Title Here: SOMETHING ABOUT DRM AND RELATIONAL ENCODING

Nicholas P. Maxwell

Midwestern State University

**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 WILL GO HERE]

Word Count: XXX

*Keywords*: Judgments of Learning; Reactivity; Single Item Lists; DRM Paradigm; Recognition

[TITLE HERE]

When learning new information, it is helpful to assess whether the knowledge one is acquiring is being sufficiently encoded. Metamemory, or the processes by which individuals regulate their learning, is a crucial aspect of learning, as these processes help individuals decide whether items need additional study or if they have been sufficiently learned (see Nelson & Narens, 1990). To investigate questions surrounding metamemory, researchers commonly use Judgments of Learning (JOL) tasks, where participants study a set of items and predict their likelihood of correctly remembering them on a later memory test (see Rhodes, 2016; Schwartz & Metcalf, 2017, for reviews). Thus, JOLs provide researchers with a simple measure for assessing how various manipulations affect the study process, including perceptual manipulations (e.g., font; Rhodes & Castel, 2008; Maxwell, Perry, & Huff, 2022) and the presence of associations between items in cue-target pairs (e.g., Koriat & Bjork, 2005; Castel, McCabe, & Rhodes, 2007).

Historically, research involving JOLs has considered these ratings to be neutral measures with no effect on memory, particularly when participants make them concurrently with or immediately following study (though see Spellman & Bjork, 1992, who posited that JOLs made following a delay improved memory for judged items). As such, early JOL studies often focused on factors influencing their accuracy, rather than the potential effects of these judgments on memory. However, over the last decade, a growing body of research has demonstrated that JOLs are *reactive* on learning (e.g., Janes, Rivers, & Dunlosky, 2018; Maxwell & Huff, 2022; Soderstrom, Clark, Halamish, & Bjork, 2015; see Double, Birney, & Walker, 2018, for review). Based on these accounts, the act of making JOLs modifies participants memory for studied items, likely by making certain features of the stimuli more salient at encoding (Ericsson & Simon, 1993). Thus, providing JOLs at encoding could potentially produce memorial benefits (i.e., *positive reactivity*) or costs (i.e., *negative reactivity*), which would be evident when comparing memory performance between participants making JOLs at encoding and a separate group of participants completing a no-JOL control task (e.g., silent reading).

Studies investigating the mechanisms behind JOL reactivity have commonly used cue-target word pairs. These studies have revealed a consistent pattern of reactivity: When pairs are semantically related (e.g., mouse – cheese), making JOLs improves memory for the target item (Janes et al., 2018; Halamish & Undorf, 2023; Maxwell & Huff, 2022; Soderstrom et al., 2015). However, this memorial benefit does not extend to unrelated pairs (e.g., mouse – cup), particularly when memory is assessed via cued-recall testing (but see Mitchum, Kelley, & Fox, 2016, who showed no reactivity on related pairs and *negative* reactivity on unrelated pairs). To explain why relatedness moderates JOL reactivity patterns, Soderstrom et al. proposed that two conditions must be met for reactivity to occur on cue-target pairs. First, studied items must contain intrinsic cues which participants use to inform their JOLs (see Koriat, 1997). Second, participants must be tested using a method that is sensitive to any cues that are strengthened due to making JOLs. Based on this *cue-strengthening account*, when participants are tested via cued-recall, the cue-strengthening account predicts a memory benefit on related pairs, given that cued-recall testing is highly sensitive to pair relatedness. For unrelated pairs, however, JOLs do not produce a memorial benefit, as intrinsic cue-target relations are not available for this pair type.

While the cue-strengthening account makes no specific claims regarding the particular cues that JOLs strengthen, it is likely that JOLs primarily strengthen pre-existing cue-target relations, as pair relatedness is a highly salient cue for later remembering (Mueller, Tauber, & Dunlosky, 2013). Given this possibility, recent studies have explored the degree to which relatedness contributes to positive reactivity. For example, Maxwell and Huff (2022) compared recall for participants making JOLs to three additional encoding groups: A no-JOL control group, a group of participants completing a shallow vowel-counting task instead of providing JOLs, and, importantly, a group of participants who engaged in a deep relational encoding task. Compared to both the no-JOL and shallow encoding groups, JOL participants showed positive reactivity for all related pair types. Critically, participants in the relational encoding group also demonstrated this pattern, even though they were not required to provide JOLs at encoding. Thus, Maxwell and Huff concluded that providing JOLs at encoding encouraged participants process cue-target relations, which led to the implicit adoption of a relational encoding strategy.

