Title here: Something about mediated pairs and relational encoding

Nicholas P. Maxwell1 & Mark J. Huff 2

1 Midwestern State University, 2 The University of Southern Mississippi

**Author Note**

Correspondence concerning this article should be addressed to Nicholas P. Maxwell, Department of Psychology, Midwestern State University, 3410 Taft Blvd, Wichita Falls, TX, United States. Study materials, data files, and *R* code used for analyses have been made available via OSF (LINK).

Abstract

[ABSTRACT HERE]

*Keywords*: Judgments of Learning; Reactivity; Mediated Associates; Cued-Recall; Recognition

Title Here

Memory researchers have long been interested in whether individuals can accurately gauge their learning progress. Countless studies have used judgments of learning (JOLs) to investigate questions related to memory monitoring, such as whether participants can successfully modify their study strategies based on task-demands (see Nelson & Narens, 1990). In a traditional JOL study, participants are presented with a study list (often cue-target paired associates) and are instructed to rate their likelihood of correctly recalling some aspect of the studied material on a later test. When using cue-target word pairs, these judgments are typically framed as the probability of recalling the target at test if prompted by the cue. While particiants can provide JOLs using a variety of scales (see Hanczakowski, Zawadzka, Pasek, & Higham, 2013), they are commonly instructed to make these ratings using a continuous 0-100 scale, which is framed as the probability of correctly recall the pair’s target item at test. Thus, JOLs provide a convenient tool for assessing metamemory accuracy, as the correspondence between predicted and actual recall can be readily assessed through a simple comparison process (Rhodes, 2016).

Early research using JOLs often treated these ratings as neutral measures which had no influence on memory. However, a growing body of evidence suggests the opposite, indicating instead that JOLs are *reactive* on learning (see Double et al., 2018, for review). Such reactivity occurs whenever the presence of a measure influences participants’ performance in some way (Ericsson & Simon, 1993). Regarding JOLs, the simple act of providing these judgments at study alters participants’ memory for studied materials, likely by calling attention to aspects of the stimuli that participants might not otherwise have attended to, potentially producing memorial benefits (i.e., *positive reactivity*) or costs (i.e., *negative reactivity*). Testing for potential reactivity effects is a straightforward process, requiring simply that memory performance for participants making JOLs be compared to a separate group of participants completing a no-JOL control task (e.g., silent reading). However, until recently, studies investigating metamemorial processes via JOLs often omitted this additional comparison group, either because JOLs were assumed to be neutral on memory or because they were primarily concerned with JOL accuracy (e.g., Koriat & Bjork, 2005; Rhodes & Castel, 2008, etc.) rather than directly investigating potential effects of these judgments on later memory.

Studies investigating JOL reactivity commonly test for potential memory changes by having participants study a mix of related and unrelated cue-target word pairs. A consistent finding is that making JOLs produces positive reactivity, but only when pairs are strong forward associates (e.g., cat – dog). For unrelated pairs (e.g., cat – sky), no reactivity is observed, with cued-recall rates not differing between JOL and control groups (e.g., Janes, Rivers, & Dunlosky, 2018; Maxwell & Huff, 2022; Soderstrom, Clark, Halamish, & Bjork, 2015; etc.). In notable exception, Mitchum, Kelley, and Fox (2016) reported a discrepant pattern in which JOLs were not reactive on related pairs and, instead, produced negative reactivity on unrelated pairs. However, subsequent studies have been unable to replicate this pattern, and furthermore, a meta-analysis conducted by Double, Birney, and Walker (2018), which analyzed results from 17 JOL studies, found evidence of positive reactivity on related pairs but no reactivity on unrelated pairs. Thus, it is evident that JOLs are reactive when participants study cue-target word pairs. However, these effects are moderated by relatedness, such that positive reactivity is generally observed on related but not unrelated pairs.

To explain this pattern, Soderstrom et al. (2015) proposed that JOL reactivity occurs whenever two criteria are met. First, the act of providing JOLs must direct participants’ attention towards aspects of the study pairs which they might otherwise overlook. For instance, when making JOLs, participants use intrinsic properties of the stimuli as indicators of future recall ability (i.e., intrinsic cues; see Koriat, 1997). Because pair relatedness is a strong predictor of later test performance, participants often use this cue when assigning JOLs. In doing so, the act of making JOLs strengthens these cues, but only for related pairs, in which these cues are readily available. However, because unrelated pairs lack inherent relatedness cues, no cue-strengthening occurs for this pair type. Second, any cues JOLs strengthen at encoding must additionally be made available at test. Related pairs, therefore, demonstrate positive reactivity when cued-recall testing is used, as this test type naturally prompts recall of the missing target. As a result, related word pairs demonstrate a positive reactivity effect, as naturally occurring relatedness cues are both strengthened at encoding and utilized at test. However, unrelated pairs show no reactivity given their lack of relatedness cues.

