Judgment of Learning Reactivity Likely Reflects Relational Encoding: Evidence from Mediated Paired Associates and Repetition Priming

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 (https://osf.io/mfbnz/).

Abstract

[ABSTRACT HERE]

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

Judgment of Learning Reactivity Likely Reflects Relational Encoding: Evidence from Mediated Paired Associates and Repetition Priming

Understanding how individuals assess their own memorial abilities is critical for understanding memory. Termed metamemory, these processes influence several aspects of learning, such as how individuals decide which items have been sufficiently learned versus which items require further study (see XXX). As a result, countless studies have investigated whether participants can successfully modify their encoding strategies based on task-demands to maximize later retention (see Nelson & Narens, 1990). Often, studies have investigated metamemory processes using Judgments of Learning (JOLs), which present participants with a set of study items (often cue-target paired associates) and instruct them to rate their ability to correctly recall each item on a later test. When using cue-target word pairs, JOLs are often framed as the probability of later recalling the target if prompted by the cue. While participants can provide JOLs using a variety of scales (see Hanczakowski, Zawadzka, Pasek, & Higham, 2013 for review), these ratings are often elicited via a continuous 0-100 scale representing the probability of correctly recalling the pair’s target item at test. Thus, JOLs provide a convenient tool for assessing metamemory accuracy, as scale JOLs allow researchers to assess the correspondence between predicted and actual recall through a simple comparison process (e.g., Koriat & Bjork, 2005; Maxwell & Huff, 2021; Rhodes & Castel, 2008; see Rhodes, 2016).

Early studies often regarded JOLs as neutral measures which had no influence on later memory. However, a growing body of evidence suggests that JOLs are instead *reactive* on learning (see Double, Birney, & Walker, 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 encoding alters participants’ memory for studied materials, likely by directing attention to aspects of the stimuli that participants might not otherwise have closely attended to. As a result, providing JOLs produces changes in memory, which can potentially manifest as 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 is 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 commonly omitted this additional comparison group, either because JOLs were assumed to be neutral on memory or because such studies were primarily concerned with factors influencing JOL accuracy (e.g., associative direction, Koriat & Bjork, 2005; Maxwell & Huff, 2021; Font size, Rhodes & Castel, 2008; JOL timing, Dunlosky & Nelson, 1994; Nelson & Dunlosky, 1991; 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 mixed lists of related and unrelated cue-target word pairs. These studies have consistently found that JOLs produce positive reactivity but only when pairs are related (e.g., cat – dog). For unrelated pairs (e.g., cat – sky), no reactivity is typically observed, with cued-recall rates being equivalent between JOL and control group participants for this pair type (e.g., Janes, Rivers, & Dunlosky, 2018; Maxwell & Huff, 2022; Soderstrom, Clark, Halamish, & Bjork, 2015; etc.). However, in a 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. Subsequent studies, however, have been unable to reproduce this pattern, and furthermore, a meta-analysis conducted by Double, Birney, and Walker (2018), which analyzed results from 17 JOL studies, found strong evidence of positive reactivity on related pairs but no evidence of negative reactivity on unrelated pairs. Thus, it is evident that JOLs are reactive on cue-target pairs, but, importantly, this effect is moderated by pair relatedness.

To explain this pattern, Soderstrom et al. (2015) proposed a cue-strengthening account of reactivity that is based on Koriat’s (1997) cue-utilization framework. According to Soderstrom et al.’s account, JOLs reactivity is contingent on two conditions being met. First, the act of providing JOLs must direct participants’ attention towards specific aspects of the study pairs which might otherwise be overlooked. 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 relatedness 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 which are JOLs strengthen at encoding must additionally be easily accessible at test. Related cue-target pairs, therefore, demonstrate positive reactivity when cued-recall testing is used, as this test type naturally prompts recall of the missing target. Thus, related word pairs demonstrate a positive reactivity effect, as naturally occurring relatedness cues are both strengthened at encoding and utilized at test. Unrelated pairs, however, show no reactivity given their lack of intrinsic relatedness cues.

