (33.82). All comparisons differed statistically, $t_s \geq 2.04$, $d_s \geq 0.29$.

A significant Measure \times Direction interaction was also found, $F(3, 297) = 22.64$, $MSE = 168.97$, $\eta_p^2 = 0.19$, which confirmed the presence of an illusion of competence pattern. Across encoding groups, mean JOLs were underconfident for forward associates (66.58 vs. 73.72), $t(101) = 2.59$, $SEM = 2.79$, $d = 0.38$ but were overconfident for backward associates (66.55 vs. 59.16), $t(101) = 2.54$, $SEM = 2.92$,

$d = 0.35$ and unrelated pairs (39.01 vs. 28.64), $t(101) = 3.41$, $SEM = 3.08$, $d = 0.43$. However, for symmetrical associates, JOLs did not differ from recall (71.22 vs. 75.99), $t(101) = 1.68$, $SEM = 2.84$, $p = 0.10$, $p_{BIC} = 0.76$.

Importantly, a significant three-way interaction confirmed that illusion of competence pattern differed as a function of Encoding Group. Beginning with backward associates, a robust illusion of competence was detected in the read group, such that JOLs greatly exceeded later recall (65.86

Fig. 4 Calibration plots as a function of pair direction in the Read Group (top panel), Item-Specific Group (middle panel), and Relational Group (bottom panel) in Experiment 2. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence interval

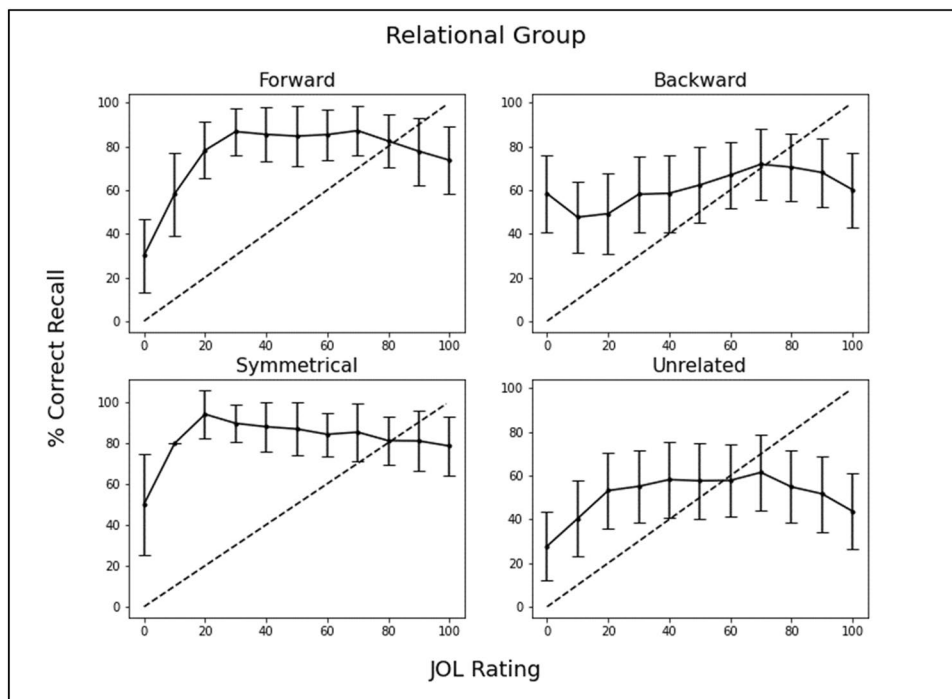


vs. 45.26), $t(35) = 4.09$, $SEM = 5.20$, $d = 1.03$. However, this pattern did not extend to the item-specific group, as JOLs and recall did not significantly differ (62.22 vs. 67.70), $t(33) = 1.34$, $SEM = 4.26$, $p = 0.19$, $p_{BIC} = 0.70$. Relational encoding similarly eliminated the illusion of competence

(71.92 vs. 65.73), $t(31) = 1.28$, $SEM = 5.03$, $p = 0.21$, $p_{BIC} = 0.71$.

Regarding forward associates, no illusion of competence patterns emerged across any of the encoding groups. JOLs did not statistically differ from recall for participants in the read group (65.11 vs. 61.22), $t < 1$, $SEM = 4.30$, $p = 0.35$,

Fig. 4 (continued)



$p_{\text{BIC}} = 0.79$. Furthermore, consistent with Experiment 1, JOLs in the item-specific group were lower than subsequent recall (62.96 vs. 80.71), $t(33) = 3.53$, $SEM = 5.24$, $d = 0.97$. Finally, JOLs in the relational group marginally underestimated later recall (72.06 vs. 80.38), $t(31) = 1.84$, $SEM = 4.61$, $p = 0.08$, $p_{\text{BIC}} = 0.52$.