Additionally, Halamish and Undorf (2023) recently investigated relatedness effects on JOL reactivity using related, unrelated, and identical cue-target pairs. In addition to testing differences in cued-recall across pair types, participants also made relatedness judgments at test, in which participants indicated whether cue items had been previously paired with a related, unrelated, or identical target word. Overall, JOLs produced positive reactivity on related but not unrelated pairs, a finding consistent with previous reactivity studies (e.g., Janes et al., 2018; Soderstrom et al., 2015, etc.). Positive reactivity also extended to identical cue-target pairs, further suggesting that perceived relatedness is necessary for reactivity to occur. Importantly, making JOLs also improved the accuracy of relatedness judgments, but only on related pairs. Taken together, these findings suggest that JOLs encourage participants to process relatedness but only for pairs with pre-existing relations.

**JOL Reactivity and List Relatedness**

As noted above, much work investigating JOL reactivity has tested for potential memory changes using various types of cue-target word pairs. This is likely because factors influencing JOLs are often studied within the context of related versus unrelated cue-target pairs, rather than via list-wise manipulations of relatedness (see Chang & Brainard, in press). However, like cue-target relatedness, list relatedness has similarly been shown to affect the magnitude of JOLs. For example, Matvey, Dunlosky, & Schwartz (2006) had participants make item-level JOLs for words presented in either categorized (i.e., related) or uncategorized (i.e., unrelated) single word lists. Overall, a classic relatedness effect emerged, as JOLs were higher for categorized lists versus uncategorized lists. More recently, Chang and Brainard replicated this general pattern for JOLs while also demonstrating that items in categorized lists were consistently recalled at a greater rate relative to uncategorized lists. Thus, relatedness effects that are observed on JOLs with cue-target pairs readily extend to categorized and uncategorized single word lists.

While relatedness has been shown to affect the magnitude of JOLs regardless of whether it is manipulated inter-item (i.e., related and unrelated cue-target pairs) or list-wise (i.e., categorized and uncategorized word lists), the mechanisms driving these effects likely differ based on the type of stimuli. According to Koriat’s (1997) cue-utilization account, making JOLs for cue-target pairs promotes the use of intrinsic cues (i.e., inherent item properties) which participants use to inform the magnitude of their JOLs. Pre-existing cue-target relations within word pairs reflect a primary example of intrinsic cues, as the relation between items is inherent to the pair and easily processed at encoding. However, by nature, single-word lists display study items to participants one item at a time. Thus, any relatedness cues would reflect *extrinsic cues*, as participants must assess how the currently presented item relates back to previously presented items within the same list. Thus, extrinsic cues reflect the specific characteristics in which encoding occurs, rather than inherent properties of the stimuli.

Regarding JOL reactivity, few studies have directly assessed the effects of making JOLs on memory for single-item lists. As a result, current theories of JOL reactivity primarily focus on explaining the role of intrinsic cues as a driving factor of reactivity on cue-target pairs (e.g., Soderstrom et al.’s, 2015 cue-strengthening account). Thus, it remains unclear whether the relatedness effects underlying JOL reactivity on cue-target pairs would similarly extend to categorized word lists. In a recent exception, however, Senkova and Otani (2021) assessed the effect of list-wise relatedness on JOL reactivity by testing for reactivity on categorized and uncategorized word lists for participants making JOLs, pleasantness ratings, or a control task in which participants simply assigned a random number to each item (Experiment 1) and participants making JOLs, completing an imagery task, or the control task (Experiment 2). Across experiments, participants making JOLs had greater free-recall relative to participants in the control group. Importantly, recall benefits were greater for categorized lists, suggesting that the presence of list-wise relatedness facilitated reactivity. Additionally, both deep encoding comparison groups also improved free-recall relative to participants in the no-JOL control group. Because both pleasantness ratings and imagery tasks are classic item-specific tasks based on the item-specific/relational framework (Einstein & Hunt, 1980; Hunt & Einstein, 1981), Senkova and Otani argued that positive JOL reactivity reported on categorized lists reflected an item-specific process rather than a relational encoding process. Considered alongside reactivity findings with cue-target pairs which have suggested the role of cue-target relations (e.g., Halamish & Undorf, 2023; Maxwell & Huff, 2022), it may be the case that JOL reactivity reflects different underlying processes based on the type of stimuli that participants study.