Soderstrom et al.’s (2015) cue-strengthening account aligns with the general pattern of reactivity 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; 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. For example, 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 testing, in which no cues are provided, reactivity failed to occur, regardless of whether pairs were related. These findings were replicated by Chang and Brainard (2023; 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 a combination of cue-utilization at encoding and cue accessibility at retrieval.

**Cue-Strengthening and Relational Encoding**

While previous research has directly investigated the relationship between cue-strengthening and reactivity by manipulating the type of test participants that complete at retrieval, few studies have assessed the types of cues which JOLs are purported to strengthen. Instead, previous has operated under the assumption that JOLs modify memory by primarily calling attention to relatedness cues, rather than other intrinsic cues 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 relatedness cues is likely because studies investigating JOL reactivity commonly use cue-target pairs at encoding. While cue-target pairs can contain several intrinsic cues, pair relatedness is typically the most salient, especially when pairs are strong associates.

Recently, Maxwell and Huff (2022) investigated the role of relational encoding processes on reactivity by comparing JOLs with 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. Furthermore, both ratings were moderately-to-strongly correlated with JOLs on related and unrelated pairs (*r*s ≥ .70 and .41, respectively), providing further evidence that participants consider relatedness when providing JOLs. As a result, the authors concluded that JOL reactivity likely reflects participants’ use of a relational encoding strategy, which is triggered whenever participants are asked to judge aspects of cue-target pairs pertaining to relatedness. However, because related pairs have stronger relatedness cues, they more readily benefit from this process. Thus, only related pairs benefit from the requirement to provide JOLs at encoding.

To test the selective nature of this account, Maxwell and Huff (2022) included an additional experiment comparing recall for participants making JOLs to a separate group of participants who either completed an explicit relational encoding task or a shallow vowel-counting task. Unlike JOLs in which relational processes are assumed to be moderated by pair type, participants in the relational encoding group were explicitly instructed to relate all paired items together, regardless of whether or not they were related. Overall, participants in the JOL group replicated previously reported reactivity patterns, such that related but not unrelated pairs received a memorial benefit. The relational encoding task similarly improved recall of related pairs, though unrelated pairs also showed a recall improvement, given that this task was applied globally to all pair types. For participants in the vowel-counting group, however, no memorial benefits were observed, regardless of pair type. Thus, the authors concluded that JOL reactivity patterns likely reflect the selective use of a relational encoding strategy.

Similarly, Halamish and Undorf (in press) assessed the influence of pair relatedness on JOL reactivity by having participants by comparing reactivity for related, unrelated, and identical cue-target pairs. Importantly, the authors also had participants complete a relatedness judgment task at retrieval in which participants indicated at test whether the cue item had presented alongside a related, unrelated, or identical target. Consistent with previous findings, participants making JOLs demonstrated positive reactivity on related pairs but no reactivity on related pairs. Furthermore, Halamish and Undorf found that positive reactivity patterns additionally extended to identical cue-target pairs, providing further evidence that cue-target relatedness is a requisite for JOL reactivity on cue-target pairs. Finally, making JOLs improved the accuracy of participants’ relatedness judgments, but only for relatedness pairs. Unrelated pairs and identical pairs, however, showed no accuracy improvement. Taken together, findings from both Maxwell and Huff (2022) and Halamish and Undorf provide converging evidence that cue-strengthening primarily operates on intrinsic relatedness cues.