The illusion of competence was again moderated by encoding task for symmetrical associates. Starting with the read group, JOLs exceeded later recall (74.32 vs. 64.06), $t(35) = 2.27$, $SEM = 4.71$, $d = 0.60$. However, item-specific encoding produced a noticeable underestimation pattern on this pair type, with JOLs below recall (65.13 vs. 82.52), $t(33) = 4.03$, $SEM = 4.49$, $d = 0.96$, while relational encoding produced JOLs that were marginally underestimated (74.19 vs. 82.44), $t(31) = 1.80$, $SEM = 4.65$, $p = 0.08$, $p_{\text{BIC}} = 0.53$.

Finally, the illusion of competence was observed on unrelated pairs for participants in both the read (28.58 vs. 16.99), $t(35) = 2.71$, $SEM = 4.42$, $d = 0.64$, and the item-specific groups (39.78 vs. 17.63), $t(33) = 4.33$, $SEM = 5.33$, $d = 1.17$. However, these JOL overestimations did not extend to participants in the relational group, as JOLs and recall did not statistically differ (49.84 vs. 53.29), $t < 1$, $SEM = 5.89$, $p = 0.54$, $p_{\text{BIC}} = 0.82$, replicating patterns observed in

Experiment 1 and providing additional evidence that relational encoding uniquely benefits unrelated pairs.

Calibration plots

Figure 4 displays calibration plots for Experiment 2. Starting with related pairs, both item-specific and relational encoding strategies influenced the correspondence between JOLs and recall. Forward and symmetrical associates were generally well calibrated for participants in the read group; however, participants completing the item-specific/relational encoding tasks greatly underestimated later recall, with recall approximating JOLs at only the highest JOL increments (> 80%). Consistent with Experiment 1, the illusion of competence pattern occurred at higher JOL increments for participants in the item-specific and relational groups relative to the read group. Furthermore, item-specific and relational encoding were equally effective at increasing the JOL increment in which the illusion of competence emerged on backward associates (> 70% for both groups), while relational encoding was most effective for unrelated pairs (> 60%).

These patterns were supported via a 3 (Encoding Group: Item-Specific vs. Relational vs. Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) \times 11 (JOL Increment) mixed ANOVA. This analysis yielded significant

effects of Pair Type, $F(3, 294) = 67.33$, $MSE = 2191.31$, $\eta_p^2 = 0.03$, and JOL Increment, $F(10, 980) = 33.02$, $MSE = 917.76$, $\eta_p^2 = 0.03$. Additionally, a significant effect of Encoding Group was detected, $F(2, 98) = 6.07$, $MSE = 6404.05$, $\eta_p^2 = 0.11$. Next, a significant interaction Pair Type \times JOL Increment confirmed the presence of the illusion of competence, $F(30, 2940) = 8.56$, $MSE = 1433.22$, $\eta_p^2 = 0.08$. The JOL Increment \times Encoding Group interaction was marginally significant, $F(20, 980) = 1.53$, $MSE = 2191.31$, $p = 0.06$, $p_{BIC} > 0.99$, while the three-way interaction was non-significant, $F(60, 2940) = 1.11$, $MSE = 1008.06$, $p = 0.26$, $p_{BIC} > 0.99$.

Discussion

Experiment 2 tested whether our findings in Experiment 1 that item-specific and relational encoding strategies reduce the illusion of competence would replicate. Importantly, participants in Experiment 2 completed the item-specific and relational tasks aloud rather than silently, which allowed us to ensure that they were consistently and correctly applying their respective strategies at encoding. Consistent with Experiment 1, item-specific and relational encoding tasks each reduced the illusion of competence relative to participants in the read-only control group. Specifically, item-specific encoding was effective at reducing the illusion on backward and symmetrical associates, but not on unrelated pairs. Relational encoding, however, was similarly effective on backward and symmetrical associates while also reducing the illusion of competence on unrelated pairs. These patterns were further explored via calibration plots, which qualified these patterns. Thus, findings from Experiment 2 largely replicated patterns reported in Experiment 1, further indicating that item-specific and relational encoding strategies effectively reduce the illusion of competence. Additionally, the replication of these patterns provides additional evidence that participants were correctly applying their respective encoding strategies in Experiment 1.

Finally, because of the additional encoding due to participants vocalizing their encoding strategies versus completing the task silently, it is likely that the inclusion of the think-aloud protocol further benefitted recall (i.e., a production effect; McLeod et al., 2010), though it is unclear whether this task would affect JOL magnitudes. To account for this possibility, the Appendix includes a set of cross-experimental analyses which compare differences in the illusion of competence between Experiments 1 and 2. While the full set of analyses is reported in the Appendix, we note memory

benefits from the think-aloud procedure were moderated by pair type, with only the more difficult backward and unrelated pairs showing an improvement from Experiment 1 to Experiment 2. Importantly, the magnitude of JOLs did not differ between experiments. Thus, our inclusion of the think-aloud procedure in Experiment 2 successfully ensured that participants remained attentive to their respective encoding tasks without influencing the magnitude of their JOLs.

