Distinctive encoding of a subset of DRM lists yields not only benefits, but also costs and spillovers

Mark J. Huff, Glen E. Bodner & Matthew R. Gretz

Psychological Research
An International Journal of Perception, Attention, Memory, and Action

ISSN 0340-0727
Volume 85
Number 1

Psychological Research (2021)
85:280-290
DOI 10.1007/s00426-019-01241-y

Springer

Psychological Research (2021) 85:280–290
https://doi.org/10.1007/s00426-019-01241-y

ORIGINAL ARTICLE

Distinctive encoding of a subset of DRM lists yields not only benefits, but also costs and spillovers

Mark J. Huff[1] · Glen E. Bodner[2] · Matthew R. Gretz[1]

Received: 3 June 2019 / Accepted: 20 August 2019 / Published online: 28 August 2019

Abstract

Prior research has emphasized that performing distinctive encoding on a subset of lists in the DRM paradigm suppresses false recognition; we show that its benefits can be mitigated by costs and spillover effects. Within groups read half the DRM lists and solved anagrams for the other half using a strategy that emphasized either item-specific or relational processing. Their recognition was compared to three pure-list control groups (read, item-specific generation, relational generation). Correct recognition in the within groups showed a benefit for generate items and a cost for read items, resulting in little net improvement relative to pure reading. False recognition in the within groups was reduced following item-specific vs. relational generation, but there was again little net improvement. Most surprisingly, false recognition in the within groups was *greater* for generate than read lists. This pattern suggests that relational processing of read lists spilled over to generate lists, boosting false recognition for generate lists. Distinctive encoding of a subset of items does not appear to globally improve memory accuracy.

Introduction

Distinctive information is often better remembered than non-distinctive information (Hunt, 2006; Schmidt, 1991). Distinctiveness can arise at the item level, such as when items are bizarre or perceptually unique (McDaniel & Einstein, 1986, see Hunt & Worthen, 2006 for review), or can be induced by study tasks that emphasize the encoding of unique, item-specific features (Hunt & Einstein, 1981). Studies using the Deese–Roediger–McDermott (DRM) false memory paradigm (Deese, 1959; Roediger & McDermott, 1995) have shown that distinctive item-specific processing can also improve memory accuracy by reducing false recognition (Huff & Bodner, 2013; McCabe, Presmanes, Robertson, & Smith, 2004Smith & Hunt, 1998; see Huff, Bodner, & Fawcett, 2015 for review).

In the DRM paradigm, participants study lists of related words (e.g. *bed*, *rest*, *tired*, etc.). Each list converges on a single non-presented critical lure (e.g. *sleep*), which participants falsely recognize or recall at a high rate and with high confidence (Roediger & McDermott, 1995; see Gallo, 2006 for a review). Distinctive encoding can reduce this DRM illusion. For example, Israel and Schacter (1997) found that the illusion was reduced in a group that saw a picture corresponding to each presented list word relative to a group that merely read each list word. This between-group pattern has been replicated with various distinctive tasks such as generating words from anagrams (Gunter, Bodner, & Azad, 2007; McCabe & Smith, 2006), creating mental images of each list word (Foley, Wozniak, & Gillum, 2006; Gunter et al. 2007; Oliver, Bays, & Zabrucky, 2016), and distinctive processing instructions (Huff & Bodner, 2013; 2019; Huff & Aschenbrenner, 2018; McCabe et al., 2004).

Two primary mechanisms have been proposed to account for the reduction of false recognition in the DRM paradigm induced by distinctive processing: one that operates at encoding, and the other at test. According to the *impoverished relational encoding account* (Hege & Dodson, 2004), performing distinctive processing at study disrupts the thematic consistency or the associative strength between the list items and their associated critical lures (Brainerd & Reyna, 2002; Roediger, Balota, & Watson, 2001). According to the *distinctiveness heuristic account*, participants adopt a memory-monitoring strategy at test in which they endorse only those items for which they are able to recollect having

✉ Mark J. Huff
mark.huff@usm.edu

1 School of Psychology, The University of Southern Mississippi, Hattiesburg, MS 39406, USA

2 College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia

performed the distinctive processing at study, thus helping them avoid endorsing the critical lures (e.g. Gallo, 2004; 2010).

Within-group condition

A within-group condition has often been used to distinguish between these accounts (for a review of other methods and issues surrounding their use, see Huff et al., 2015). Within groups typically study half the DRM lists in a distinctive manner, and the other half in a non-distinctive manner such as by reading silently (i.e. the standard DRM task). Impoverished relational encoding should occur only for the distinctive lists, thus by this account, the DRM illusion should be lower for the distinctive lists. This pattern was obtained by Arndt and Reder (2003). However, most studies have instead found similar rates of false recognition for both list types (e.g. Dodson & Schacter, 2001; Gunter et al., 2007; McCabe & Smith, 2006; Schacter, Israel, & Racine, 1999). The latter pattern is consistent with participants applying a distinctiveness heuristic to all test items, thus enabling them to reject critical lures from both list types.

Relying on a within-group comparison is potentially problematic for identifying the cause of distinctive processing effects. For instance, performing a distinctive task on some DRM lists may encourage participants to process the read lists in a more distinctive fashion than they otherwise would have done. In the present work, we refer to the possibility that the type of processing applied to one set of items might then be applied to another set of items as a *spillover effect*. Spillover effects can potentially be bidirectional. Therefore, similar levels of false recognition for both list types in a within group might reflect impoverished relational encoding of distinctive lists spilling over to read lists, and/or relational encoding of read lists spilling over to distinctive lists—rather than indicating the use of a distinctiveness heuristic at test.

Signal-detection approach

As an alternative to using a within group to contrast the two accounts of distinctive processing effects, Huff and Bodner (2013) and Gunter et al. (2007) advocated for a signal-detection approach that provides separate estimates of encoding and test-based influences on recognition in the DRM paradigm. Signal detection separates memory for studied versus non-studied information (i.e. discriminability) from the tendency to report that an item was studied (i.e. response bias). When applied to the DRM paradigm, the amount of memory information encoded for list items and for critical lures can be estimated (using d'), as can the amount of memory monitoring at test (such as use of a distinctiveness heuristic) for list items and for critical lures (using lambda, λ).

