Original Article

QJEP

Quarterly Journal of Experimental Psychology
2022, Vol. 75(8) 1571–1582

DOI: 10.1177/17470218211056498
qjep.sagepub.com
SAGE

Drawing individual images benefits recognition accuracy in the Deese–Roediger–McDermott paradigm

Jacob M. Namias, Mark J Huff, Allison Smith and Nicholas P Maxwell

Abstract

We examined the effects of drawing on correct and false recognition within the Deese–Roediger–McDermott (DRM) false memory paradigm. In Experiment 1, we compared drawing of a word's referent using either a standard black pencil or coloured pencils relative to a read-only control group. Relative to reading, drawing in either black or coloured pencil similarly boosted correct recognition and reduced false recognition. Signal-detection analyses indicated that drawing reduced the amount of encoded memory information for critical lures and increased monitoring, indicating that both processes contributed to the false recognition reduction. Experiment 2 compared drawing of individual images of DRM list items relative to drawing integrated images using sets of DRM list items. False recognition was lower for drawing of individual images relative to integrated images—a pattern that reflected a decrease in encoded memory information but not monitoring. Therefore, drawing individual images improves memory accuracy in the DRM paradigm relative to a standard read-control task and an integrated-drawing task, which we argue is due to the recruitment of item-specific processing.

Keywords

Drawing; DRM paradigm; false memory; distinctiveness; item-specific; relational processing

Received: 20 March 2021; revised: 30 September 2021; accepted: 3 October 2021

Memory scholars have long been interested in encoding techniques that can improve the quality of memory output. Approaches commonly target specific tasks that are generally classified as "deep" processing tasks according to the levels-of-processing framework (Craik, 2002; Craik & Lockhart, 1972). Examples include pleasantness ratings (Hunt & Einstein, 1981), generation (Bertsch et al., 2007; Slamecka & Graf, 1978), production (Conway & Gathercole, 1987; MacLeod & Bodner, 2017), survival processing (Nairne et al., 2007), and more recently, drawing an image of a word's referent (Wammes et al., 2016, 2017). Although the benefits of these encoding tasks on memory for studied information are well supported using a variety of study materials (e.g., Fernandes et al., 2018; Ozubko et al., 2012), it is equally important to gauge task effectiveness on memory errors when considering overall memory accuracy. The aim of our current study was to evaluate whether the correct memory benefits of drawing would extend to associative false memory errors and evaluate whether drawing individual images reflects recruitment of distinctive item-specific processing.

Several processes have been proposed to support drawing benefits on memory. For instance, Fernandes et al. (2018) suggested that drawing benefits reflected the integration of three separate encoding-based processes (elaboration, motor action, and pictorial processing) to produce a cohesive and powerful memory trace. In addition to encoding processes, drawing also appears to facilitate recollection-based processes at test. For instance, when using the remember–know–new procedure to estimate recollection and familiarity processes (Yonelinas, 2002), correct recognitions of drawn items are more likely to be accompanied by "remember" responses, an estimate of recollection processes, than familiarity-based "know" responses. Of course, contributions of encoding and test-based recollection processes are not

School of Psychology, The University of Southern Mississippi, Hattiesburg, MS, USA

Corresponding author:
Mark J Huff, School of Psychology, The University of Southern Mississippi, Hattiesburg, MS 39406, USA.
Email: mark.huff@usm.edu

mutually exclusive, as participants likely recollect source information that is generated by the elaborative, motoric, and pictorial processes deployed at encoding.

While the encoding and recollective processes provide a reasonable account of why drawing benefits memory for studied items, it is less clear how drawing processes affect memory errors which can negatively affect overall memory accuracy. A common method for evaluating the effectiveness of encoding tasks on both correct memory and memory errors is through the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). Participants in this paradigm are presented with lists of associates (e.g., butter, dough, oven, etc.) that converge on a single critical lure (e.g., bread). Under standard intentional encoding instructions (i.e., reading or listening), the *DRM illusion* approximates 40% to 50% in free recall (Huff & Bodner, 2019; Roediger & McDermott, 1995) and false recognition approximates hit rates (see Gallo, 2006, 2010; Huff et al., 2015, for reviews). Furthermore, the DRM illusion is persistent, appearing weeks to months following initial study (Seamon et al., 2002), and only shows a modest reduction following warnings that are presented prior to study (Gallo et al., 1997; McCabe & Smith, 2002).

One tact that reliably produces benefits to overall memory accuracy in the DRM paradigm is the use of distinctive encoding tasks. Creating mental images of individual list items (Foley et al., 2006), studying pictures of a word's referent (Schacter et al., 1999), studying words in unique fonts (Arndt & Reder, 2003), rating list words based on pleasantness (Huff & Bodner, 2013, 2019), generating list words from anagrams (McCabe & Smith, 2006), and silently thinking about unique characteristics of list items (Huff & Bodner, 2013) have all reduced the DRM illusion relative to a standard read-only control task. Unlike explicit warnings, many of these encoding tasks can benefit memory accuracy by reducing the DRM illusion and increasing correct memory for studied list items—a pattern termed a *mirror effect* (Glanzer & Adams, 1990). A common thread between these distinctive tasks is that they facilitate processing of item-specific information (Hunt & Einstein, 1981) which mechanistically could either disrupt the thematic consistency of the list (e.g., Fuzzy-trace theory; Brainerd & Reyna, 2002) or disrupt the spread of implicit spreading activation (Roediger, Balota, & Watson, 2001), both of which may contribute to the formation of the DRM illusion (see Huff et al., 2021, for review).

Distinctive encoding benefits have been chalked up to processes that operate during encoding of list items or increased monitoring at test. For instance, the *impoverished relational encoding* account (Hege & Dodson, 2004; Hockley & Cristi, 1996) posits that distinctive tasks interfere with the encoding of thematic/associative information. Separately, distinctive encoding may also encourage participants to deploy a global memory monitoring strategy at test termed the *distinctiveness heuristic* (Schacter et al., 1999). According to this strategy, participants will only report/endorse information from memory that includes recollection of distinctive details that originated from the encoding task. Because critical lures do not possess distinctive details as they were never studied, the absence of such details can be diagnostic that the lure was not studied, leading to its rejection at test (Gallo, 2004).