General discussion

The goal of the present study was to reduce the illusion of competence by improving the accuracy of JOL ratings on predicting later recall of forward, backward, and symmetrical associates, and unrelated cue–target word pairs. Previous research has consistently found that JOLs over predict recall of unrelated and deceptive backward associates, resulting in an illusion of competence pattern (Koriat & Bjork, 2005; Maxwell & Huff, 2021). Manipulations to remedy this metacognitive illusion have often focused on improving JOL calibration via manipulations primarily affecting the magnitude of participants' JOLs (i.e., reducing JOLs to more closely approximate recall; see Koriat & Bjork, 2006). However, because previous work has shown memory benefits when deep processing is used in conjunction with JOLs (Tekin & Roediger, 2020), Experiment 1 first sought to further qualify deep-processing effects by comparing item-specific and relational encoding—separate encoding tasks that each promote deep processing. Experiment 2 was then designed to replicate Experiment 1 while also including a think-aloud procedure to ensure that participants applied their instructed encoding strategies.

Across experiments, forward associates did not produce an illusion of competence pattern across any of the three encoding groups. However, consistent with previous research (e.g., Koriat & Bjork, 2005; Maxwell & Huff, 2021), a robust illusion of competence emerged for backward pairs, and smaller illusion patterns were found for symmetrical associates and unrelated pairs. Importantly, these illusions were moderated by encoding task. In Experiment 1, participants in the standard, read-only JOL group showed a robust illusion of competence on backward associates, which extended to symmetrical associates and unrelated pairs. For the item-specific group, JOLs underpredicted later recall of forward associates (cf. Castel et al., 2007; Koriat & Bjork, 2005). The illusion of competence again occurred on backward associates, though it was reduced compared to both

the read and relational encoding groups, a pattern consistent with our prediction that item-specific encoding would be most beneficial in reducing the illusion of competence for related pairs (cf. Huff & Bodner, 2014). For symmetrical associates, item-specific encoding eliminated the illusion of competence. Unrelated pairs also showed an illusion of competence pattern, though this was reduced relative to the read group. Finally, for the relational encoding task, the illusion of competence pattern was eliminated for unrelated pairs, but as with item-specific encoding, the pattern was reduced but not eliminated for backward associates versus participants in the read group. For symmetrical associates, the illusion of competence was again reduced.

These patterns extended to Experiment 2, in which participants completed the item-specific/relational encoding tasks aloud. Both strategies eliminated the illusion of competence on backward and symmetrical pairs, and, consistent with Experiment 1, only relational encoding was effective on unrelated pairs. Furthermore, relative to Experiment 1, each encoding strategy was more effective, often producing underconfidence patterns in which JOLs underestimated recall. These patterns likely resulted due to the additional encoding afforded by the think-aloud procedure in Experiment 2, which further elevated recall rates relative to JOLs. Additionally, our finding of similar data patterns in Experiments 1 and 2 suggests that participants were effectively applying their respective encoding strategies in Experiment 1, even when required to complete these tasks silently. Taken together, findings from both experiments indicate that both item-specific and relational encoding tasks improve JOL accuracy versus a read task, though their relative effectiveness depends upon the associative direction of the pair type.

Following analysis of mean JOL and recall rates, we constructed a series of calibration plots which assessed differences in absolute accuracy between JOLs and recall for each pair type as a function of encoding task. These plots reflected qualitative differences in JOL overestimation between encoding groups, particularly for backward associates and unrelated pairs. Starting with Experiment 1, read group participants were overconfident for unrelated pairs at all JOL increments and for backward associates at all increments above 50%, a pattern consistent with Maxwell and Huff (2021). Thus, overestimation was most likely to occur for pairs in which relatedness cues used at encoding were not readily available at retrieval, replicating previous work on the illusion of competence (e.g., Koriat & Bjork, 2005; Maxwell & Huff). For the item-specific group, participants were again overconfident for unrelated pairs at almost all JOL increments, but overconfidence of backward associates occurred only at higher JOL increments relative to reading (80% vs. 50%, respectively). Finally, for the relational group, overconfidence for backward associates emerged at

JOL increments greater than 60%, and for unrelated pairs at increments greater than 50%. Thus, compared to the read and item-specific tasks, relational encoding greatly improved participants' abilities to accurately predict their own recall for unrelated pairs, suggesting that unrelated pairs are particularly benefitted by relational encoding strategies. Finally, across all groups, participants were generally well calibrated for forward and symmetrical associates.