Huff and Bodner (2013) applied the signal-detection approach to recognition in the DRM paradigm to examine two overlapping issues. First, they examined the locus of the effects for processing that is *item-specific* in nature, that is, when the encoding task emphasized the processing of unique features of each DRM list item (Hunt & Einstein, 1981). To this end, they compared recognition in between groups that performed item-specific processing on all lists to pure-read groups who simply read all lists. Across experiments, the item-specific groups either solved an anagram of each list word, rated the pleasantness of each list word, or were given instructions to focus on the unique features of each list word.

Second, Huff and Bodner (2013) examined the influence of encoding tasks that emphasized *relational* processing of shared characteristics of list items (Hunt & Einstein, 1981). To this end, the anagram task, pleasantness rating task, or task instructions were modified to emphasize use of relational processing. Thus, each experiment compared recognition across a pure item-specific group, a pure relational group, and a pure-read group.

Huff and Bodner (2013) consistently found that item-specific processing reduced false recognition whereas relational processing did not. Moreover, the effect of item-specific processing was driven by a combination of reduced encoding of information about critical lures (consistent with impoverished relational encoding) and an increase in test-based monitoring (consistent with use of a distinctiveness heuristic). Relational processing also increased memory monitoring at test, but that monitoring was ineffective because the relational encoding tasks had already led to encoding of the critical lures (see also Bodner, Huff, Lamontagne, & Azad, 2017).

Revisiting the within-group condition

Huff et al. (2015) confirmed these dual effects of distinctive encoding (i.e. at encoding and at retrieval) in a meta-analysis of previous DRM studies that manipulated distinctive processing in both within groups and pure-list groups. Relevant to the present study, they also examined the effects of performing distinctive item-specific processing on half of the DRM lists (i.e. in a within group) relative to pure-list groups. Huff et al.'s meta-analyses suggested that performing distinctive item-specific processing on some lists spilled over to influence the encoding of the read lists. Consistent with this possibility, less memory information was encoded about the critical lures from read lists in the within groups compared to the pure-read groups. This difference suggests that distinctive item-specific processing spilled over to the read lists, thereby reducing the encoding of critical lures for those lists relative to the pure-read groups.

Huff et al.'s (2015) analyses also found evidence of the converse spillover effect: The processing applied to read lists spilled over to distinctive lists. Specifically, more information was encoded for critical lures from distinctive lists in the within groups than in the pure item-specific groups. Thus, distinctive processing was less effective when applied to only a subset of lists. This pattern suggests that relational processing of read lists, which should commonly occur given the associative structure of DRM lists (Huff & Bodner, 2014; Hunt & Seta, 1984), might spill over to the distinctive lists. In sum, although comparisons of within and pure-list groups may help adjudicate between accounts of distinctiveness effects, the full pattern of costs, benefits, and spillover effects in a within-group context has yet to be carefully evaluated.

The present study

Building on Huff et al.'s (2015) meta-analysis, our study had two aims. Our first aim was to establish the pattern of benefits, costs, and spillover effects that result when distinctive item-specific processing is applied to a subset of DRM lists. To this end, we compared a *within-item-specific generation group* that generated half the lists and read half the lists to both a *pure item-specific generation group* and a *pure-read group*. These comparisons allowed us to assess the potential benefits and costs of performing distinctive item-specific processing for a subset of lists (Bodner, Taikh, & Fawcett, 2014), as well as potential spillover effects across list types. Benefits refer to an improvement in memory for a list type in a within group relative to a pure group. Costs refer to an impairment in memory for a list type in a within group relative to a pure group. Benefits and costs could reflect a spillover in the processing applied from one list type to the other list type, but they could also arise for other reasons. For example, participants may encode the read lists more superficially in a within group, yielding costs (e.g. Begg & Snider, 1987; Bodner et al., 2014).

To date, spillover effects have only been examined for within groups performing item-specific processing (see Huff et al., 2015). Therefore, our second aim was to examine the benefits, costs, and spillover effects that result from *relational* encoding. To this end, we also tested a *within-relational generation group* that performed a variant of the anagram generation task for half of the lists that emphasized relational processing (Huff & Bodner, 2013). In all prior within-group studies, a task emphasizing item-specific processing of list words was always compared to a standard read task which recruits relational processing given the related nature of DRM lists. With two relational tasks, we expected that spillover effects in the within-relational generation group would be less prominent or absent given their redundancy in processing.

In sum, we compared correct and false recognition, as well as our signal-detection indices of encoding and monitoring, across five groups. The within-item-specific generation, within-relational generation, and pure-read groups tested here were compared to each other and to the pure item-specific generation and pure relational generation groups. The pure groups were taken from Huff and Bodner (2013, Experiment 3), which were tested under identical conditions.

For correct recognition, we expected generation effects in both within groups—higher correct recognition for generate vs. read list items (Slamecka & Graf, 1978; Bertsch, Pesta, Wiscott, & McDaniel, 2007), complementing the between-subject generation effects reported in Huff and Bodner (2013). Based on Huff and Bodner's (2015) meta-analysis, these within-subject generation effects were expected to reflect both a benefit for generate items (relative to their respective pure-generation group) and a cost to read items (relative to the pure-read group).

We expected lower false recognition in the item-specific vs. relational within group, based on Huff and Bodner's (2013) pure group findings. We also expected that false recognition would be low and equivalent for generate and read lists in the within-item-specific generation group, in keeping with the distinctiveness heuristic account of the findings reviewed in Huff et al.'s (2015) meta-analyses. In the within-relational group we expected high levels of false recognition for both list types. We expected our signal-detection analyses to show that within-item-specific generation reduces the DRM illusion through both encoding and monitoring processes, whereas relational generation would increase monitoring but would be ineffective at reducing false recognition because encoding of critical lures had already occurred.