The cue-strengthening account aligns with the general pattern of reactivity that is consistently reported for cue-target pairs (i.e., positive reactivity on related pairs, no reactivity on unrelated pairs; e.g., Janes et al., 2018; Maxwell & Huff, 2022; Maxwell & Huff, in press; Rivers, Janes, & Dunlosky, 2021). Additional research supports Soderstrom et al.’s (2015) claim that reactivity only occurs when the test utilizes cues strengthened at encoding. To test this, Myers, Rhodes, and Hausman (2020) compared reactivity effects between cued-recall and free-recall testing (Experiments 1 and 2) and recognition testing (Experiments 3 and 4). Overall, reactivity patterns observed with cued-recall testing extended to recognition testing, as this test type similarly emphasizes cues used at encoding. However, for free-recall, in which no cues are provided, reactivity failed to occur, regardless of whether pairs were related. These findings were additionally replicated by Chang and Brainard (in press; Experiment 3), who similarly showed no positive reactivity on related pairs when free-recall testing was used. Thus, positive reactivity on related cue-target pairs is likely driven by cue utilization at encoding, as these effects are only observed when cues are available at test.

**Cue-Strengthening and Relational Encoding**

Whereas previous research has directly assessed the relationship between cue-strengthening and reactivity by manipulating the type of test participants complete at encoding, comparatively few studies directly assessed the types of cues which JOLs are purported to strengthen. Instead, previous studies investigating reactivity have often assumed that JOLs modify memory by primarily calling attention to relatedness cues, rather than other intrinsic which participants could potentially utilize when forming their JOLs (e.g., concreteness, item frequency, relatedness, etc., see Dunlosky & Matvey, 2001; Koriat, 1997). This emphasis on pair relatedness cues is likely because studies investigating JOL reactivity generally present participants with cue-target pairs at encoding. While cue-target pairs contain several intrinsic cues, pair relatedness is typically the most salient, particularly for strongly related paired associates. Thus, it is likely that when participants provide JOLs for cue-target pairs, their ratings largely reflect perceived pair relatedness (see XXX).

Recently, Maxwell and Huff (2022) investigated the role of relational encoding processes on reactivity by comparing JOLs to two additional judgment tasks—judgments of associative memory (JAMs; Maki, 2007; Maxwell & Buchanan, 2020) and frequency of co-occurrence judgments—each of which similarly emphasized cue-target relations while removing the metacognitive component associated with JOLs (i.e., neither judgment required making a memory prediction). Like JOLs, JAMs and frequency judgments each produced reactivity patterns mirroring JOLs (i.e., positive reactivity on related pairs, no reactivity on unrelated pairs; e.g., Janes et al., 2018; Soderstrom et al., 2015), providing further evidence that JOLs encourage participants to process cue-target relations at encoding. Based on these findings, the authors concluded that JOL reactivity likely reflects the use of a relational encoding strategy, which is triggered by when participants are asked to judge aspects of cue target pairs. However, relational processing from JOLs is applied selectively to pairs as a function of relatedness. Thus, positive reactivity occurs on related pairs, while no reactivity occurs whenever pairs are unrelated.

To test the strategic nature of this account, Maxwell and Huff (2022) included an additional experiment which compared JOLs to an explicit relational encoding task, which instructed participants to apply relational encoding across all pair types, regardless of relatedness. [HOW IS THIS DIFFERENT FROM SELECTIVE ENCODING?]

Separately, [HALAMISH PAPER]

Taken together, [WHAT DO THESE STUDIES TELL US ABOUT THE ROLE OF RELATEDNESS?]

**The Present Study**

[PLAY UP IMPORTANCE OF RELATEDNESS; TRANSITION TO MEDIATED ASSOCIATES]

[THE PRESENT STUDY] [RECALL VS. RECOGNITION]

**Experiment 1: Cued-Recall Testing**

The goal of Experiment 1 was to test the relational component of JOL reactivity. Overall, we expected that previously reported reactivity patterns would replicate on forward associates and unrelated pairs (i.e., positive reactivity on forward associates, no reactivity on unrelated pairs). Additionally, [WHAT ARE THE MEDIATED PAIR PREDICTIONS]. [WHY DOES THIS MATTER?]