These patterns then extended to Experiment 2. Relative to the read-only group, item-specific encoding again improved the correspondence between JOLs and recall for backward associates and unrelated pairs, thus reducing the illusion of competence. Similarly, relational encoding was again most effective at improving calibration on unrelated pairs. Therefore, across experiments, item-specific and relational encoding strategies benefitted calibration whenever the target word did not readily converge upon the cue at retrieval.

Although the present study is the first to employ the item-specific/relational framework to improve JOL accuracy, we note that Senkova and Otani (2021) compared recall performance for words receiving JOLs relative to lists studied using two item-specific encoding tasks (pleasantness ratings and single-mental imagery) that did not provide concurrent JOLs. Overall, neither item-specific encoding task increased recall relative to JOLs on either related or unrelated word lists. However, compared to a read-only task that did not provide JOLs, both JOLs and item-specific encoding tasks boosted correct recall, leading the authors to conclude that the act of providing JOLs at study recruits item-specific processing. The present study, however, showed that when combined with JOLs, item-specific encoding strategies boost correct recall relative to standard, read-only JOLs. Thus, it is possible that item-specific encoding may produce an additional memory benefit when combined with JOLs. Of course, the test type also differed between Senkova and Otani and our experiment (free-recall vs. cued-recall), so it is unknown whether the benefit of JOLs combined with item-specific encoding compared to JOLs alone would occur using other test types.

Finally, we note that our results complement work by Tekin and Roediger (2020), who showed that JOLs facilitated recognition memory for levels-of-processing (LOP) encoding tasks. Specifically, JOLs were particularly beneficial to memory when participants engaged in them alongside shallow encoding tasks (e.g., an e-counting task), such that the LOP effect (i.e., memory benefits of deep vs. shallow encoding; see Craik & Lockhart, 1972) was eliminated. While the present study was not designed to test the separate effects of JOLs on memory (i.e., JOL reactivity; Maxwell & Huff, 2022; Rivers et al., 2021; Soderstrom et al., 2015) or specifically test recall within the LOP framework as we did not include a shallow encoding group, we note that both item-specific and relational encoding strategies constitute

deep encoding tasks. Our finding that recall was greater when participants combined JOLs with these encoding strategies relative to making JOLs alone is consistent with previous research on the benefits item-specific/relational strategies (e.g., Hunt & Einstein, 1981) and, furthermore, suggests that Tekin and Roediger's findings may extend to cued recall. However, given that the present study did not include control groups who only engaged in only item-specific or relational encoding without also making JOLs, more research will be needed to test the potential for additivity with JOLs. Regardless, however, the benefits of item-specific/relational encoding appear to interact with different types of associative pairs which indicates that there may be boundaries in which relational and item-specific encoding tasks can reduce illusions of competence.

Taken together, our findings that item-specific and relational encoding strategies consistently reduced the illusion of competence via improved recall are consistent with previous research indicating that these tasks modify recall without affecting the magnitude of JOLs (e.g., Tekin & Roediger, 2020). This lack of change in JOL magnitude is somewhat surprising when considered alongside Koriat's (1997) cue-strengthening account. Because item-specific and relational strategies emphasize each emphasize qualitative aspects of the stimuli, they likely direct participants' attention towards intrinsic cues which inform the magnitude of their JOLs. Thus, these tasks would be expected to influence JOLs, particularly when participants engage in relational tasks at encoding, given that these tasks specifically emphasize pre-existing cue–target relations. However, across experiments, item-specific and relational encoding tasks each increased JOLs for unrelated pairs, though the magnitude of JOLs for related cue–target pairs remained unchanged. Thus, when pairs contain pre-existing relations, the use of additional encoding strategies likely provides little benefit on intrinsic cues, particularly when cue–target relations are strong. Alternatively, when participants are required to complete these tasks when studying pairs with weak intrinsic cues (e.g., unrelated pairs), JOLs are increased, as the additional retrieval cues afforded by the item-specific/relational encoding tasks likely enhance participants' predictions of later memory. Thus, future JOL studies may wish to explore the interactive effects of item-specific and relational encoding strategies and cue–target relations on JOL accuracy.

Conclusion

The present study found that the illusion of competence can be reduced when participants are directed to engage in item-specific or relational strategies at encoding. Across experiments, we found that the illusion of competence for backward and symmetrical associates can be reduced via item-specific encoding and that overestimation of unrelated pairs is reduced when participants use a relational encoding

strategy. Calibration plots revealed that item-specific and relational encoding tasks generally improved the correspondence between JOLs and recall across pair types. Thus, item-specific and relational tasks can be used to reduce, but not eliminate, the illusion of competence for backward, symmetrical, and unrelated word pairs, which appears to be the product of enhanced calibration between JOLs and subsequent recall.