Finally, given the potential for benefits, costs, and/or spillover effects to all be in play in the within groups, we also report net analyses of correct and false recognitions. The net analyses averaged across the two list types in the within groups, enabling us to assess whether performing item-specific or relational processing on half the lists improved overall recognition relative to both a pure-read group and the corresponding pure-generation group.

Methods

Participants

Undergraduates at the University of Calgary participated for course credit and were randomly assigned to either a within-item-specific generation group ($N = 24$), a within-relational generation group ($N = 24$), or a pure-read group ($N = 24$). The pure item-specific generation group ($N = 30$) and pure relational generation group ($N = 30$) reported in

Huff and Bodner (2013, Experiment 3) were included to allow us to establish the full pattern of possible costs, benefits, and spillovers in our within-group conditions. Those data as well as the present data were collected in the same year and in the same lab, using the same computers, instructions, items, and procedure. Moreover, the recognition means for the pure-read group in Huff and Bodner (2013, Experiment 3; see their Table 3) were between 0.01 and 0.02 of the means for our pure-read group (see our Table 1; all *t*s < 1.07, *p*s > 0.29). The similarity of these two pure-read groups supports, by triangulation, the comparisons we report using Huff and Bodner's pure-generate groups.

Materials

The 20 DRM lists used in Huff and Bodner (2013), originally from the Roediger, Watson, McDermott, and Gallo (2001) norms, were divided into two equal sets. The set that was studied was counterbalanced across participants. Each list consisted of 12 words (minimum length of 4 letters), presented in descending backward-associative strength from the critical lure. Anagrams of the list words were constructed by swapping letters one and three or two and four (counterbalanced across participants). Each relational list included a single related cue word (also from Huff and Bodner); this was the highest available associate in terms of backward-associative strength from the norms for that list that was not one of the 12 words on that study list. The same related cue word was presented in capital letters next to each word from the corresponding list during study (e.g. FEATHER—rahd; for *hard*). The 10 lists were presented in the same once-randomized order for all participants. In the within-subject generation groups, 5 lists were read aloud, and 5 lists required generation. Generation vs. reading alternated across the 10 lists and was counterbalanced across participants.

The 80-item recognition test was presented in a newly randomized order for each participant. For the pure groups, the test comprised of 30 list items (from positions 1, 8, and 10 of each studied list), 30 item controls (from the same list positions from non-studied lists), 10 critical lures (one per studied list), and 10 critical lure controls (one per non-studied list). For the within groups, the test comprised of 15 read list items, 15 generate list items, 5 read list critical lures, 5 generate list critical lures, 30 list item controls, and 10 critical lure controls.

Procedure

The procedure followed Huff and Bodner (2013, Experiment 3). Participants were tested individually via computer. They read over the study instructions, then the experimenter verbally reviewed them. Depending on their group assignment, participants were instructed to read each word aloud for the read lists and/or to solve the anagrams for the generate lists and to read the solutions aloud. By not providing a strategy for solving the anagrams, it was expected that the item-specific generation groups would largely focus on solving each anagram (i.e. item-specific processing). In contrast, the relational generation groups were instructed to use the related cue word to help them solve the anagrams, thereby encouraging relational processing based on the list theme. To foster this goal, relational generation groups were required to read both the related word cue and the anagram solution aloud on each trial. If the participant could not solve the anagram within a few seconds they were given a hint (the first letter of the solution). The experimenter scored each anagram trial as "correct", "hint", or "pass." If a solution

Table 1 Mean (SE) proportions of "old" responses and signal-detection indices by item type and group

Item type/index	Within item-specific generation			Within relational generation			Pure read	Pure item-specific generation[a]	Pure relational generation[a]
	Read lists	Generate lists	Avg.	Read lists	Generate lists	Avg.			
List items	0.60 (0.03)	0.90 (0.02)	0.75 (0.02)	0.65 (0.03)	0.91 (0.01)	0.78 (0.02)	0.74 (0.02)	0.84 (0.02)	0.84 (0.02)
List item controls	–	–	0.12 (0.02)	–	–	0.08 (0.02)	0.09 (0.01)	0.06 (0.01)	0.06 (0.01)
List item *d*′	1.53 (0.12)	2.59 (0.12)	2.06 (0.12)	1.96 (0.12)	2.93 (0.12)	2.45 (0.11)	2.11 (0.09)	2.73 (0.11)	2.79 (0.11)
List item λ	–	–	1.27 (0.09)	–	–	1.54 (0.10)	1.44 (0.09)	1.67 (0.08)	1.65 (0.07)
Critical lures	0.46 (0.06)	0.60 (0.05)	0.53 (0.04)	0.64 (0.06)	0.73 (0.05)	0.69 (0.04)	0.62 (0.04)	0.39 (0.04)	0.54 (0.05)
Critical lure controls	–	–	0.13 (0.03)	–	–	0.14 (0.02)	0.16 (0.02)	0.09 (0.02)	0.05 (0.02)
Critical lure *d*′	1.07 (0.17)	1.44 (0.13)	1.25 (0.13)	1.52 (0.14)	1.74 (0.14)	1.63 (0.10)	1.42 (0.13)	1.03 (0.12)	1.61 (0.16)
Critical lure λ	–	–	1.18 (0.11)	–	–	1.13 (0.08)	1.05 (0.09)	1.34 (0.09)	1.47 (0.08)

[a]From Huff and Bodner (2013, Experiment 3). Control item means in the within-generation groups appear in the average column because they were not studied and thus could not be assigned to list type