**Method**

**Participants**

Participant recruitment was conducted using two platforms. First, XX undergraduate psychology students were recruited from the University of Southern Mississippi and completed the study in exchange for partial course credit. Next, an additional XX participants completed Experiment 1 via Prolific (www.prolific.co) and were compensated at a rate of $4.50/half-hour. Participants were randomly assigned to either the JOL (*n* = XX) or no-JOL (*n* = XX) encoding groups. Our final sample of XX participants was based on an a priori power analysis conducted using *G\*Power 3.1* (Faul, Erdfelder, Lang, & Buchner, 2007), which indicated that 28 participants would be required to detect medium main effects/interactions (Cohen’s *d* = 0.50). However, we extended participant recruitment due to an anticipated increase in response variability from online testing. All participants were native English speakers, and Prolific participants were additionally required to have obtained at least a high school degree or equivalent.

**Materials**

To create the stimuli pairs, 60 cue-target word pairs were taken from Maxwell and Huff (2021). These pairs consisted of 30 forward associates originally developed using the University of South Florida Free Association norms (e.g., mouse – cheese; Nelson, McEvoy, & Schreiber, 2004) and 30 unrelated pairs (e.g., muffin – floor). Next, an additional 30 mediated associates (e.g., beach – box) were taken from Balota and Lorch (1986) and Jones (2010), resulting in a total of 90 cue-target pairs the final stimuli set. The stimuli were then split into two lists, which contained 15 pairs of each type (forward, unrelated, and mediated), resulting in a total of 45 pairs in each study list. For both lists, pair types were matched on several lexical variables which could potentially influence recall, including SUBTLEX frequency (Brysbaert & New, 2009), concreteness values derived from the English Lexicon Project (Balota et al., 2007), and word length. Additionally, forward pairs within each list were matched on FAS (see Appendix Tables AX-AX for associative strength and lexical properties for each pair type). To account for primacy and recency effects, study lists were arranged such that the 45 tested pairs were always preceded and followed by an additional five non-tested buffer items. The final set of stimuli pairs is available at [OSF LINK]. Finally, two cued-recall tests were created by taking the cue items from the 45 tested pairs and replacing the target item with a question mark (e.g., mouse – ?).

**Procedure**

Experiment 1 was administered online using Collector, an open-source program for presenting web-based psychological experiments (Garcia & Kornell, 2015). Following informed consent, participants in both the JOL and no-JOL groups were told that they would be studying a series of cue-target word pairs and that their memory for the target item in each pair would later be tested. After receiving this initial set of instruction, participants in the JOL groups were further informed that while studying, they would be asked to rate their likelihood later recalling the target item if prompted by the cue. Participants in the JOL group were instructed to make these ratings using a 0-100 scale, such that a rating of 0 represented the participant being sure they would not remember the target, while a rating of 100 indicated that the participant was completely sure they would remember the correct target item. Participants in the no-JOL group were simply instructed to read each pair silently. After receiving their respective encoding instructions, participants were presented with the first study list. For both groups, encoding was self-paced, with participants pressing the ENTER key to move to the next pair. Participants in the JOL group provided their ratings concurrently with study, such that JOLs were elicited while the pair was displayed on the computer screen. List items were randomized for each participant, with the caveat that each list always began and ended with the same five buffer items across participants.

After completing the first study list, participants were given a two-minute filler task which required them to alphabetize the 50 US states. Once the time-limit was reached, participants immediately began the cued-recall test, which individually presented them with the first word from each of the previously studied pairs (minus buffer items) in a randomized order. Each item was structured as a cue-target pair, with the missing target item represented by a question mark. Participants were instructed to type the missing word for each pair and were additionally informed that if they could not retrieve an item, they could advance to the next pair. The cued-recall test was self-paced, with participants typing their response and pressing the ENTER key to move to the next pair. Once the cued-recall test was completed, participants immediately began the second block, which structured identically as the first. Thus, participants completed two study/test cycles. To account for potential block effects, block order was counterbalanced across participants. Following completion of the second study/test block, participants were debriefed. For both groups, the experiment took approximately 30 minutes to complete.