Importantly, researchers have found that both impoverished relational encoding and the distinctiveness heuristic are not mutually exclusive. Huff and Bodner (2013) used a signal-detection approach to derive separate estimates of encoding and monitoring processes on correct and false recognition in three experiments using a between-group design. Across experiments, two deep processing variants of an encoding task were compared with a read-only control task. One task variant emphasised the processing of item-specific features of individual list items. The other variant emphasised the processing of relational or shared features of the study list items. Item-specific and relational variants were compared using different tasks, including processing instructions, pleasantness ratings, and generation. Across task types, both item-specific and relational variants increased correct recognition relative to a read-control group, but only the distinctive item-specific variants reduced false recognition relative to the read group. A signal-detection analysis was then applied, which allowed for the separation of memory experiences for studied versus non-studied information (discriminability, d') from test-based monitoring, or the likelihood that non-studied information is correctly rejected. Discriminability serves as an estimate for the amount of information encoded for studied items and critical lures (i.e., impoverished relational encoding). Separately, correct rejection rates serve as a quantitative estimate of the amount of monitoring at test (computed as lambda, λ), which estimates the use of a distinctiveness heuristic. When applied to Huff and Bodner's (2013) recognition data, item-specific variants reduced false recognition relative to the read-control group which was due to a reduction in encoded memory information (lower d') and an increase in monitoring (greater lambda). Thus, both impoverished relational encoding and the distinctiveness heuristic contribute to item-specific reductions in the DRM illusion—a pattern that has been supported in a meta-analysis (Huff et al., 2015) and when applying the drift-diffusion model (Ratcliff, 1978) to estimate encoding and monitoring parameters using response latencies (Huff & Aschenbrenner, 2018).

Although distinctive encoding tasks produce robust and reliable mirror effects in the DRM paradigm, a recent study by Meade et al. (2020) reported that drawing may not always benefit memory accuracy. In their study, participants studied DRM lists using a standard single-image drawing task which was compared within-subjects to lists that were studied through writing the list word (Experiment

1), drawing versus generating a mental image of the word's referent (Experiment 2), or drawing versus a feature listing task in which participants listed the physical characteristics of each word's referent (Experiment 3). On a final recognition test, drawing increased correct recognition relative to both the writing and mental imagery tasks, but drawing did not increase correct recognition relative to feature listing, although recognition on both tasks was equivalent and at ceiling. Importantly, however, when compared to writing and mental imagery tasks, drawing produced an increase in false recognition—a memory accuracy cost. However, compared with the feature listing task, drawing lists showed a reduction in false recognition. Based on these results, the authors concluded that although drawing generally produces correct recognition benefits, it can be accompanied by accuracy costs relative to writing and mental imagery.

While determining the costs and benefits of encoding tasks is informative regarding the utility of a task in practice (e.g., Begg & Snider, 1987; Bodner et al., 2014), the drawing costs that were reported by Meade et al. (2020) may have been due, at least in part, to the context in which the drawing comparisons were made. First, the use of a within-design to compare the drawing task with other tasks may have provided an ambiguous context with which to evaluate the effects of drawing on correct and false recognition. In particular, when encoding subsets of study lists using different encoding tasks, processing from one encoding task may *spillover* to another task and vice versa which may affect correct and false recognition rates in a within-group context. As an example of encoding spillovers in the DRM paradigm, Huff et al. (2021) examined correct and false recognition in a within-group in which half of the DRM lists were encoded using a standard read task and the other half were encoded using a distinctive item-specific generation task. These within-group read and generate lists were then compared with pure groups in which DRM lists were only studied using either a read task or an item-specific generation task. The comparison between read and generate within-group lists relative to their corresponding pure groups allowed for the estimation of spillover effects present in a within-group context. In Huff et al.'s study, a robust pattern of spillovers on false recognition for read and generate within-group lists emerged. Specifically, false recognition for within-group read lists was 16% lower than false recognition in the pure read group, suggesting that item-specific processing from the generation task spilled over to read lists which reduced false recognition relative to the pure read group. Similar spillovers were found on generate lists. False recognition for within-generation lists was 21% greater than false recognition for the pure generate group, indicating that processing within-group read lists, which likely encouraged relational processing due to the strong association across DRM list items (cf. Huff & Bodner, 2014; Hunt & Seta, 1984), spilled over to generate lists, disrupting the generation task's effectiveness at reducing the DRM illusion.[1] When considering Meade et al.'s (2020) results, the within-group drawing lists may have been similarly affected by spillovers. As an example, false recognition rates for drawing lists varied 10% to 15% across experiments despite the drawing task remaining the same. This variability may be due to processing spillovers from the other within-subject tasks which may have affected false recognition rates of the drawing task. Evaluating the costs and benefits of drawing is therefore challenging in a within-subjects context, as one cannot determine whether encoding processes can be ascribed to the drawing task itself or whether processing applied to one task spilled over to the other.

Second, Meade et al. (2020) also compared the effects of drawing individual images to rewriting, individual mental imagery, and feature listing tasks, all of which are "deep" processing tasks that have been shown to affect both correct and false memory in the DRM paradigm relative to a read-only control task. Importantly, tasks, such as rewriting/typing (i.e., production; Bodner et al., 2016) and individual mental imagery, are classified as item-specific tasks, which have been shown to be effective at reducing DRM illusion relative to a standard read-only control task (Dodson & Schacter, 2001; Foley et al., 2006). Thus, the drawing costs reported by Meade et al. were found relative to comparison tasks that have already been shown to be effective at reducing the DRM illusion rather than a standard read-control group which is typically used to determine baseline correct and false memory rates (see Gallo, 2006; Huff et al., 2015 for reviews). It is perhaps unsurprising then that drawing was found to produce elevated false recognition rates relative to production and individual mental imagery tasks (an accuracy cost) given the effectiveness of these tasks at reducing false recognition.