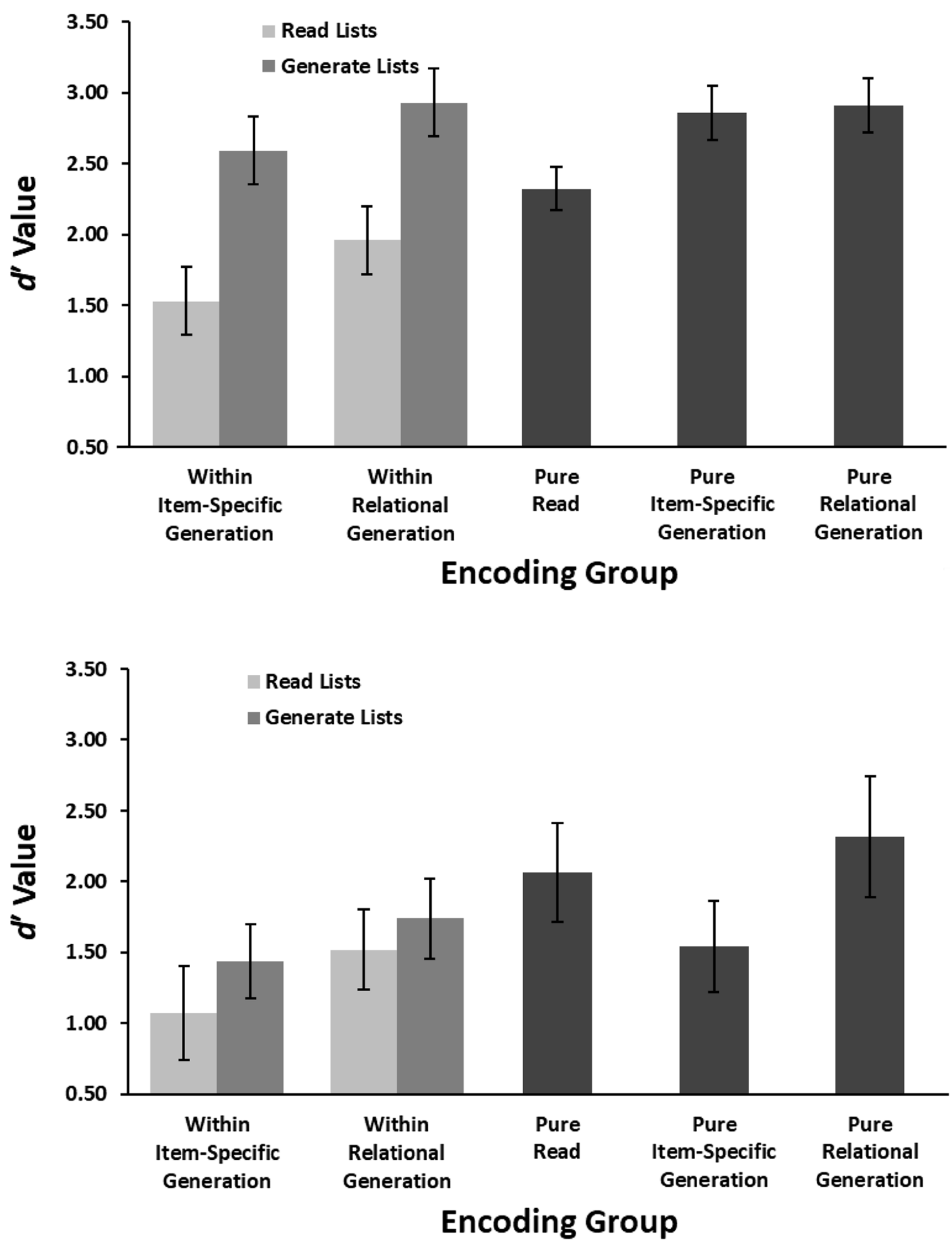
could not be generated following a hint, participants were provided with the solution and said it aloud. At the end of each list, the words "next list" appeared.

Participants completed the old/new recognition test immediately after the study phase. They were instructed to quickly and accurately classify an item as "old" if it had been presented on a study list, or "new" if the item had not been presented on a study list. They responded by pressing labelled "old" or "new" buttons on a response box.

Results

Table 1 presents the mean proportions of "old" responses and mean signal-detection indices as a function of item type and group. To highlight our primary findings, we include a summary of d' values for correct and false recognition (Fig. 1). The mean anagram completion rate across generation groups was 95% and, therefore, analyses were not conditionalized on correct generation (as in Huff & Bodner, 2013). A $p < 0.05$ level of significance was used. For significant comparisons, effect sizes using partial eta squared

Fig. 1 Mean (bars = ± 95% CI) correct and false recognition d' values by encoding group



(η_p^2) are reported for analyses of variance (ANOVAs), and Cohen's *d* for *t* tests.

Correct recognition

Huff and Bodner (2013, Experiment 3) established generation effects in both their generate groups relative to their pure-read group; we do not repeat their analyses here. To evaluate the generation effects in our within groups, we conducted separate 2(list type: generate vs. read) × 2(within-generation task: item-specific vs. relational) mixed ANOVAs on (a) hits and (b) our signal-detection measure of memory information for list items (*d′* in which list items were hits, and list item controls were false alarms). Both ANOVAs revealed a main effect of list type, indicating a higher hit rate for generate than read lists (0.91 vs. 0.62), $F(1, 46) = 209.72$, MSE = 0.01, $\eta_p^2 = 0.82$, and more encoded memory information for list items from generate than read lists (2.76 vs. 1.74), $F(1, 46) = 280.73$, MSE = 0.09, $\eta_p^2 = 0.86$—a within-subject generation effect. The main effect of generation task was not significant on hits, $F < 1$, but was significant on *d′*, $F(1, 46) = 5.71$, MSE = 0.63, $\eta_p^2 = 0.11$; the latter difference reflected more encoding of memory information for list items during relational than item-specific generation (2.45 vs. 2.06). The within-subject generation effect was similar for both within groups; the interaction was not significant in either ANOVA, *F*s < 1.

We next compared correct recognition of *read items* in each within group to the pure-read group, to test whether read lists suffered a cost when interleaved with generate lists. Indeed, there was a cost to recognition of read items in the within-item-specific group relative to the pure-read group for hits (0.60 vs. 0.74), $t(46) = 3.86$, SEM = 0.04, $d = 1.14$, and for *d′* (1.53 vs. 2.11), $t(46) = 3.78$, SEM = 0.15, $d = 1.11$. This robust cost to non-distinctively processed items in a distinctive processing context replicates previous work (e.g. Begg & Snider, 1987; Bertsch et al., 2007; Bodner et al., 2014; Huff et al., 2015; McDaniel & Einstein, 1986). We also observed a cost to read items in the within-relational group relative to the pure-read group for hits (0.65 vs. 0.74), $t(46) = 2.12$, SEM = 0.04, $d = 0.63$, but it was not significant for *d′* (1.96 vs. 2.11), $t(46) = 1.02$, *SEM* = 0.15, $p = 0.31$.