**The Present Study**

Although it is evident that making JOLs benefits recall of related versus unrelated cue-target pairs, less is known about the effects of these judgments on memory for categorized and uncategorized lists. As such, the present study sought to first replicate findings from Senkova and Otani (2021) demonstrating that JOLs produce a greater memorial benefit on categorized versus uncategorized lists using free-recall testing (Experiment 1A) and whether this pattern extends to recognition testing (Experiment 1B). Next, Experiment 2 used the Deese-Roediger-McDermott paradigm (Deese; 1959; Roediger & McDermott, 1995), which allowed for an assessment of the effects of item-specific and relational processes on JOL reactivity. To preview, across experiments [RELATED VS UNRELATED FINDINGS FOR FREE-RECALL] [RECOGNITION FINDINGS?] Finally, [ITEM-SPECIFIC VS. RELATIONAL FOR DRM]

**Experiment 1A: Related versus Unrelated Lists and Free-Recall Testing**

The goal of Experiment 1A was to replicate findings initially reported by Senkova and Otani (2021) showing that JOLs improve free-recall of categorized but not uncategorized single-word lists. In doing so, free-recall was compared between three groups: Participants who provided concurrent, item-level JOLs, participants who made a global JOL immediately following the presentation of each study list, and a no-JOL control group where participants silently read each item without providing JOLs. Based on previous research, it was expected that making JOLs would benefit recall of categorized but not uncategorized word lists. However, [item-specific vs. relational stuff]

**Method**

**Participants**

XX participant were recruited from XX University and completed the study in exchange for partial course credit. Next, an additional XX participants were recruited from Prolific (www.prolic.co) and participated at a rate of $3.00 per 20-minute session. The final sample contained XX participants and was informed by an a priori power analysis conducted with *G\*Power 3.1* (Faul, Erdfelder, Buchner, & Lang, 2009), which suggested that xx participants would be needed to detect medium main effects and interactions (STATS). All participants were native English speakers, and Prolific participants were additionally required to have obtained at least a high school degree or equivalent.

**Materials**

Ninety-six English words were selected to serve as stimuli. These words were split into eight 12-word lists, which were based on Van Overschelde, Rawson, and Dunlosky’s (2004) categorical word norms. Of these lists, four were categorized, such that each word was an exemplar of a given category. The remaining four lists were uncategorized and, as such, all words within the list were semantically unrelated. These lists were created by randomly selecting words from unused categories in the Van Overschelde et al. norms. Appendix Table X displays lexical properties for all lists.

**Procedure**

Experiment 1A was administered online using Collector, an online platform for conducting browser-based psychology experiments (Garcia & Kornell, 2015). Following informed consent, participants were told that they would be viewing a series of words and that their memory for each word would later be tested. Participants in the two JOL groups then received additional instructions regarding their respective judgments. Specifically, participants in the item-based JOL group were instructed to rate their ability to remember each word, with JOLs elicited concurrent with study, such that JOLs were provided while each word was displayed. Separately, participants in the global JOL group were informed that following the last item in each list, they would be asked to provide a single JOL representing their ability to correctly remember the previously presented set of items on a later test. For both JOL groups, JOLs were framed as the percent likelihood of correctly items at test and were elicited via a continuous 0-100 scale (i.e., 0 = definitely will not remember, 100 = definitely will remember). To mitigate potential anchoring effects, participants were instructed to be as accurate as possible when providing their JOLs and were encouraged to use the full range of the response scale. Participants in the no-JOL group did not receive additional judgment instructions and were instead instructed to read each pair silently

After receiving their respective encoding instructions, participants were presented with the first word list. For all participants, encoding was self-paced, with participants pressing the ENTER key to advance to the next word. After completing the first study phase, participants completed a short filler task in which they were presented with a random consonant and instructed to list as many words which started with this letter as they could generate in a 30 second time span (i.e., list all words starting with the letter “M”). After the time limit had been reached, participants completed a free-recall task in which they were prompted to type as many words from memory as they could correctly recall from the previous list. Following the free-recall task, participants began the second list. This process then repeated until participants had completed all four-study lists. List presentation order was randomized for all participants, and all words were additionally randomized within lists. The full experiment took approximately 20 minutes to complete.