Appendix

For both experiments, we assessed whether item-specific or relational encoding instructions affected the resolution between JOLs and recall. Relative accuracy or *resolution* refers to the degree to which a person's JOL rating discriminates between what is and what is not remembered (Rhodes, 2016). Unlike calibration, which can be assessed through plots, resolution is commonly assessed via Goodman–Kruskal gamma correlations. The gamma coefficient represents a measure of association between -1 and $+1$, with resolution decreasing as gamma approaches zero. Positive values denote the degree that remembered items were given high JOLs and non-remembered items low JOLs, while negative gamma values denote the inverse of this pattern (Nelson, 1984). While the illusion of competence is generally assessed in terms of calibration (e.g., Koriat & Bjork, 2005), we note that item-specific and relational encoding strategies may additionally improve resolution, given that resolution is affected whenever an encoding task affords participants with an opportunity to adjust their JOLs (i.e., modifying JOLs based on previous trials). Thus, for completeness, we report a series of analyses assessing changes in resolution for each experiment.

Experiment 1: resolution

Following the procedure used by Nelson and colleagues (Dunlosky & Nelson, 1992, 1994; Nelson, 1984), we computed Goodman–Kruskal gamma correlations (G) between JOLs and recall for each participant for each of the four pair types (forward, backward, symmetrical, and unrelated; Table 5 reports mean G s and 95% CI s as functions of pair type and encoding group). To test for changes in resolution, we assessed differences in mean G using a 3 (Encoding Group: Item-Specific vs. Relational vs. Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA. Overall, main effects/interactions were only marginally significant, F s ≥ 1.94 ; p s ≤ 0.07 , $p_{BICs} > 0.99$; however, planned follow-up analyses were still carried out.

For forward pairs, both item-specific and relational encoding resulted in reduced resolution compared to silent

Table 2 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 |

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)

reading (0.10 vs. 0.13 vs. 0.35, respectively). All comparisons differed significantly ($t_s \geq 2.56$, $d_s \geq 0.64$), except for the comparison between item-specific and relational encoding, $t < 1$, $p = 0.97$, $p_{BIC} = 0.88$. This pattern subsequently extended to backward pairs (0.12 vs. 0.07 vs. 0.24), though only the comparison between the relational encoding and read groups was significant, $t(57) = 2.34$, $SEM = 0.07$, $d = 0.60$, and all other comparisons for backward pairs were non-significant, $t_s \leq 1.63$, $p_s \geq 0.11$, $p_{BIC} \geq 0.67$. For symmetrical pairs, *G* was again lower for item-specific and relational encoding relative to the read group (0.15 vs. 0.13 vs. 0.23), however, all comparisons failed to reach conventional significance, $t_s \leq 1.53$, $p_s \geq 0.13$, $p_{BIC} \geq 0.70$. Finally, for unrelated pairs, resolution was increased for participants who completed item-specific (0.26) and relational encoding tasks (0.33) relative to participants in the read group (0.20). However, again, all comparisons failed to reach significance, $t_s \leq 1.06$, $p_s \geq 0.29$, $p_{BIC} \geq 0.81$. Thus, while item-specific and relational encoding strategies are effective at reducing the illusion of competence, this reduction appears to occur primarily due to changes in calibration rather than resolution.

Table 1 Mean associative strength summary statistics forward, backward, and symmetrical pairs

| Condition | Variable | <i>M</i> | <i>SD</i> | Min | Max |
|-------------|----------|----------|-----------|------|------|
| Forward | FAS | 0.37 | 0.21 | 0.05 | 0.81 |
| | BAS | 0.00 | 0.00 | 0.00 | 0.00 |
| Backward | FAS | 0.00 | 0.00 | 0.00 | 0.00 |
| | BAS | 0.37 | 0.21 | 0.05 | 0.81 |
| Symmetrical | FAS | 0.19 | 0.13 | 0.01 | 0.46 |
| | BAS | 0.19 | 0.13 | 0.02 | 0.52 |

FAS (forward associative strength) and BAS (backward associative strength) values for unrelated pairs as these items share zero associative overlap