The purpose of this study was, therefore, to re-examine the costs and benefits of drawing on correct and false recognition in the DRM paradigm by curtailing potential spillover effects of processing using between-subject groups and evaluating recognition relative to a read group baseline. In Experiment 1, the single-image drawing task used by Meade et al. (2020) was compared with a *read-only control group* that was equated in encoding duration. We term this drawing task as the *black-pencil group*, given participants were instructed to generate drawings using a standard black pencil. We expected that the black-pencil group would produce a mirror effect pattern relative to reading based on the well documented benefits of drawing on correct memory (Fernandes et al., 2018) and reductions in false memory following distinctive/item-specific encoding (Huff et al., 2015). These groups were further compared with a second drawing group which similarly drew a single image but were required to use two or more different coloured pencils when producing an image. This *coloured-pencil group* was included to evaluate whether drawing images using different colours would enhance the visual distinctiveness of the drawing versus the standard black-pencil task. We predicted that coloured-pencil drawing may exaggerate the recognition accuracy benefits relative to the black-pencil group due to enhanced distinctive processing. Finally, we applied a signal-detection measure of sensitivity to separate contributions from impoverished relational encoding and the distinctiveness heuristic that may underlie potential costs and benefits of drawing.

In Experiment 2, we further tested drawing effects on memory accuracy by testing whether increases in correct recognition previously observed in the drawing task (Meade et al., 2020) could have reflected the recruitment of distinctive item-specific processing. To this end, we compared the single-image black-pencil drawing task from Experiment 1 with a relational-drawing variant in which participants were tasked with drawing several interactive images of DRM list items simultaneously. We similarly applied a signal-detection analysis to assess the encoding/monitoring processes that are consistent with item-specific and relational processing differences that have been reported previously (e.g., Huff & Bodner, 2013). Collectively, our experiments evaluated whether drawing affects memory accuracy relative to a standard read-only control task in a between-group design, and whether these benefits are consistent with item-specific (vs. relational) processing effects.

Experiment 1: black- and coloured-pencil tasks versus reading

In Experiment 1, black-pencil and coloured-pencil drawing tasks were compared. A read-only control group was included to gauge costs and benefits on correct and false recognition. We expected that both drawing groups would produce an increase in correct recognition relative to the read group, but that the coloured-pencil group would produce an exaggerated increase due to additional colours that may enhance the perceptual distinctiveness of the drawn list items. Our second prediction, based on Foley et al. (2006), who used a single mental imagery task, was that drawing would also reduce false recognition relative to the read group (i.e., a mirror effect) and that this reduction would be greater in the more distinctive coloured-pencil group. Using a signal-detection analysis, drawing benefits to accuracy were expected to reflect contributions from both impoverished relational encoding and a distinctiveness heuristic.

Methods

Participants. Overall, 111 English-proficient participants were recruited from the greater Hattiesburg, MS community and volunteered to participate in the experiment. Participants were randomly assigned to the read-control group ($n=37$), the black-pencil group ($n=36$), or the coloured-pencil group ($n=37$). One participant in the coloured-pencil group was eliminated due to excessive "old" recognition responses across trials resulting in high hits and false alarms (i.e., $>70\%$ of false alarms to control items, which indicated that they did not adhere to task instructions), leaving 110 participants for analysis. A sensitivity analysis using G*Power 3(Faul et al., 2007) indicated that our sample had sufficient power (.80) to detect medium effect sizes or larger (effect size $f>.30$). Mean participant age was 27.99 years ($SD=12.90$; range = 18–74). All participants reported normal or corrected-to-normal vision.

Materials. Eight DRM lists were taken from Meade et al. (2020) which included highly concrete DRM items that were relatively easy to draw. These lists contained items from the window, car, chair, sweet, spider, shirt, needle, and foot critical lure lists. Two versions were then created from these lists: Version A (lists from window, chair, sweet, and needle critical lures) and Version B (lists from car, shirt, spider, and foot critical lures). Each list was ordered from the highest to the lowest backward associative strength to the critical lure based on the strength values reported by Roediger, Watson, et al. (2001). A 40-item recognition test was presented in a newly randomised order for each participant. This test contained 16 studied list words (taken from list positions 1, 3, 5, and 7 in each list), 4 critical lures from each study list, 16 list control words (taken from the same list positions as non-studied lists), and 4 critical lure controls (from each non-studied list).

Procedure. Participants were tested individually on a computer running PowerPoint with an experimenter present. Following consent, all participants were instructed that they would view multiple lists of words and that they would be tested on these words later. They were told that each word would be presented on the screen individually and that each word would only be presented for 10 s. Participants in the drawing groups (black- and coloured-pencil groups) were then presented with a sheet of paper containing eight boxes and instructed to quickly sketch an image of each word's referent in each of the boxes. The coloured-pencil group was further told that they would need to use at least two colours when sketching their images for the purpose of drawing a more elaborative image. Time was measured by an experimenter and started when the participant first touched a pencil to the paper. Participants were further prompted by the experimenter to begin their drawings as soon as a word appeared on the screen. Verbal cues were used when time was up to ensure that participants stopped drawing and were ready to continue onto the next item. Separately, participants in the read-control group were instructed to "read each list item

Table 1. Experiment 1: Mean (±95% CI) Proportion of "Old" Responses and Signal-Detection Indices for the Read, Black-Pencil, and Coloured-Pencil Drawing Groups.

Item type/index	Drawing groups		Read group
	Black pencil	Coloured pencil	
List items	0.97 (0.02)	0.98 (0.01)	0.86 (0.05)
List item controls	0.02 (0.01)	0.01 (0.01)	0.09 (0.04)
List items d'	3.48 (0.13)	3.58 (0.07)	2.69 (0.26)
List items λ	1.77 (0.06)	1.83 (0.03)	1.45 (0.17)
Critical items	0.31 (0.09)	0.30 (0.08)	0.59 (0.10)
Critical item controls	0.02 (0.03)	0.03 (0.04)	0.10 (0.04)
Critical items d'	0.65 (0.22)	0.59 (0.20)	1.17 (0.24)
Critical items λ	1.11 (0.07)	1.08 (0.09)	0.95 (0.10)

silently to yourself" for 10 s.[2] At the end of each list, the words "next list" appeared on the screen, indicating that a new list would begin.

Following completion of the four study lists, all participants completed a math filler task for 2 min followed by a recognition test. The recognition test was presented on a sheet of paper, and participants were asked to place a checkmark into two columns labelled either "old" or "new" which specified their memory decision. Participants were instructed to place a checkmark in the "old" column for those items that were remembered from the study lists and "new" for all words that were not from the study lists. They were further specified to place a checkmark in either the old or new columns but not both. Participants were instructed to respond quickly, but not at a cost to accuracy. No warning regarding the critical lures was provided. Following completion of the recognition test, all participants completed a demographic questionnaire and debriefed. The experiment took 10 min to complete.