**The Present Study**

While it is evident JOL reactivity on cue-target pairs is contingent upon relatedness, it is less clear whether reactivity is purely limited to situations in which relatedness cues are readily apparent at encoding (i.e., forward associates like *mouse* – *cheese*). If JOL reactivity instead reflects relation encoding, reactivity would still be expected to occur on items that are associatively related yet appear unrelated at encoding. To test this, Experiments 1A and 1B compared reactivity on forward associates and unrelated pairs with mediated associates (e.g., *lion* – *stripes*). Unlike traditional forward associates, mediated pairs are not directly related via traditional measures of word association (e.g., forward association strength; FAS; Nelson, McEvoy, & Schreiber, 2004). Instead, their relatedness reflects an indirect relationship between concepts, such that paired items are linked by a related but not presented item (e.g., *lion* – *stripes* is mediated through *tiger*; see Huff and Hutchison, 2011). Based on a spreading activation account, the presentation of mediated pairs activates the non-presented mediator item, which provides a link between otherwise disparate items (Balota & Lorch, 1986; Jones 2010). Thus, if making JOLs at encoding strengthens pre-existing relations between cue-target pairs, positive reactivity would be expected on mediated pairs. However, if JOL reactivity is reliant on intrinsic relatedness cues being readily available at encoding, no reactivity would be expected to occur. Thus, by comparing between forward and mediated pairs, Experiments 1A and 1B provided a stronger test of the cue-strengthening account while also directly testing the relational account of JOL reactivity.

**Experiment 1A: Mediated Associates and Cued-Recall Testing**

The goal of Experiment 1A was to test the relational and cue-strengthening accounts of JOL reactivity. In doing so, we compared cued-recall performance between JOL and no-JOL groups on forward and mediated paired associates and unrelated pairs. Based on previous reactivity studies, we expected any JOL reactivity effects would be moderated by pair type. Specifically, making JOLs should produce positive reactivity, but only on related pairs. For unrelated pairs, no reactivity was expected. Regarding mediated pairs, the relational and cue-strengthening accounts lead to diverging predictions. The relational account of reactivity predicts JOLs would be reactive on mediated pairs, as the making JOLs should strengthen the pre-existing links between mediated pairs, leading to enhanced memory for this pair type versus a control group. However, because the links between mediated pairs are weaker relative to forward pairs, it is likely that any observed reactivity effects on mediated pairs would be reduced compared to forward associates. The cue-strengthening account, however, predicts no reactivity on mediated pairs, as intrinsic relatedness cues are likely not available at encoding. Thus, any positive reactivity on mediated pairs would be taken as evidence in favor of a relational account of reactivity.

**Method**

**Participants**

Participant recruitment was informed by an a priori power analysis conducted using *G\*Power 3.1* (Faul, Erdfelder, Lang, & Buchner, 2007), which indicated that 86 participants would be required to detect large main effects/interactions (Cohen’s *d* = 0.50). However, because we conducted Experiment 1 online, we extended participant recruitment due to an anticipated increase in response variability. Participant recruitment occurred simultaneously via two platforms. Eighty-four undergraduate psychology students were recruited from the University of Southern Mississippi and completed the study in exchange for partial course credit. An additional 48 participants completed Experiment 1 via Prolific (www.prolific.co) and were compensated at a rate of $4.50/half-hour. To be eligible for participation, participants were required to be native English speakers, and Prolific participants were additionally required to have obtained at least a high school degree or equivalent. For both recruitment sources, particpants were randomly assigned to either the JOL or no-JOL encoding groups. Data from 12 participants were omitted due to low recall rates (i.e., recall < 5%, which suggested participants were distracted at encoding), recall rates exceeding 95% for all pair types (which suggested that participants were cheating on the cued-recall test, or JOL ratings that consistently anchored on scale extremes, which suggested that participants were not following encoding directions. Our final sample contained 120 participants (JOL *n* = 60; no-JOL *n* = 60).