Experiment 2: resolution

Next, we assessed whether item-specific or relational encoding instructions influenced the resolution between JOLs and recall (see Table 6 for Mean *G*s and 95% *CI*s for all comparisons). A 3 (Encoding Group: Item-Specific vs. Relational vs. Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA was used to test for differences in resolution as functions of encoding group and pair type. Overall, this analysis yielded a significant main effect of encoding group, $F(2, 99) = 3.59$, $MSE = 0.24$, $\eta_p^2 = 0.07$. Collapsed across pair types, resolution was greater for participants in the read group (0.19) relative to the item-specific (0.10) and relational encoding groups (0.03). All comparisons were non-significant, $t_s \leq 1.04$, $p_s \geq 0.30$, $p_{BIC} \geq 0.82$, except for the comparison between the read and relational groups, $t(66) = 3.01$, $SEM = 0.05$, $d = 0.74$. Additionally, this analysis revealed a significant effect of pair type, $F(3, 297) = 4.29$, $MSE = 0.19$, $\eta_p^2 = 0.04$. Post hoc testing indicated that resolution was greatest for unrelated pairs (0.19), followed by symmetrical pairs (0.17), forward pairs (0.08), and backward pairs (0.01). Resolution for backward pairs was significantly lower relative to symmetrical and unrelated pairs, $t_s \geq 3.22$, $d_s \geq 0.37$, though comparison between all other pair types were non-significant, $t_s \leq 1.49$, $p_s \geq 0.14$, $p_{BIC} \geq 0.72$. Additionally, the Encoding Group \times Pair Type interaction was non-significant, $F(6, 297) = 1.69$, $MSE = 0.19$, $p = 0.12$, $p_{BIC} > 0.99$. Thus, like Experiment 1, item-specific and relational encoding reduced the illusion of competence primarily through improved calibration than resolution.

Cross-experimental analysis

Because participants in the item-specific and relational encoding groups in Experiment 2 were required to verbalize their encoding processes, it is possible that this procedure

Table 3 Comparison of mean JOL ratings and correct recall percentages across pair directions for each encoding group in Experiment 1

| Encoding group | Task | Direction | <i>M</i> | 95% CI | <i>F</i> | <i>B</i> | <i>S</i> |
|----------------|--------|-------------|----------|--------|----------|----------|----------|
| Item specific | JOL | Forward | 68.67 | 5.95 | 0.05 | 0.13 | 1.74* |
| | | Backward | 69.55 | 6.41 | | | |
| | | Symmetrical | 71.62 | 5.24 | | | |
| | | Unrelated | 40.64 | 7.49 | | | |
| | Recall | Forward | 78.84 | 5.47 | 1.16* | 1.08* | 4.42* |
| | | Backward | 59.01 | 6.85 | | | |
| | | Symmetrical | 78.24 | 6.05 | | | |
| | | Unrelated | 14.35 | 4.35 | | | |
| | | | | 4.75* | | | |
| Relational | JOL | Forward | 72.96 | 4.86 | 0.08 | 0.29* | 2.66* |
| | | Backward | 71.55 | 5.52 | | | |
| | | Symmetrical | 75.77 | 4.82 | | | |
| | | Unrelated | 36.59 | 5.90 | | | |
| | Recall | Forward | 77.22 | 6.09 | 1.44* | 1.30* | 1.95* |
| | | Backward | 50.49 | 6.96 | | | |
| | | Symmetrical | 74.41 | 5.94 | | | |
| | | Unrelated | 32.52 | 8.08 | | | |
| | | | | 2.07* | | | |
| Read | JOL | Forward | 70.04 | 3.89 | 0.13 | 1.00* | 4.11* |
| | | Backward | 68.62 | 4.39 | | | |
| | | Symmetrical | 80.22 | 4.20 | | | |
| | | Unrelated | 24.85 | 5.68 | | | |
| | Recall | Forward | 62.23 | 6.96 | 1.40* | 1.64* | 3.51* |
| | | Backward | 37.78 | 5.91 | | | |
| | | Symmetrical | 64.85 | 6.34 | | | |
| | | Unrelated | 14.76 | 3.96 | | | |
| | | | | 3.11* | | | |

Mean JOL and recall rates for each associative direction condition across each encoding group. The three right-most columns indicate Cohen's *d* effect sizes for post hoc comparisons, * $p < 0.05$

affected the magnitude of the JOLs and/or their recall performance. We tested this possibility using a 2 (Experiment) \times 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. The only reliable interaction that emerged was the Experiment \times Measure \times Direction interaction, $F(3, 552) = 3.94$, $MSE = 128.35$, $\eta_p^2 = 0.02$. All other interactions with Experiment, including the four-way interaction, were non-significant, $F_s \leq 2.02$ $p_s \geq 0.06$, $p_{BICs} \geq 0.64$.