We next compared correct recognition for *generate items* in each within group to the corresponding pure-list group from Huff and Bodner (2013). For the within-item-specific generation group, generate items showed a benefit relative to the pure item-specific generation group for hits (0.90 vs. 0.84), $t(52) = 2.20$, SEM = 0.03, $d = 0.61$, but not in *d′* (2.59 vs. 2.73), $t < 1$. The same was true when comparing generate items in the within-relational generation group relative to the pure relational generation group—the benefit was reliable for hits (0.91 vs. 0.84), $t(52) = 2.47$, SEM = 0.03, $d = 0.69$, but not *d′* (2.93 vs. 2.79), $t < 1$.

The presence of within-group costs to read items raises the question of whether these groups showed a net increase in correct recognition. Did performing item-specific or relational generation on some lists improve correct recognition, overall? To examine this novel question, net recognition in the within groups was calculated by averaging correct recognition across the read and generate lists (see Avg. columns in Table 1). One-way ANOVAs revealed significant differences across the 5 groups for both hits, $F(4, 127) = 5.10$, *MSE* = 0.04, $\eta_p^2 = 0.14$, and *d′*, $F(4, 127) = 9.98$, MSE = 0.31, $\eta_p^2 = 0.24$. To follow-up these differences, we compared recognition in each within group, in turn, to the pure-read group and corresponding pure-generate group.

For the within-item-specific generation group, overall correct recognition was similar to the pure-read group, both for hits (0.75 vs. 0.74) and *d′* (2.06 vs. 2.11), *t*s < 1, and was lower than the pure item-specific generation group, both for hits (0.75 vs. 0.84), $t(52) = 3.18$, SEM = 0.03, $d = 0.88$, and *d′* (2.06 vs. 2.73), $t(52) = 4.20$, SEM = 0.16, $d = 1.16$. Thus, there was no net improvement in correct recognition to performing item-specific generation on a subset of lists relative to a standard DRM baseline condition. Moreover, net correct recognition in the item-specific within group was much lower than in the pure item-specific generation group.

Turning to the within-relational generation group, net correct recognition was similar to the pure-read group for hits (0.78 vs. 0.74), $t(46) = 1.28$, SEM = 0.03, $p = 0.21$, but was greater than the pure-read group for *d′* (2.45 vs. 2.11), $t(46) = 2.34$, SEM = 0.14, $d = 0.69$. The within-relational group's hit rate was marginally lower than the pure relational generation group (0.78 vs. 0.84), $t(52) = 1.90$, SEM = 0.03, $p = 0.06$, $d = 0.53$, and their *d′* was significantly lower (2.45 vs. 2.79), $t(52) = 2.20$, SEM = 0.15, $d = 0.61$. Therefore, although performing relational generation on a subset of lists increased memory information for list items relative to a pure-read group, correct recognition was worse than when all lists received relational generation.

Finally, we measured the impact of the within-group design on the overall level of memory monitoring for list items using our signal-detection index, λ (computed as the standardized false alarm rate to list item controls). Monitoring at test differed across the five groups, $F(4, 127) = 3.66$, MSE = 0.20, $\eta_p^2 = 0.10$. Monitoring was lower in the within-item-specific group than the within-relational group (1.27 vs. 1.54), $t(46) = 2.00$, SEM = 0.14, $d = 0.59$. Monitoring in the within-item-specific generation group was similar to the pure-read group (1.27 vs. 1.44), $t(46) = 1.37$, SEM = 0.13, $p = 0.18$, but lower than in the pure item-specific generation group (1.27 vs. 1.67), $t(52) = 3.24$, SEM = 0.12, $d = 0.90$. In contrast, monitoring in the within-relational generation group was similar to both the pure-read group (1.54 vs. 1.44), $t < 1$, and the pure-relational generation group (1.54 vs. 1.65), $t < 1$. In sum, monitoring in the within groups was similar to the pure-read group, and was lower in the item-specific within group than in the pure item-specific generation group.

False recognition

Huff and Bodner (2013) compared their pure-generation groups to a pure-read group, so we do not repeat their analyses here. In brief, they found that pure item-specific generation reduced the DRM illusion, but pure relational generation did not. The reduction was due to less information having been encoded for critical lures following item-specific generation than in the pure-read condition, as assessed by the *d′* index. Both generate groups showed increased memory monitoring, as assessed by the λ index, thus monitoring differences did not drive the dissociation. Here, we focus on the effects of performing item-specific or relational generation on half of the lists.

To begin, the within-subject generation effect on critical lure false recognition was evaluated using separate 2(list type: generate vs. read) × 2(generation task: item-specific vs. relational) mixed ANOVAs on (a) critical lure false alarms and (b) critical lure *d′* (in which critical lure false alarms were treated as *hits* and critical lure controls as false alarms). Contrary to our expectations, the list type main effect reflected *greater* false recognition for generate than read lists (0.67 vs. 0.55), $F(1, 46) = 6.54$, MSE = 0.01, $\eta_p^2 = 0.13$, and *greater* critical lure *d′* for generate than read lists (1.59 vs. 1.29), $F(1, 46) = 5.99$, MSE = 0.36, $\eta_p^2 = 0.12$. This pattern was surprising given that false recognition is typically low and equivalent between distinctive and non-distinctive lists in a within group (Dodson & Schacter, 2001; Gunter et al., 2007; McCabe & Smith, 2006; Schacter et al., 1999). This pattern fits with neither the distinctiveness heuristic account (which predicts low and equivalent DRM effects for both list types) or the impoverished relational encoding account (which predicts a smaller DRM effect for generate lists). Below we further examine the basis of this novel finding. As expected, relational generation resulted in more false recognition than item-specific generation (0.69 vs. 0.53), $F(1, 46) = 6.73$, MSE = 0.09, $\eta_p^2 = 0.13$, and was associated with greater critical lure *d′* (1.63 vs. 1.25), $F(1, 46) = 4.99$, MSE = 0.67, $\eta_p^2 = 0.10$. The interactions were not significant, $F\text{s} < 1$.