**Results**

For all analyses, significance was set at the *p* < .05 level. For all significant main effects and interaction, partial eta-squared (*ηp*2)is reported, and Cohen’s *d* effect sizes are reported for significant post-hoc comparisons. Additionally, for all non-significant main effects and post-hoc comparisons, a Bayesian estimate of the strength of the evidence supporting the null hypothesis is reported (Masson, 2011; Wagenmakers, 2007). This analysis compares two models (one assuming a significant effect and one assuming a null effect), which allows for the computation of a probability estimate (a *p*-value termed *p*BIC; Bayesian Information Criterion). This provides a probability estimate of the null hypothesis being retained, which is sensitive to sample size, and provides increased confidence in reported null effects

Figure X (top panel) depicts changes in free-recall as functions of encoding group and list construction. To test for reactivity effects, the data was analyzed using a 3(Encoding Group: Item-JOL vs. List-JOL vs. No-JOL) × 2(List Type: Categorized vs. Uncategorized) mixed-measures ANOVA. Overall, [MAIN EFFECT OF ENCODING GROUP?] [PATTERN]. Next, [MAIN EFFECT OF LIST-TYPE] [PATTERN]. Importantly, [INTERACTION] Starting with categorized lists, [PATTERN]. However, for uncategorized lists, [PATTERN]. Thus, [SUMMARY]

**Experiment 1B: Related versus Unrelated Lists and Recognition Testing**

Experiment 1B tested [MAIN GOAL – FREE RECALL]. Based on [WHAT DO WE PREDICT?]

**Method**

**Participants**

An additional XX undergraduate students were recruited from Midwestern State University and completed Experiment 1B in exchange for partial course credit. Participants were randomly assigned to one of the three encoding groups used in Experiment 1A. Initial sample sizes for each group were based on the previous experiment, and the same exclusion criteria were applied. This process removed XX participants from the final dataset, leading to a total of XX participants in the final sample (Item-wise JOLs, *n* = XX; List-wise JOLs, *n* = xx; No-JOL control group, *n* = xx). This sample was based on Experiment 1A, and a sensitivity analysis conducted with *G\*Power 3.1* suggested that this sample had sufficient power to detect [SIZE OF EFFECT] [STATS].

**Materials and Procedure**

Experiment 1B used the same stimuli and materials as Experiment 1A, with the following exceptions. First, the four free-recall tests used in Experiment 1A were replaced with a single, 96-item old/new recognition test. This test included all 48 items from the previously studied lists, as well as the 48 items from the counterbalanced lists, which served as non-presented control items. Second, participants studied all lists back-to-back, rather than completing filler tasks in between lists. All other materials, including the categorized lists and the three sets of encoding instructions (item-JOL, global JOL, and silent reading), were identical to Experiment 1A.

Following the presentation of the final word list, participants completed a 60-second filler task in which they were given a random consonant and were tasked with generating as many words as possible which started with this letter. Immediately afterwards, participants began the recognition test, which presented all 96 items in a randomized order. Specifically, participants were informed that they would be viewing a series of words and were instructed to indicate whether each word had been previously studied (“old”) or had not been previously studied (“new”). Test performance was self-paced; however, participants were instructed to respond as quickly as possible without compromising their accuracy. Like Experiment 1A, Experiment 1B took approximately 20 minutes to complete.

**Results**

Figure X (bottom panel) [REFERENCE FIGURE AND APPENDIX] [INTRODUCE MODEL]

**Signal Detection**

[WORDS]

**Discussion**

[WORDS HERE]

[TRANSITION TO DRM]

**Experiment 2A: DRM Lists and Free-Recall Testing**

[WORDS HERE] [PROBABLY PROLIFIC?]

**Method**

**Participants**

[WORDS HERE] [PROBABLY PROLIFIC]

**Materials**

[WORDS HERE]

**Procedure**

[WORDS HERE]

**Results**

[WORDS HERE]

**Experiment 2B: DRM Lists and Recognition Testing**

[WORDS HERE]

**Method**

**Participants**

[WORDS HERE] [MSU TEXAS]

**Materials and Procedure**

[WORDS HERE] with the exception that the free-recall tests used in Experiment 2A were replaced with a single, 80-item old/new recognition test, which was presented following the final study list. This test consisted of 30 previously studied items (taken from positions 2, 8, and 10 from each list), 10 critical lures from each studied list, 30 non-presented items taken from counterbalanced lists, and 10 critical lure controls which were taken from the non-studied set of lists. Participants studied all 10 lists back-to-back, with the order of list presentation randomized across participants. All JOL instructions were identical to the previous experiments, and depending on their encoding group, participants either made JOLs concurrent with study, made list-wise JOLs following the completion of each list, or read each word silently. Like the previous experiments, all encoding was self-paced.

**Results**

[WORDS HERE]

**Discussion**

[WORDS HERE]

**General Discussion**

[WORDS HERE]

**Conclusion**

[WORDS HERE]

**Open Practices Statement**

[WORDS HERE]

**Compliance with Ethical Practices**

[WORDS HERE]

**References**

[FIRST REF HERE]

[FIGURE 1]

[FIGURE 2]

**Appendix**

Table AX