**Materials**

To create the stimuli pairs, 60 cue-target word pairs were generated using the University of South Florida Free Association norms. These pairs included 30 forward associates (e.g., mouse – cheese; Nelson, McEvoy, & Schreiber, 2004) and 30 unrelated pairs (e.g., muffin – floor). 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 (i.e., forward, unrelated, and mediated). This resulted in 45 cue-target pairs per 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 (see Appendix Table A1 for lexical properties for all pair types). Additionally, forward pairs within each list were matched on FAS (see Table A2). 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. Thus, each list contained a total of 55 pairs. The final set of stimuli pairs HAS been made available at https://osf.io/mfbnz/. 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. For all significant analyses of variance (ANOVAs) and t-tests, we report partial eta squared (*ηp*2) and Cohen’s *d* effect size measures, respectively. 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 supplemental analysis compares a model in which a significant effect is assumed to a secondary model assuming a null effect and returns an estimated probability 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). This scoring process followed Maxwell et al.’s (2022) guidelines for processing cued-recall data, such that participant responses were allowed to vary by one character before being counted as incorrect.

Figure 1 (top panel) plots mean cued-recall rates for participants in the JOL and no-JOL groups. For completeness, all comparisons are reported in the Appendix (Table A3). To test for JOL reactivity effects, 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 effect of Encoding Group, *F*(1, 118) = 14.20, *MSE* = 731.18, *η*p2 = .11, such that mean cued-recall rates were higher for participants in the JOL group versus the no-JOL group (45.29 vs 34.56, respectively). Additionally, this analysis yielded a significant effect of pair type, *F*(2, 236) = 778.00, *MSE* = 111.03, *η*p2 = .87. Collapsed across encoding groups, mean recall was highest for forward pairs (69.47), followed by mediated pairs (33.22), and unrelated pairs (17.08). Post-hoc testing confirmed that all comparisons between all pair types differed significantly, *t*s ≥ 6.75, *d*s ≥ 0.74. Importantly, a significant Encoding Group × Pair Type interaction was detected, *F*(2, 236) = 27.07, *MSE* = 111.03, *η*p2 = .19. Starting with forward pairs, a robust reactivity effect was detected, such that recall rates for participants making JOLs greatly exceeded participants in the no-JOL group (79.94 vs. 59.00; *t*(118) = 6.73, *SEM* = 3.16, *d* = 1.23). Critically, this positive reactivity pattern extended to mediated pairs, as making JOLs similarly facilitated recall of this pair type (38.55 vs. 27.89; *t*(118) = 2.82, *SEM* = 3.82, *d* = 0.51). However, no reactivity was observed on unrelated pairs, as cued-recall rates did not differ between participants in the JOL and no-JOL groups (17.39 vs. 16.67; *t*(118) < 1, *SEM* = 2.81, *p* = .83, *p*BIC = .88). Thus, the requirement to provide JOLs at encoding benefited cued-recall performance, but only when pairs contained a pre-existing relationship.

**Experiment 1B: Mediated Associates and Recognition Testing**

Experiment 1B closely followed the previous experiment with the exception that participants were tested via recognition rather than cued-recall. We selected this test type given previous studies have shown that JOL reactivity effects on cued-recall testing similarly occur on this test type (see Myers et al., 2020). Thus, our use of recognition testing in Experiment 1B provided a further test of JOL reactivity effects on recognition memory while additionally testing whether reactivity effects observed on mediated pairs in the previous experiment would replicate. Given that Myers et al. reported that JOL reactivity patterns for cued-recall testing extend to recognition, testing we expected that forward pairs would again demonstrate a positive reactivity pattern, such that providing JOLs at encoding would improve correct recognition relative to the no-JOL group. Additionally, based on the previous experiment, we anticipated that any positive reactivity patterns observed on forward pairs extend to mediated pairs, though again, this effect was expected to be smaller for mediated pairs versus forward pairs. Finally, we expected that no reactivity would occur on unrelated pairs.

**Method**

**Participants**

An additional XX participants completed Experiment 1B. All participants were undergraduate students recruited from the University of Southern Mississippi (*n* = xx) or Midwestern State University (*n* = xx) who completed the study in exchange for partial course credit. Consistent with Experiment 1A, participants were randomly assigned to the JOL encoding group or the no-JOL control group. Data screening followed the same procedure outlined in Experiment 1A, and data from xx participants were excluded from the final analyses. Thus, our final sample contained XX participants (*n* JOL = XX, *n* no-JOL = XX).