We next compared false recognition for *read lists* across groups. False recognition for read lists differed across the three groups, $F(2, 69) = 3.69$, MSE = 0.07, $\eta_p^2 = 0.10$, though the difference in critical lure *d′* was marginal, $F(2, 69) = 2.54$, MSE = 0.53, $p = 0.09$, $\eta_p^2 = 0.07$. False recognition for read lists was lower in the item-specific versus relational within group (0.46 vs. 0.64), $t(46) = 2.70$, SEM = 0.07, $d = 0.67$, as was *d′* (1.07 vs. 1.52), $t(46) = 2.06$, SEM = 0.14, $d = 0.61$. False recognition for read lists was also lower in the within-item-specific generation group than in the pure-read group (0.46 vs. 0.62), $t(46) = 2.50$, SEM = 0.07, $d = 0.65$—a within-subject benefit for read lists (i.e. decreased false recognition)—but the *d′* difference did not reach significance (1.07 vs. 1.42), $t(46) = 1.61$, SEM = 0.14, $p = 0.11$. False recognition and critical lure *d′* were equivalent for the within-relational generation and pure-read groups (0.64 vs. 0.62; 1.52 vs. 1.42), $t\text{s} < 1$. Thus, the item-specific processing applied to the generate lists appears to have spilled over to the read lists, resulting in reduced false recognition for the read lists—a beneficial spillover effect.

We next compared false recognition for *generate lists* in the within groups to the corresponding pure-generation group from Huff and Bodner (2013). For item-specific generation, false recognition for generate lists was far greater in the within group than the pure group (0.60 vs. 0.39), $t(52) = 3.27$, SEM = 0.06, $d = 0.91$, a pattern also found in *d′* (1.44 vs. 1.03), $t(52) = 2.34$, $d = 0.65$—a cost (i.e. increased false recognition) to generate lists in the within group. For relational generation, false recognition of critical lures was also greater in the within group relative to the pure group (0.73 vs. 0.54), $t(52) = 2.61$, SEM = 0.07, $d = 0.72$, but this pattern was non-significant in *d′* (1.74 vs. 1.61), $t < 1$. These costs suggest that the relational processing used for read lists spilled over to the generate lists, resulting in increased false recognition.

We next examined whether there was a net difference in false recognition in the within groups, averaged across read and generate lists, relative to the pure-list groups. One-way ANOVAs revealed significant differences in false recognition, $F(4, 127) = 6.41$, MSE = 0.05, $\eta_p^2 = 0.17$, and critical lure *d′*, $F(4, 127) = 3.81$, MSE = 0.47, $\eta_p^2 = 0.11$. Net false recognition in the within-item-specific generation group was similar to the pure-read group (0.53 vs. 0.62), $t(46) = 1.44$, SEM = 0.06, $p = 0.16$, and *d′* (1.25 vs. 1.42), $t < 1$, and was greater than in the pure item-specific generation group (0.53 vs. 0.39), $t(52) = 2.30$, SEM = 0.06, $d = 0.67$, though not for *d′* (1.25 vs. 1.03), $t(52) = 1.28$, SEM = 0.18, $p = 0.21$. Net false recognition in the within-relational generation group was also similar to the pure-read group (0.69 vs. 0.62), $t(46) = 1.09$, SEM = 0.06, $p = 0.28$; and critical lure *d′* (1.63 vs. 1.42), $t(46) = 1.25$, SEM = 0.17, $p = 0.22$, and was also greater than in the pure relational generation group (0.69 vs. 0.54), $t(52) = 2.18$, SEM = 0.14, $d = 0.60$, though again not for *d′* (1.63 vs. 1.61), $t < 1$. In sum, there was no net decrease in false recognition when item-specific or relational generation was performed on half the lists relative to the pure-read group. Instead, the within design increased false recognition relative to pure generation (though it did not specifically increase encoding of critical lures in the *d′* measure).

In prior work, pure item-specific and pure relational generation increased memory monitoring for critical lures relative to a pure-read group (Huff & Bodner, 2013; Huff et al., 2015). Here, a one-way ANOVA indicated that monitoring for critical lures differed across the five groups, $F(4, 127) = 3.65$, $MSE = 0.22$, $\eta_p^2 = 0.10$. Monitoring was similar in the item-specific and relational within groups (1.18 vs. 1.13), $t < 1$. Monitoring in the within-item-specific generation group was similar to both the pure-read group (1.18 vs. 1.05), $t < 1$, and the pure item-specific generation group (1.18 vs. 1.34), $t(52) = 1.18$, $SEM = 0.14$, $p = 0.25$. Monitoring in the within-relational generation group was similar to the pure-read group (1.13 vs. 1.05), $t < 1$, but was reduced relative to the pure relational generation group (1.13 vs. 1.47), $t(52) = 2.99$, $SEM = 0.11$, $d = 0.83$. In short, monitoring for critical lures did not increase when generation was performed for only half of the lists, unlike when generation occurred for all lists (Huff & Bodner).

Discussion

To parse the effects of encoding manipulations, memory researchers have often compared them across within- and between-subject designs. In the DRM literature, effects in these two designs have largely been examined in isolation (e.g. Arndt & Reder, 2003; Dodson & Schacter, 2001; Gunter et al., 2007; McCabe & Smith, 2006; Schacter et al., 1999), unlike in other areas of memory research (e.g. Begg & Snider, 1987; Bodner et al., 2014; McDaniel & Einstein, 1986). Huff et al.'s (2015) meta-analysis of distinctive encoding across within- and between-subject DRM experiments revealed that the within groups showed a mixture of benefits and costs. Our study confirmed and extended their findings in several ways. First, we identified spillover effects from each type of list to the other type of list. Second, following Huff and Bodner (2013), we also explored the effects of relational processing of a subset of lists. And third, we introduced measures of net recognition, which revealed little evidence of overall increased memory accuracy in the within groups relative to a pure-read group. As reviewed below, generation of half the lists did little to improve overall recognition accuracy—and indeed, it resulted in recognition costs and spillover effects.