Results

Mean proportions of correct recognition, false recognition, and signal-detection indices for each of the three encoding groups are presented in Table 1. We adopt a $p < .05$ significance criterion for all results reported. For concision, we report p-values only for significant comparisons and include effect size estimates using omega squared (ω^2) for analyses of variance (ANOVAs) and Hedge's g for t-tests for all reliable comparisons. For non-significant comparisons, an additional test using a Bayesian estimate of the strength supporting the null hypothesis was included (Masson, 2011; Wagenmakers, 2007). This analysis compares a null model with one that assumes an effect. A p-value is then computed, which corresponds to a probability estimate that the null effect is retained (termed p_{BIC}; Bayesian Information Criterion). We therefore supplement all null effects found using standard null hypothesis significance testing with a p_{BIC} analysis. Finally, linear mixed models were also completed for analyses of raw recognition data. These analyses could not be conducted on our signal-detection estimates as these indices cannot be computed at the trial level as required by the models. We therefore report standard univariate analyses to remain consistent across our analyses but relegate our mixed model analyses on raw recognition to our Supplemental Materials.

Correct recognition. A one-way ANOVA was used to compare correct recognition of list items across the three groups; a significant difference was found, $F(2, 107) = 15.60$, $MSE = .01$, $\omega^2 = .23$. Correct recognition was near ceiling for the coloured-pencil and black-pencil groups, which were equivalent (0.98 vs 0.97), $t < 1$, $p_{BIC} = .85$, and both were greater than the read group (0.98 vs 0.86), $t(72) = 4.51$, $SEM = .03$, $g = 1.04$ and (0.97 vs 0.86), $t(71) = 3.78$, $SEM = .03$, $g = 0.88$, for the coloured-pencil and black-pencil groups, respectively.

We then conducted a one-way ANOVA on d' values to compare the amount of information encoded for list items relative to control items. The d' index was calculated by taking the z-score of the hit rate for the list items minus the z-score of list item controls for each participant. There was a significant difference in d' across groups, $F(2, 107) = 30.33$, $MSE = .29$, $\omega^2 = .36$. Like correct recognition, d' was equivalent between the coloured-pencil and black-pencil groups, (3.58 vs 3.48), $t(71) = 1.31$, $SEM = .08$, $p = .19$, $p_{BIC} = .78$, and was greater in both drawing groups relative to the read group, (3.58 vs 2.69), $t(72) = 6.74$, $SEM = .14$, $g = 1.49$; and (3.48 vs 2.69), $t(71) = 5.57$, $SEM = .15$, $g = 1.22$, for the coloured-pencil and black-pencil groups, respectively. A final one-way ANOVA compared estimates of quantitative memory monitoring for critical lures using lambda (λ). Lambda was calculated by taking the z-score of 1 minus the false alarm rate of list item controls (cf. Huff & Bodner, 2013) in which a higher lambda value indicates more conservative responding, which can be taken as evidence of the amount of monitoring applied at test (also see Gunter et al., 2007; Huff & Aschenbrenner, 2018; Wickens, 2002). Memory monitoring was also found to differ significantly across groups $F(2, 107)=14.60$, $MSE=.11$, $\omega^2=.21$. Monitoring was equivalent between the coloured-pencil and black-pencil groups (1.83 vs 1.78), $t(71)=1.47$, $SEM=.03$, $p=.15$, $p_{BIC}=.74$, but was greater in both drawing groups relative to the read group (1.83 vs 1.45), $t(72)=4.27$, $SEM=.09$, $g=0.98$; and (1.78 vs 1.45), $t(71)=3.57$, $SEM=.09$, $g=0.83$, for the coloured-pencil and black-pencil groups, respectively.

False recognition. False recognition was also found to differ across the three groups, $F(2, 107)=12.51$, $MSE=.08$, $\omega^2=.19$. Like correct recognition, false recognition was equivalent between the coloured-pencil and black-pencil groups (0.30 vs 0.31), $t<1$, $p_{BIC}=.89$. Importantly, both drawing groups showed reduced false recognition of critical lures relative to the read-control group (0.30 vs 0.59), $t(72)=4.33$, $SEM=.07$, $g=1.00$, and (0.31 vs 0.59), $t(71)=3.98$, $SEM=.07$, $g=0.92$, for the coloured-pencil and black-pencil groups, respectively.

Next, we computed d' values for critical lures as an estimate of encoded memory information. For critical lures, d' was computed by taking the z-score of the false alarm rate to critical lures minus the z-score of the false alarm rate for critical lure controls. Note that in this analysis, false alarms to critical lures are treated as hits. A difference was again found across groups, $F(2, 107)=8.37$, $MSE=.46$, $\omega^2=.14$. Encoded memory information was equivalent between the black-pencil and coloured-pencil groups (0.59 vs 0.65), $t<1$, $p_{BIC}=.89$, but both were lower than the read group (0.59 vs 1.17), $t(72)=3.70$, $SEM=.16$, $g=0.85$; and (0.65 vs 1.17), $t(71)=3.18$, $SEM=.16$, $g=0.74$, for coloured-pencil and black-pencil groups, respectively. The lambda index for critical lures was similarly computed by taking 1 minus the z-scored false alarm rate to critical lure controls. Reliable differences were found across groups, $F(2, 107)=3.60$, $MSE=.07$, $\omega^2=.06$. Monitoring was marginally greater in the coloured-pencil group than the read group (1.08 vs 0.95), $t(72)=1.88$, $SEM=.07$, $p=.07$, $g=0.43$, $p_{BIC}=.60$, but significantly greater in the black-pencil group than the read group (1.11 vs 0.94), $t(71)=2.53$, $SEM=.06$, $g=0.59$. Monitoring was equivalent between the black-pencil and coloured-pencil groups (1.08 vs 1.11), $t<1$, $p_{BIC}=.88$.

Discussion

Drawing images of DRM list items increased correct recognition relative to a read-control group—a pattern consistent with other drawing studies (Meade et al., 2020; Wammes et al., 2016). This benefit was equivalent for both the standard black-pencil group and the coloured-pencil group, although correct recognition was at ceiling which may have masked potential differences between the two drawing groups. Signal-detection analyses showed that increases in both encoded memory information and monitoring contributed to drawing benefits over reading.