Overall, collapsed across encoding groups, mean JOL ratings did not differ between Experiments 1 and 2 for forward pairs (70.23 vs. 66.58, respectively), $t(188) = 1.67$, $SEM = 2.23$, $p = 0.10$, $p_{BIC} = 0.77$, or backward pairs (69.26 vs. 66.55), $t(188) = 1.19$, $SEM = 2.29$, $p = 0.24$, $p_{BIC} = 0.87$. For symmetrical pairs, JOLs in Experiment 1 were marginally greater than Experiment 2 (75.35 vs. 71.22), $t(188) = 1.81$, $SEM = 2.32$, $p = 0.07$, $p_{BIC} = 0.73$, while JOLs for unrelated pairs were marginally lower in Experiment 1 relative to Experiment 2 (33.69 vs. 39.01), $t(188) = 1.81$,

$SEM = 2.94$, $p = 0.07$, $p_{BIC} = 0.72$. Thus, across pair types, having participants engage in the think-aloud procedure in Experiment 2 did not affect their JOLs.

Regarding recall, no differences emerged between experiments for forward pairs (73.92 vs. 73.72), $t < 1$, $SEM = 2.87$, $p = 0.92$, $p_{BIC} = 0.93$, or symmetrical pairs (72.70 vs. 75.99), $t(188) = 1.22$, $SEM = 2.64$, $p = 0.22$, $p_{BIC} = 0.87$. However, for backward pairs, recall was greater in Experiment 2 than Experiment 1 for backward pairs (49.27 vs. 59.16), $t(188) = 3.01$, $SEM = 3.33$, $d = 0.44$, and unrelated pairs (20.91 vs. 28.64), $t(188) = 2.27$, $SEM = 3.41$, $d = 0.33$. Thus, the additional encoding afforded by the think-aloud task boosted recall, but only for more challenging backward and unrelated pairs. Importantly however, the item-specific and relational encoding effects produced similar effects on reducing the illusion of competence on both experiments, demonstrating that participants were indeed applying item-specific and relational processing tasks effectively in Experiment 1 when encoding was completed silently.

Table 4 Comparison of mean JOL ratings and correct recall percentages across all pair directions for each encoding group in Experiment 2

| Encoding group | Task | Direction | <i>M</i> | 95% CI | <i>F</i> | <i>B</i> | <i>S</i> | |
|----------------|------------|-------------|-------------|--------|----------|----------|----------|-------|
| Item specific | JOL | Forward | 62.96 | 6.87 | 0.03 | 0.13 | 1.13* | |
| | | Backward | 62.23 | 6.85 | | | | |
| | | Symmetrical | 65.13 | 7.59 | | | | |
| | | Unrelated | 39.78 | 7.41 | | | | |
| | Recall | Forward | 80.71 | 5.47 | 0.74* | 0.94* | 4.71* | |
| | | Backward | 67.70 | 6.26 | | | | |
| | | Symmetrical | 82.52 | 4.05 | | | | |
| | | Unrelated | 17.63 | 5.14 | | | | |
| | Relational | JOL | Forward | 72.03 | 4.76 | 0.01 | 0.16 | 1.49* |
| | | | Backward | 71.92 | 4.82 | | | |
| | | | Symmetrical | 74.19 | 4.92 | | | |
| | | | Unrelated | 49.84 | 6.36 | | | |
| Recall | | Forward | 80.38 | 5.98 | 0.69* | 0.82* | 1.28* | |
| | | Backward | 65.73 | 8.31 | | | | |
| | | Symmetrical | 82.45 | 5.39 | | | | |
| | | Unrelated | 53.29 | 9.73 | | | | |
| Read | | JOL | Forward | 65.11 | 4.28 | 0.05 | 0.64* | 2.87* |
| | | | Backward | 65.86 | 4.15 | | | |
| | | | Symmetrical | 74.32 | 4.50 | | | |
| | | | Unrelated | 28.58 | 5.84 | | | |
| | Recall | Forward | 61.23 | 7.33 | 0.67* | 0.82* | 2.45* | |
| | | Backward | 45.26 | 8.19 | | | | |
| | | Symmetrical | 64.06 | 6.60 | | | | |
| | | Unrelated | 16.99 | 5.94 | | | | |

Mean JOL and recall rates for each associative direction condition across each encoding group. The three right-most columns indicate Cohen’s *d* effect sizes for post hoc comparisons, **p* < 0.05

Table 5 Mean (\pm 95% CI) Goodman–Kruskal gamma correlations between JOLs and recall for each encoding group as a function of pair type in Experiment 1

| Encoding Group | Forward | Backward | Symmetrical | Unrelated |
|----------------|--------------|--------------|--------------|--------------|
| Read | 0.35 (0.12)* | 0.24 (0.13)* | 0.23 (0.10)* | 0.20 (0.18)* |
| Item specific | 0.10 (0.11) | 0.12 (0.13) | 0.15 (0.16) | 0.26 (0.16)* |
| Relational | 0.13 (0.11)* | 0.07 (0.07) | 0.13 (0.14) | 0.33 (0.10)* |