For correct recognition, performing item-specific or relational generation for half the lists both produced within-subject generation effects. This effect was driven in part by both a within-subject cost to recognition of read list items and a within-subject benefit to recognition of generate list items. Among other possibilities, the read-item cost could be due to the within-subject groups performing less relational processing (or due to them performing more item-specific processing of the read lists), paying less attention or giving less priority to the read lists (see Begg & Snider, 1987), or suffering a switch cost in shifting from generate lists to read lists. The generate-item benefit likely reflects generate items being rendered more distinctive in memory in a context of non-distinctive items than in a pure list of distinctive items (Hunt & Worthen, 2006). Another possibility is that this benefit resulted from a spillover of relational processing from the read lists to the generate lists. Since the within-relational generation group would already be using relational processing for the generate lists, making any spillover of relational processing redundant, they should have shown a reduced generation benefit. Contrary to that possibility, the benefit for generate items was very similar in the item-specific and within-relational groups (0.06 vs. 0.07).

The costs (to read items) and benefits (to generate items) in each within-subject group resulted in net correct recognition that was generally similar to the pure-read group, but worse than in the pure-generate groups. Distinctiveness effects such as generation (Bertsch et al., 2007), production (Fawcett, 2013; Forrin, Groot, & MacLeod, 2016), and bizarre imagery (McDaniel & Einstein, 1986) are often larger (or only present) in within- versus between-subject designs. Our net recognition analyses suggest that researchers need to assess whether larger effects of encoding manipulations in within groups reflect costs (see Bodner et al., 2014).

It is important for researchers to evaluate whether larger within-subject benefits come with correspondingly larger within-subject costs. As an example, McDaniel and Einstein (1986, Experiment 1) found a recall advantage of bizarre encoding over common encoding in a within-subject design (0.62 vs. 0.38), but not in a pure-list design (0.52 vs. 0.55). Though there was a 0.10 benefit to bizarre encoding in the within design ($0.62 - 0.52 = 0.10$), there was a corresponding 0.17 cost to common encoding in the within design ($0.38 - 0.55 = - 0.17$). Applying a net memory analysis, these benefits and costs resulted in recall (averaged across item type) being no greater in the within group (0.50) than in the pure-bizarre group (0.52), or, more strikingly, than in the pure-common group (0.55). Thus, performing bizarre encoding on a subset of items did not provide an *overall* mnemonic benefit—consistent with what we reported here for generation in the DRM paradigm.

Turning to false recognition, based on our prior work, we expected that item-specific generation for half the lists would reduce the DRM illusion, whereas relational generation would not (Huff & Bodner, 2013; Huff et al., 2015). In keeping with that expectation, the DRM illusion was indeed lower in the within-item-specific group than in the within-relational group when averaged across read and generate lists overall. However, false recognition was greater for the generate lists than for the read lists in both groups. Indeed, false recognition for within-generate lists was much higher than in the pure item-specific generation group, whereas false recognition for the within-read lists was much lower than in the pure-read group. This pattern suggests that the processing applied to each list type was influenced by the processing applied to the other list type. Interleaving generate and read list types across the ten study lists may have fostered spillover effects; indeed, future studies could examine whether blocking the two sets of lists reduces these effects (see Bodner et al., 2014). It appears that relational processing of read lists spilled over to generate lists (increasing false recognition for generate lists) while item-specific processing of generate lists spilled over to read lists (decreasing false recognition for read lists). This pattern shows that performing distinctive item-specific processing for a subset of lists can backfire, resulting in increased false recognition (cf. Huff & Bodner, 2013).

In the within-relational generation group, false recognition for generate lists was again higher than for read lists in this group, as well as relative to the pure relational generation group—establishing the replicability of this counter-intuitive effect. False recognition for the read lists in this within group was very similar to the pure-read group. This similarity is not surprising, given that relational processing would normally be used to encode related words that form each DRM list. Given that both tasks in the within-relational group emphasized relational processing, spillover effects were, therefore, minimal. However, the cost to generate lists in the relational generation group is again indicative of a read-to-generate spillover effect.

Our findings suggest that spillover effects are more likely to influence recognition when the two list types in a within group differ in the type of processing they require (item-specific vs. relational) than when they induce the same type of processing (relational vs. relational). The effects of performing two different item-specific tasks across lists in a within-subject condition remain an interesting variant to be explored. Conceivably, if the two tasks engender unique forms of item-specific processing, they might induce spillover effects that amplify the reduction in false memory relative to pure-list groups. We are currently pursuing this possibility.

Just as there was little evidence that performing distinctive item-specific generation for half the lists increased correct recognition, there was little evidence that it reduced false recognition relative to a pure-read group. On the contrary, net false recognition in the within groups was higher than in the pure-list generate groups. In the within-item-specific generation group, spillover effects resulted in no net reduction in false recognition over reading all the lists. These findings highlight the need for measuring the net effects of within-subject manipulations of encoding tasks on false memory.

Huff and Bodner (2013) found that pure item-specific and pure relational generation groups showed increased memory monitoring at test relative to a pure-read group. In contrast, here, when generation lists were interleaved with read lists at study, monitoring for both list items and critical lures was similar to the pure-read group. One possibility for this similarity is that participants adapted to the presence of the read lists by globally relaxing their monitoring constraints during the test. Consistent with this possibility, Huff et al.'s (2015) meta-analysis found lower monitoring estimates in within- vs. between-subject designs.