Turning to the DRM illusion, relative to the read group, both drawing groups reduced false recognition at similar rates. When considered with the correct recognition patterns, drawing in both black-pencil and coloured-pencil formats produced a mirror effect pattern, indicating a net benefit to memory accuracy. Reduced false recognition in both drawing groups was due to a reduction in encoded memory information for critical lures and an increase in monitoring, indicating the presence of both impoverished relational encoding and a distinctiveness heuristic (cf. Huff et al., 2015).

Our finding that both drawing tasks reduce the DRM illusion relative to a read-control group is important, as it builds off Meade et al.'s (2020) results and indicates that drawing can benefit memory accuracy in the DRM paradigm. We discuss this difference in greater detail in the "General Discussion" section, but our results indicate that both the comparison task used to determine costs and benefits and the context in which the comparison is made (i.e., between-groups vs. within-groups) can contribute to whether costs or benefits are found. Indeed, our use of a read-only control group provides a standard baseline with which to evaluate the DRM illusion, and the use of pure groups prevents potential task spillovers. Under these conditions, drawing appears to operate as a cost-free strategy on correct and false recognition.

In contrast to our expectations, however, the use of coloured pencils to produce images did not procure additional recognition benefits relative to the black-pencil condition. We reasoned that the use of additional colours may increase distinctive processing of drawings at study, leading to an exaggerated mirror effect pattern relative to the black-pencil group. The recognition similarities between these groups suggest that drawing benefits over reading need not be highly detailed and can occur monochromatically.

The mirror effect benefit found for drawing is consistent with other distinctive item-specific tasks that have been compared with pure read groups. In our experiment, however, it is unclear whether drawing involves the recruitment of item-specific processing because it has not been directly compared with a relational task designed to recruit a contrasting processing type. In a relational processing task, participants are tasked with associating items together at study, rather than differentiating them in item-specific tasks. In three experiments, Huff and Bodner (2013), compared item-specific instructions, pleasantness ratings, and generation relative to separate task variants designed to recruit relational processing. Across experiments, correct recognition was generally equivalent between the task variants, but item-specific variants reduced false recognition relative to relational variants. Signal-detection analyses consistently indicated that memory monitoring was equivalent between item-specific and relational variants, but item-specific processing reduced the amount of encoded memory information for critical lures, confirming that the different task types affected encoding but not monitoring processes. The comparison between item-specific and relational task variants is therefore important regarding processing claims—a comparison we make in Experiment 2.

Experiment 2: item-specific versus relational drawing tasks

In Experiment 2, we tested whether the standard drawing task reflected the contribution of distinctive item-specific processing. The black-pencil individual-image drawing task used in Experiment 1 was compared with a novel relational-drawing variant in which participants were instructed to draw several integrated images of DRM list items simultaneously. Our relational variant was based on the integrated mental imagery task used by Foley et al. (2006) in which participants combined several items into a single mental image at study. Consistent with Huff and Bodner's (2013) item-specific/relational comparisons, Foley et al. reported a DRM false recognition reduction for lists encoded with single mental images relative to integrated mental images. Based on this finding, we predicted that drawing a single image would similarly reduce the DRM illusion relative to drawing integrated mental images and that this pattern would reflect a reduction in encoded memory information and not monitoring (cf. Huff & Bodner, 2013). For correct recognition, however, Foley et al. found that individual mental images produced a slight increase in correct recognition over integrated images. Therefore, if drawing operates similarly to imagery, we expected that drawing individual images would similarly produce a benefit over integrated drawing.

In Experiment 2, the black-pencil instructions from Experiment 1 were again used with a new group of participants while the coloured-pencil group was dropped due to similarities with the black-pencil group. Based on Foley et al.'s (2006) integrated mental imagery instructions, an integrated-drawing group was included in which participants were given four objects and tasked with drawing an image that integrated all objects. We expected that drawing integrated images would be more likely to recruit relational processing and inflate the DRM illusion relative to drawing individual images, which is more likely to encourage item-specific processing. In addition, the number of DRM lists that were presented to each participant was expanded to 10 (vs. 4 in Experiment 1) as our Experiment 2 sample consisted of undergraduate students, who received partial course credit for participation, rather than community volunteers.

Methods

Participants. Overall, 60 English-proficient participants were recruited from introductory psychology courses at the University of Southern Mississippi and participated for course credit. Participants were randomly assigned to the individual-drawing group ($n=28$) or the integrated-drawing group ($n=32$). Sample sizes were based on Huff and Bodner's (2013) item-specific and relational groups. One individual-drawing participant was eliminated due to excessive "old" recognition responses, leaving 59 participants available for analysis. A sensitivity analysis using G*Power 3 (Faul et al., 2007) indicated our sample had sufficient power (.80) to detect medium-to-large effect sizes and greater (effect size $f > .37$). Mean participant age was 19 years ($SD=3.87$; range 18–46). All participants reported normal or corrected-to-normal colour vision at the time of testing.

Materials. The materials used in Experiment 2 were similar to Experiment 1 with the exception that the number of DRM lists used was expanded to 20 total lists (vs. 8), each of which contained eight items. Lists were again taken from Meade et al. (2020) and similarly arranged to be presented in descending BAS order. In addition to the lists used in Experiment 1, the following list additions were included: Lamp, butterfly, cottage, fruit, soft, high, cabbage, whistle, doctor, music, lion, and river. Two versions were then created from these lists: Version A (lists from chair, high, butterfly, cottage, fruit, sweet, car, soft, window, and lamp critical lures) and Version B (lists from needle, cabbage, spider, whistle, shirt, doctor, foot, music, lion, and river critical lures). A 100-item recognition test was presented in a newly randomised order for each participant which contained 40 studied list words (from list positions 1, 4, 5, and 8 in each list), 10 critical lures from each study list, 40 list control words (taken from the same list positions in non-studied lists), and 10 critical lure controls (1 from each non-studied list).

Procedure. All participants were individually tested on a computer running SuperLab software (Cedrus Corp.) with an experimenter present. Following informed consent, participants were told that they would see a list of study words on the screen and that their job would be to draw an image of the word's referent in the box on a sheet of paper. Individual-drawing participants were given the same black-pencil instructions from Experiment 1 in which they were told they would have 10 s to draw an image of the word presented on the computer screen using a standard pencil. Integrated-drawing participants were given four list words simultaneously on the computer screen and were asked to draw a single image that integrated all four of the list words using a standard pencil in a box on a sheet of paper. To equate for encoding duration with the individual-drawing group, participants in the integrated-drawing group were given 40 s to create an image. Again, time was recorded from the moment when participants started drawing and a verbal cue ensured that participants ended their

Table 2. Experiment 2: Mean (± 95% CI) Proportion of "Old" Responses and Signal-Detection Indices for the Individual-Image and Integrated Image Drawing Groups.