**Materials and Procedure**

Experiment 1B used the same materials and followed the same general procedure as Experiment 1A, with the following exceptions. First, [SOMETHING ABOUT CONTROL ITEMS AND COUNTER BALANCES] [RECOGNITION TEST] [NOTHING ELSE DIFFERED]

**Results**

Figure 1 (bottom panel) plots mean correct recognition as a function of encoding group. For completeness, all comparisons are reported in Table A3. [INTRODUCE THE ANOVA]

**Discussion**

[WORDS HERE] [RECAP]

[ANOTHER PARAGRAPH HERE SETTING UP THE PRIMING STUFF?]

**Experiment 2: Repetition Priming**

[WORDS HERE]

**Method**

**Participants**

[WORDS HERE]

**Materials**

[WORDS HERE]

**Procedure**

[WORDS HERE]

**General Discussion**

[WORDS HERE]

Our inclusion of mediated associates provides a novel comparison, as to date, studies investigating reactivity effects have primarily compared forward associates and unrelated pairs (though see Maxwell & Huff, 2022; in press; and Mitchum et al., 2016, who included backward associates), and no study has assessed whether reactivity effects on related pairs extends to mediated paired associates. Additionally, by including mediated pairs, the present study provides a situation in which cue-target pairs are related yet perceived as unrelated at encoding. In doing so, the present study directly tests the cue-strengthening account’s requirement that JOLs strengthen intrinsic relatedness cues at encoding.

**Conclusion**

[WORDS HERE]

**References**

Balota, D. A. & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(3), 336–345.

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*, 977–990.

Chang, M. & Brainerd, C. J. (2023). Changed-goal or cue-strengthening? Examining the reactivity of judgments of learning with the dual-retrieval model. *Metacognition and Learning 18*, 183–217.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgments of learning. *Memory, 26*(6), 741–750.

Dunlosky, J. & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1180–1191

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*(4), 545-565.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev). Bradford Books/MIT Press.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Garcia, M. & Kornell, N. (2015). Collector [Computer software]. Retrieved April 3rd, 2020 from https://github.com/gikeymarica/Collector

Halamish, V. & Undorf, M. (in press). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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*, 429–444.

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review, 25*(6), 2356–2364.

Jones, L. L. (2010). Pure mediated priming: A retrospective semantic matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 135–146.

Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experiment Psychology: General, 126*(4), 349–370.

Koriat, A. & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187–194.

Maki, W. S. (2007). Judgments of associative memory. *Cognitive Psychology, 54*(4), 319–353.

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*, 679–690.

Maxwell, N. P. & Buchanan, E. M. (2020). Investigating the interaction of direct and indirect relation on memory judgments and retrieval. *Cognitive Processing, 21*, 41-53.

Maxwell, N. P. & Huff, M. J. (2021). The deceptive nature of associative word pairs: The effects of associative direction on judgments of learning. *Psychological Research, 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.

Maxwell, N. P. & Huff, M. J. (In Press). Is discriminability a requirement for reactivity? Comparing the effects of mixed vs. pure list presentations on judgment of learning reactivity*. Memory & Cognition*, 1-16.

Maxwell, N. P., Huff, M. J., & Buchanan, E. M. (2022). The *lrd* package: An *R* package and Shiny application for processing lexical data. *Behavior Research Methods, 54*, 2001-2024.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145*(2), 200–219.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407.

Nelson, T. O. & Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “Delayed-JOL Effect.” *Psychological Science, 2*(4), 267-270.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In: *The psychology of learning and motivation*, ed. G. Bower. American Psychologist.

Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 65–80). Oxford University Press.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 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.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 553–558.

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*, 779–804.

[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 1A and 1B*

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