*Significant from zero, *p* < 0.05

Table 6 Mean (\pm 95% CI) Goodman–Kruskal gamma correlations between JOLs and recall for each encoding group as a function of pair type in Experiment 2

| Encoding group | Forward | Backward | Symmetrical | Unrelated |
|----------------|--------------|-------------|--------------|--------------|
| Read | 0.20 (0.13)* | 0.07 (0.12) | 0.30 (0.11)* | 0.21 (0.17)* |
| Item Specific | 0.02 (0.18) | 0.11 (0.15) | 0.22 (0.18)* | 0.30 (0.19)* |
| Relational | 0.02 (0.20) | 0.01 (0.14) | 0.04 (0.17) | 0.12 (0.14) |

*Significant from zero, *p* < 0.05

Additionally, we examined experiment differences in calibration plots and resolution. First, cross-experimental differences in calibration plots were assessed via a 2 (Experiment) \times 3 (Encoding Group: Item-Specific vs. Relational vs. Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) \times 11 (JOL Increment) mixed ANOVA. Overall, this analysis yielded a significant Experiment \times Pair Type interaction, $F(3, 546) = 12.57$, $MSE = 1640.37$, $\eta_p^2 = 0.12$. However, all other interactions, including the four-way interaction, failed to reach significance, $F_s \leq 1.69$, $ps \geq 0.08$, $p_{BICs} > 0.99$. Regarding resolution, a 2 (Experiment) \times 3 (Encoding Group: Item-Specific vs. Relational vs. Read) \times 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA confirmed that mean *G* did not differ as a function of experiment, as no interactions with Experiment were detected, $F_s \leq 1.72$, $ps \geq 0.16$, $p_{BICs} > 0.99$. Thus, changes in calibration and resolution across pair types/encoding groups did not differ between experiments.

Author contributions Study design and conceptualization were completed by NPM and MJH. NPM completed all analyses and prepared

figures. NPM, EEC, and MJH all contributed to writing the manuscript. EEC wrote the first draft and NPM and MJH provided revisions.

Data availability Study materials and analyzed data are available via OSF (<https://osf.io/x9n4f/>). Supplemental Materials have been made available at <https://osf.io/svzgz8/>. This study was completed as part of the Honors Thesis requirements for EEC. NPM is now at Midwestern State University.

Declarations

Conflict of interest The studies reported were approved by the University of Southern Mississippi Institutional Review Board (Protocol #IRB-18-15) and found to be in accordance with the 1964 Helsinki Declaration ethical principles. Informed consent was obtained from all individuals who participated in this study. The authors report no competing interests.

Open practices statement The data for all experiments have been made available at <https://osf.io/x9n4f/>. Neither experiment was pre-registered.

References

- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126–131.
- Balota, D. A., Yap, M. J., 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.
- 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.
- 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 and Review*, *14*(1), 107–111.
- Craik, F. I. M. (2002). Levels of processing: Past, present ... and future? *Memory*, *10*(5–6), 305–318.
- 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.
- 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.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory and Cognition*, *20*(4), 374–380.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–565.
- Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology*, *6*(5), 588–598.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.
- 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.
- Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1246–1256. <https://doi.org/10.1037/a0031338>.
- Huff, M. J., & Bodner, G. E. (2014). All varieties of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language*, *73*, 43–58.
- 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.
- Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology*, *126*(4), 349–370.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology*, *31*(2), 187–194.
- Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners’ sensitivity to retrieval conditions at test. *Memory & Cognition*, *34*, 927–959.
- 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.
- 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*, *36*(3), 671–685.
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*(3), 679–690.
- Maxwell, N. P., & Huff, M. J. (2021). The deceptive nature of associative word pairs: Effects of associative direction on judgments of learning. *Psychological Research Psychologische Forschung*, *85*(4), 1757–1775.
- Maxwell, N. P., & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning*, *17*, 589–625.
- 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.
- Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people’s metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *The Quarterly Journal of Experimental Psychology*, *69*(4), 781–799.
- Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *The Quarterly Journal of Experimental Psychology*, *64*(8), 1543–1562.
- Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology*, *33*(2), 263–273.

- Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory and Cognition*, 28(6), 887–899.
- 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, and Computers*, 36(3), 402–407.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychonomic Bulletin*, 95(1), 109–133.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, 2, 267–270.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, 26, 125–173.
- Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from <https://www.pstnet.com>
- Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 90–117). Oxford.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology*, 137(4), 615–625.
- Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory*, 29(10), 1342–1353.
- Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory and Cognition*, 49, 955–967.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology*, 4(6), 592–604.
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology*, 41(2), 553–558.
- Tekin, E., & Roediger, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift Für Psychologie*, 228(4), 278–290.
- Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin and Review*, 14(5), 779–804.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Psychological Research is a copyright of Springer, 2024. All Rights Reserved.