Item type/index	Individual drawing	Integrated drawing
List items	0.93 (0.02)	0.86 (0.03)
List item controls	0.06 (0.02)	0.06 (0.01)
List items d'	3.24 (0.22)	2.82 (0.18)
List items λ	1.68 (0.13)	1.65 (0.10)
Critical items	0.31 (0.06)	0.49 (0.08)
Critical item controls	0.06 (0.04)	0.03 (0.02)
Critical items d'	0.88 (0.19)	1.48 (0.21)
Critical items λ	1.42 (0.13)	1.52 (0.08)

drawing on time and moved on to the next study word or set of four study words. Participants were not informed in advance that the list of words, or sets of words in the integrated-drawing group, was related. To denote separation between the study lists, the words "next list" appeared after the individual-drawing group produced images for all eight list words or after the integrated-drawing group drew integrated images for both sets of four list words. Following encoding of all 10 lists, all participants completed an old/new recognition test that was completed on a keyboard in which participants were instructed to press a labelled "old" key if the word was in the drawn set that was studied and the "new" key if the word was not. Participants were also instructed to respond as quickly as possible without compromising accuracy, and no information was provided about the presence of critical lures on the test list. Following completion of the recognition test, participants completed a demographic questionnaire and were fully debriefed. The experiment took approximately 25 min to complete.

Results

Mean proportions of correct recognition, false recognition, and their corresponding signal-detection indices for the individual-drawing and integrated-drawing groups are presented in Table 2. Correct recognition was greater in the individual-drawing group relative to the integrated-drawing group (0.93 vs 0.86), $t(57)=3.34$, $SEM=.02$, $g=0.87$, which reflected an increase in encoded memory information (d') in the individual-drawing over the integrated-drawing group (3.24 vs 2.82), $t(57)=2.85$, $SEM=.15$, $g=0.74$, but not due to differences in memory monitoring (1.68 vs 1.65), $t<1$, $p_{BIC}=.88$.

For false recognition, the DRM false memory illusion was lower in the individual-drawing group than the integrated-drawing group (0.31 vs 0.49), $t(57)=3.35$, $SEM=.05$, $g=0.86$, and reflected a decrease in encoded memory information to the individual-drawing group relative to the integrated-drawing group (0.88 vs 1.48), $t(57)=4.06$, $SEM=.15$, $g=1.05$. Like correct recognition, there was no difference between the groups in memory monitoring (1.42 vs 1.52), $t(57)=1.19$, $SEM=.08$, $p=.24$, $p_{BIC}=.79$. Taken together, individual drawing produced a mirror effect pattern relative to the integrated-drawing group, and this pattern was due to an increase in encoded memory information for list items and a reduction in encoded memory information for critical lures.

Discussion

Consistent with Foley et al. (2006), we found that drawing individual images led to an increase in correct recognition, and this increase was due to an increase in encoded memory information but not monitoring. False recognition was also reduced in the individual-drawing group relative to the integrated-drawing group, and again, this pattern was due to a reduction in encoded memory information but not monitoring. The difference between individual and integrated-drawing groups is also consistent with the item-specific/relational variants used by Huff and Bodner (2013) which suggest that drawing of individual images likely involves the recruitment of item-specific processing. Our finding that qualitative differences in how the drawing task is completed affects estimates of encoded memory information but not monitoring which suggests that our drawing tasks affect processes selectively at study.

General discussion

The primary goal of our experiments was to evaluate the effects of drawing at encoding on recognition accuracy in the DRM paradigm. In Experiment 1, we compared between-groups that were tasked with drawing referents of DRM list items with either a standard black pencil or using coloured pencils relative to a read-only control group. Consistent with the well documented benefits of drawing on correct memory (Fernandes et al., 2018; Wammes et al., 2016), we similarly found that both drawing tasks improved correct recognition over reading. Importantly, false recognition of critical lures also reflected drawing benefits in which both drawing groups reduced false recognition relative to the read-only control. Signal-detection analyses revealed consistent loci for these patterns. Benefits to correct recognition for both drawing groups were due to an increase in encoded memory information and enhanced test-based monitoring. Relatedly, benefits to false recognition for both drawing groups were due to a reduction in encoded memory information for critical lures (consistent with impoverished relational encoding; Hege & Dodson, 2004) and an increase in monitoring (consistent with a distinctiveness heuristic; Schacter et al., 1999).

In Experiment 2, we compared the black-pencil drawing task from Experiment 1 in which an individual image was drawn for each DRM list item with an integrated-drawing group that was instructed to integrate four list items into a single drawing. We expected that integrated drawings would encourage relational processing which has consistently been shown to increase false recognition relative to an item-specific processing task (Foley et al., 2006; Huff & Bodner, 2013; McCabe et al., 2004). Consistent with these previous patterns, the individual-drawing group reduced false recognition which reflected a decrease in encoded memory information relative to the integrated-drawing group but produced no difference in memory monitoring. This pattern is consistent with other item-specific/relational comparisons (e.g., Huff & Bodner, 2013), suggesting that drawing individual images promotes item-specific processing relative to reading and integrated drawing. In terms of activation monitoring and fuzzy-trace processes, drawing individual images therefore appears to reduce either the implicit activation of critical lures (Roediger, Balota, et al., 2001) or restrict the extraction of a consistent gist theme (Brainerd & Reyna, 2002) while also facilitating monitoring at test relative to a read-control task.

Our finding that individual drawing and integrated drawing produced differences in both correct and false recognition provides further evidence that the qualitative processes (i.e., item-specific vs relational) that are engaged during a task are perhaps more impactful on subsequent memory than the task itself, at least when associative lures are factored into overall memory accuracy. Although both groups engaged in drawing in Experiment 2, recognition was greater when drawing individual images. This pattern indicates that when drawings emphasise individual items, recognition may be particularly enhanced relative to when several images are drawn together. Of course, this pattern does not indicate that drawing integrated images is ineffective at enhancing memory as the integrated image task was not directly compared with a read-only control.