**Results**

For all analyses,significance was set at the *p* < .05 level. Additionally, for all significant analyses of variance (ANOVAs) and t-tests, we report partial eta squared (XX) and Cohen’s *d* effect size measures. Finally, for all non-significant main effects, interactions, and post-hoc comparisons, provide a Bayesian estimate regarding the strength of evidence in support of the null hypothesis (see Masson, 2011; Wagenmakers, 2007). This additional analysis compares a model in which a significant effect is assumed to secondary model assuming a null effect, which yields a probability estimate representing the likelihood of the null hypothesis being retained (termed *p*BIC; Bayesian information criterion). Because *p*BIC is sensitive to sample size, this measure provides additional confidence in null effects. Prior to running our analyses, participants’ recall responses were first scored in *R* using *lrd*, a package which allows for automated scoring of recall data while accounting for potential spelling and grammatical errors (Maxwell, Huff, & Buchanan, 2022). Scoring criteria was based on Maxwell et al.’s (2022) guidelines for processing cued-recall responses, such that participant responses were allowed to vary by one character before being counted as incorrect. Figure X plots mean cued-recall rates for participants in the JOL and no-JOL groups. For completeness, all comparisons are reported in the Appendix (Table AX).

To test for differences in cued-recall rates between the JOL and no-JOL groups, the cued-recall data was analyzed using a 2 (Encoding Group: JOL vs. No-JOL) × 3 (Pair Type: Forward vs. Mediated vs. Unrelated) mixed measures ANOVA. Overall, this analysis yielded a significant [MAIN EFFECT ENCODING], such that [PATTERN] Next, [MAIN EFFECT PAIR TYPE], as [PATTERN]. Importantly, [INTERACTION]. Post-hoc testing indicated that [PATTERN]

**Discussion**

[WORDS HERE]

**Experiment 2: Recognition Testing**

[WORDS HERE]

**Method**

**Participants**

[WORDS HERE]

Materials and Procedure

[WORDS HERE]

**Results**

[WORDS HERE]

**Discussion**

[WORDS HERE]

**General Discussion**

[WORDS HERE]

**Conclusion**

[WORDS HERE]

**References**

[FIRST REF HERE]

[FIGURE 1]

[FIGURE 2]

**Appendix**

Table A1

*Summary Statistics for Cue and Target Concreteness, Length, and Frequency as a function of pair type*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pair Type | Position | Variable | *M* | *SD* |
| Forward | Cue | Concreteness | 5.24 | 0.92 |
|  |  | Length | 5.20 | 1.44 |
|  |  | Frequency | 2.51 | 0.59 |
|  | Target | Concreteness | 5.44 | 0.95 |
|  |  | Length | 5.03 | 1.28 |
|  |  | Frequency | 3.53 | 0.63 |
| Mediated | Cue | Concreteness | 5.82 | 0.87 |
|  |  | Length | 4.97 | 1.30 |
|  |  | Frequency | 3.35 | 0.54 |
|  | Target | Concreteness | 5.52 | 0.95 |
|  |  | Length | 5.03 | 1.10 |
|  |  | Frequency | 3.13 | 0.58 |
| Unrelated | Cue | Concreteness | 4.97 | 1.24 |
|  |  | Length | 5.10 | 1.56 |
|  |  | Frequency | 3.22 | 0.82 |
|  | Target | Concreteness | 5.16 | 1.00 |
|  |  | Length | 5.17 | 1.58 |
|  |  | Frequency | 3.05 | 0.78 |

*Note*: Frequency ratings were derived from SUBLTEX (Brysbaert & New, 2009). Concreteness ratings were derived from the English Lexicon Project (Balota et al., 2007).

Table A2

*Associative Strength Summary Statistics for Forward Associates in each Study List*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| List | *M* | *SD* | *Min.* | *Max.* |
| List 1 | 0.445 | 0.234 | 0.141 | 0.808 |
| List 2 | 0.448 | 0.211 | 0.101 | 0.808 |

*Note:* Cells reflect FAS (forward associative strength) values derived from the University of South Florida Free Association Norms (Nelson et al., 2004).

Table A3

*Comparisons of Mean Recall Percentages for each Encoding Group as a function of Pair Type in Experiments 1 and 2*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Experiment | Encoding Task | Pair Type | *M* | *± 95% CI* | F | M |
| Exp. 1 | JOL | Forward |  |  |  |  |
|  |  | Mediated |  |  |  |  |
|  |  | Unrelated |  |  |  |  |
|  | No-JOL | Forward |  |  |  |  |
|  |  | Mediated |  |  |  |  |
|  |  | Unrelated |  |  |  |  |
| Exp. 2 | JOL | Forward |  |  |  |  |
|  |  | Mediated |  |  |  |  |
|  |  | Unrelated |  |  |  |  |
|  | No-JOL | Forward |  |  |  |  |
|  |  | Mediated |  |  |  |  |
|  |  | Unrelated |  |  |  |  |

Note: The three left-most columns denote Cohen’s d effect sizes for post-hoc comparisons. \* = *p* < .05