Our experiments were also designed to follow-up experiments by Meade et al. (2020) who reported a drawing cost to false recognition when drawing was compared with writing and mental imagery tasks. Specifically, we included a pure read-only control group to obtain a baseline false recognition comparison which was then used to gauge potential drawing costs and benefits. This baseline was critical for understanding how drawing would compare to a standard DRM encoding task that is often used in the literature. In addition, we used a between-group design to curtail spillover effects that have been shown to affect false recognition rates in the DRM paradigm (Huff et al., 2021). Finally, we included a signal-detection analysis to provide quantitative estimates of the contributions of encoding and monitoring processes following drawing. Our results indicate that compared with a standard DRM baseline, drawing of individual images does not produce a cost to recognition accuracy, as evidenced by a mirror effect pattern. Moreover, drawing benefits reflect enhancements to both encoding and monitoring processes.

Although our experiments were designed to eliminate contributions of task spillovers using a between-subject design, we acknowledge that an additional method for evaluating spillovers would be to measure them directly. As Huff et al. (2021) showed, spillovers can be gauged by comparing tasks completed within-subjects with the same tasks completed in a pure group. A potential area for future research would be to directly assess how drawing processes spillover to comparison tasks in a within-group and vice versa, given spillovers have been shown to be bidirectional. This type of comparison would not only be important for directly gauging spillovers but it also could be informative of when drawing tasks are more or even less effective. For instance, drawing benefits on memory may be particularly affected by the presence of some within-subject tasks but not others, which could provide important boundaries on when drawing should or should not be used in practice.

Interestingly, however, having participants draw individual images with multiple colours did not procure additional benefits on recognition. In contrast to our prediction, coloured drawings were equivalent to the standard black-pencil drawing group on both correct and false recognition. One possibility for this similarity may be due to the black-pencil group already maximising recognition benefits. Indeed, correct recognition exceeded 90% in both experiments and false recognition was 31%, a rate considerably lower than false recognition rates found in other item-specific tasks (cf. Huff et al., 2015). Of course, in the present experiments, encoding durations were generally longer (10 s vs self-paced or a few seconds) and study lists had fewer items in each list than these other studies (8 vs. 12–15 items), which could also contribute to the lower false recognition rates found across our experiments (Coane et al., 2007; Robinson & Roediger, 1997, for list-length differences; see Supplemental Materials for 10 s and 1.5 s read group duration comparisons). An interesting question is therefore whether drawing is equally effective or perhaps even more effective at improving recognition accuracy relative to other item-specific tasks that have been tested. Another possibility is that participants may not have had enough time to generate drawings with sufficient colour detail to yield additional improvements. Although participants were instructed to use at least two colours, most drawings (greater than 80%) only used two colours and participants were required to switch colours in the middle of their drawings. Therefore, this limited use of colour combined with switching may have diminished any benefits of colour drawings over the black-pencil drawing group.

Our study also provides important information regarding potential processes that underly drawing benefits. We argue that the standard individual-drawing task likely promotes item-specific processing given correct and false recognition rates mimic tasks that promote item-specific processing at study. While we triangulate our processing claims here based on comparisons with tasks that are more likely to recruit relational processing (reading related word lists and drawing integrated images), we do not have independent measures of item-specific or relational processing for these tasks (see Burns, 2006 for review) which can include computing adjusted ratio of clustering scores on recall (Huff & Bodner, 2014; Hunt & Seta, 1984; Roenker et al., 1971) and cumulative recall curves (Burns et al., 2007). However, the reliability of false recognition of DRM critical lures following item-specific and relational encoding suggests that the DRM paradigm serves as an important tool for gauging processing in various encoding tasks and that individual/integrated-drawing tasks are consistent with the item-specific and relational processing patterns in the literature.

Given the potential processing benefits of drawing and the potential utility of drawing in applied contexts (see Fernandes et al., 2018, for discussion), an important question for future research is whether individual drawing as a study technique may similarly reduce related memory errors in applied contexts, such as educational settings. Indeed, Wammes et al. (2017) found that drawing images of educationally relevant definitions of terms can benefit memory relative to producing a written transcription of the definition (but not paraphrasing). Whether drawing can also reduce the frequency of memory errors may also be another avenue for improving educational performance. For instance, when designing multiple-choice type exam questions, instructors frequently choose distractors that share some relation to the target test option. If drawing individual images of educational concepts reduces errors for related distractors, students may be better served by drawing images in isolation. Of course, while this possibility remains to be tested, our data patterns reported here suggest that drawing may be more effective in a pure-drawing condition (i.e., all materials are encoded through drawings, not just a subset) and if drawings are made of individual versus integrated concepts.

Conclusion

Our results indicate that drawing individual images of list items at encoding can produce a benefit to memory accuracy relative to both reading and drawing integrated images in a pure-group context. Given possible drawing costs that have been reported when drawing was used to encode a subset of lists (Meade et al., 2020), our results provide evidence that spillovers may emerge when drawing is manipulated within-subjects which may affect whether drawing produces a cost or benefit on recognition accuracy. Both the type of experimental design and the type of comparison task used are therefore critical for evaluating the effects of a given encoding task on memory accuracy. In addition, our results suggest that drawing individual images is more likely to recruit item-specific processing which can produce a mirror effect pattern in correct and false recognition relative to a standard read-only control and integrated-drawing tasks. Individual drawing therefore appears to operate as a cost-free task by enhancing overall recognition accuracy, at least when compared with a read-control task in a pure-group context.

Acknowledgements

The authors thank Dr. Myra Fernandes for providing study materials for their experiments.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Mark J Huff https://orcid.org/0000-0002-0155-7877

Nicholas P Maxwell https://orcid.org/0000-0003-4517-8323

Data accessibility statement

Data used in analyses and Supplementary Materials are available through our OSF page (osf.io/r4yh9/).

Supplementary material

The supplementary material is available at: qjep.sagepub.com.

Notes

1. The spillover effects on false recognition were found to driven by encoding-based processes (i.e., d' estimates) and not monitoring processes, consistent with the notion that spillovers are driven by changes at study. In addition, Huff et al. (2021) included a relational generation within-group variant and a pure relational group. The within-relational group similarly showed false recognition patterns consistent with relational processing spillovers, indicating that spillover effects can occur for both item-specific and relational processing.
2. A 10 s per item study duration was used to equate encoding time across the drawing and reading groups. One possibility, suggested by an anonymous reviewer, is that the relatively long duration may have encouraged participants to disengage at encoding, possibly due to boredom. If so, read participants may have "loafed" at study, which could have reduced memory accuracy contributing to the mirror effect patterns found in the drawing groups. We reasoned that if participants loafed, their performance would mimic that of a group with a lower encoding duration. We therefore tested a second control group using a standard 1.5 s

encoding duration (cf. Roediger & McDermott, 1995). Recognition accuracy was lower in the 1.5 s group, indicating that 10 s read group participants were unlikely to be loafing. Analyses comparing the two read groups and the 1.5 s read group to the drawing groups are reported in our Supplemental Materials (osf.io/r4yh9/).

References

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, *48*, 1–15. https://doi.org/10.1016/s0749-596x(02)00518-1

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 553–563. https://doi.org/10.1037/0278-7393.13.4.553

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201–210. https://doi.org/10.3758/BF03193441

Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, D.-L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology*, *70*, 93–98. https://doi.org/10.1037/cep000082

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*, *21*, 149–154. https://doi.org/10.3758/s13423-013-0485-1

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164–169. https://doi.org/10.1111/1467-8721.00192

Burns, D. J. (2006). Assessing distinctiveness: Measures of item-specific and relational processes. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 109–130). Oxford University Press.

Burns, D. J., Jenkins, C. L., & Dean, E. E. (2007). Falsely recalled items are rich in item-specific information. *Memory & Cognition*, *35*, 1630–1640. https://doi.org/10.3758/BF03193497

Coane, J. H., McBride, D. M., Raulerson, B. A., III, & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*, 62–70. https://doi.org/10.1027/1618-3169.51.1.62

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341–361. https://doi.org/10.1016/0749-596X(87)90118-5

Craik, F. I. M. (2002). Levels of processing: Past, present. and future? *Memory*, *10*, 305–318. https://doi.org/10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22. https://doi.org/10.1037/h0046671

Dodson, C. S., & Schacter, D. L. (2001). “If I had said it I would have remembered it”: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161. https://doi.org/10.3758/BF03196152

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*, *43*, 175–191. https://doi.org/10.3758/BF3193146

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, *27*, 302–308. https://doi.org/10.1177/0963721418755385

Foley, M. A., Wozniak, K. H., & Gillum, A. (2006). Imagination and false memory inductions: Investigating the role of process, content and source of imaginations. *Applied Cognitive Psychology*, *20*, 1119–1141. https://doi.org/10.1002/acp.1265

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 120–128. https://doi.org/10.1037/0278-7393.30.1.120

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848. https://doi.org/10.3758/MC.38.7.833

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271–276. https://doi.org/10.3758/BF03209405

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 5–16. https://doi.org/10.1037/0278-7393.16.1.5

Gunter, R. W., Bodner, G. E., & Azad, T. (2007). Generation and mnemonic encoding induce a mirror effect in the DRM paradigm. *Memory & Cognition*, *35*, 1083–1092. https://doi.org/10.378/BF03193480

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 787–795. https://doi.org/10.1037/0278-7393.30.4.787

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, *24*, 202–216. https://doi.org/10.3758/BF0320081

Huff, M. J., & Aschenbrenner, A. J. (2018). Item-specific processing reduces false recognition in older and younger adults: Separating encoding and retrieval using signal detection and the diffusion model. *Memory & Cognition*, *46*, 1287–1301. https://doi.org/10.3758/s13421-018-0837-1

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*, 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. https://doi.org/10.1016/j.jml.2014.02.004

Huff, M. J., & Bodner, G. E. (2019). Item-specific and relational processing both improve recall accuracy in the DRM paradigm. *Quarterly Journal of Experimental Psychology*, *72*, 1493–1506. https://doi.org/10.1177/17470-21818-801427

Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin & Review*, *22*, 349–365.

Huff, M. J., Bodner, G. E., & Gretz, M. R. (2021). Distinctive encoding of a subset of DRM lists yields not only benefits, but also costs and spillovers. *Psychological Research*, *85*, 280–290. https://doi.org/10.1007/s00426-019-01241-y

Huff, M. J., Di Mauro, A., Coane, J. H., & O'Brien, L. M. (2021). Mapping the time course of semantic activation in mediated false memory: Immediate classification, naming, and recognition. *Quarterly Journal of Experimental Psychology*, *74*, 483–496.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, *20*, 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 454–464. https://doi.org/10.1037/0278-7393.10.3.454

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*, 390–395. https://doi.org/10.1177/0963721417691356

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679–690. https://doi.org/10.3758/s13428-010-0049-5

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review*, *11*, 1074–1079. https://doi.org/10.3758/BF03196739

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*, 1065–1077. https://doi.org/10.3758/BF03194324

McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory*, *14*, 570–583. https://doi.org/10.1080/096582106006245-64

Meade, M. E., Klein, M. D., & Fernandes, M. A. (2020). The benefit (and cost) of drawing as an encoding strategy. *Quarterly Journal of Experimental Psychology*, *73*, 199–210. https://doi.org/10.1177/1747021819869188

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 263–273. https://doi.org/10.1037/0278-7393.33.2.263

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*, 717–727. https://doi.org/10.1080/09658211.2012.699070

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Robinson, K. J., & Roediger, H. L., III (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*, 231–237. https://doi.org/10.1111/j.1467-9280.1997.tb00417.x

Roediger, H. L., III, Balota, D., & Watson, J. (2001). Spreading activation and arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). American Psychological Association.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., III, Watson, J. M., McDermott, K. A., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407. https://doi.org/10.3458/BF03196177

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45–48. https://doi.org/10.1037/h0031355

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1–24. https://doi.org/10.1006/jmla.1998.2611

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition*, *30*, 1054–1064. https://doi.org/10.3758/BF03194323

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804. https://doi.org/10.3758/BF03194105

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, *69*, 1752–1776. https://doi.org/10.1080/17470218.2015.1094494

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017). Learning terms and definitions: Drawing and the role of elaborative encoding. *Acta Psychologica*, *179*, 104–113. https://doi.org/0.1016/j.actpsy.2017.07.008

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517. https://doi.org/10.1006/jmla.2002.2864