However, we recommend that monitoring estimates for within-subject conditions be treated with caution, given that each group provides only a single index of monitoring, based on control item false recognition. These point estimates of monitoring cannot distinguish contributions of separate retrieval operations, such as early selection versus late correction processes (cf. Hunt, Smith, & Dunlap, 2011). Participants in a within group are likely able to infer whether a given test item had appeared on a generate or read list (i.e. given that all the items on a given list were thematically related and studied using the same task). As a result, participants could potentially monitor differentially for critical lures from generate versus read lists at test. Our single index of monitoring cannot detect such differences. Given the indications of spillover effects in the within groups, the global level of memory monitoring captured by this index will be influenced by spillover effects. Obtaining separate indices of monitoring for each list type in a within-subject design remains an important step for the development of the signal-detection approach to analysing recognition memory in the DRM paradigm.

Our study provides novel evidence that within-subject designs do not invariably lead to low and equivalent DRM illusions for each list type (cf. Dodson & Schacter, 2001; Gunter et al., 2007; McCabe & Smith, 2006; Schacter et al., 1999; cf. Arndt & Reder, 2003). This equivalence has been taken as evidence that participants apply a distinctiveness heuristic at test, in which they only endorse critical lures if they remember performing a distinctive encoding task on them. Because critical lures from both list types were not studied, participants reject critical lures from both list types. At least one notable exception to this pattern was reported by Arndt and Reder, who found lower false recognition for lists presented in distinctive fonts. Their pattern supported the claim that distinctive item-specific encoding led to impoverished relational encoding of the distinctive lists, thus reducing the DRM illusion for those lists. Here, we found the opposite pattern, namely *higher* false recognition for generate lists—which we have attributed to a spillover of relational processing from read lists. This pattern challenges both the distinctiveness heuristic and impoverished relational encoding accounts.

A remaining question is why distinctive and non-distinctive tasks produce equivalent false recognition rates under some conditions, and differences that favour either the distinctive or non-distinctive list type under other conditions.

One relevant factor may be the extent of item-specific processing induced by the distinctive task. As we have noted (Huff & Bodner, 2013), item-specific and relational task variants are not process pure, thus some variants may be more successful in inducing a given type of processing than others. For instance, our item-specific generation task may have encouraged more item-specific processing than reading words aloud (Dodson & Schacter, 2001), solving pig-Latin anagrams which involved a single rule (McCabe & Smith, 2006), or looking at a picture of each DRM list word (Schacter et al., 1999). Measuring task-level differences in item-specific or relational processing was outside the scope of this study, but in work outside the DRM paradigm, we have explored the utility of categorical clustering scores toward this aim (see Huff & Bodner, 2014).

Conclusion

Our results show that encoding tasks can have different effects on correct and false recognition, depending on whether that processing is item-specific vs. relational in nature, and depending on whether it is applied to all or a subset of study materials. Whenever encoding strategies are varied within subjects, a complex pattern of benefits, costs, and spillover effects can result. To gauge these effects, researchers need to compare each within-group condition, as well as their net effects, to pure-list counterparts.

Compliance with ethical standards

Conflict of interest The authors report no competing interests.

Ethical standards The studies reported were approved by the University of Calgary Research Ethics Board (Protocol #6684) and found to be in accordance with the 1964 Helsinki Declaration ethical principles.

Informed consent Informed consent was obtained from all individuals who participated in this study.

References

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language, 48*(1), 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*(4), 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*(2), 201–210. https://doi.org/10.3758/BF03193441.

Bodner, G. E., Huff, M. J., Lamontagne, R. W., & Azad, T. (2017). Getting at the source of distinctive encoding effects in the DRM paradigm: evidence from signal-detection measures and source judgments. *Memory, 25,* 642–655. https://doi.org/10.1080/09658211.2016.1205094.

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review, 21*(1), 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*(5), 164–169. https://doi.org/10.1111/1467-8721.00192.

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*(1), 155–161. https://doi.org/10.3758/BF03196152.

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica, 142,* 1–5. https://doi.org/10.1016/j.actpsy.2012.10.001.

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*(9), 1119–1141. https://doi.org/10.1002/acp.1265.

Forrin, N. D., Groot, B., & MacLeod, C. M. (2016). The d-prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42,* 1090–1111. https://doi.org/10.1037/xlm0000214.

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology. Learning, Memory & Cognition, 30*(1), 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*. New York: Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*(7), 833–848. https://doi.org/10.3758/MC.38.7.833.

Gunter, R. W., Bodner, G. E., & Azad, T. (2007). Generation and mnemonic encoding induce a mirror effect in the DRM paradigm. *Memory & Cognition, 35*(5), 1083–1092. https://doi.org/10.3758/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 & Cognition, 30*(4), 787–795. https://doi.org/10.1037/0278-7393.30.4.787.

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/10.1177/1747021818801427.

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*(2), 349–365. https://doi.org/10.3758/s13423-014-0648-8.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). Oxford: Oxford University Press.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 497–514. 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.

Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language, 65,* 378–389. https://doi.org/10.1016/j.jml.2011.06.003.

Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.001.0001.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4*(4), 577–581. https://doi.org/10.3758/BF03214352.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review, 11*(6), 1074–1079. https://doi.org/10.3758/BF03196739.

McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory, 14*(5), 570–583. https://doi.org/10.1080/09658210600624564.

McDaniel, M., & Einstein, G. (1986). Bizarre imagery as an effective memory aid. The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 54–65. https://doi.org/10.1037/0278-7393.12.1.54.

Oliver, M. C., Bays, R. B., & Zabrucky, K. M. (2016). False memories and the DRM paradigm: Effects of imagery, list, and test type. *The Journal of General Psychology, 143*(1), 33–48. https://doi.org/10.1080/0022139.2015.1110558.

Roediger, H. L., III, Balota, D., & Watson, J. (2001a). Spreading activation and arousal of false memories. In H. L. I. I. I. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington DC: 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*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803.

Roediger, H. L., III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8,* 385–407. https://doi.org/10.3758/BF03196177.

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.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*(6), 523–542. https://doi.org/10.3758/BF03197149.

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.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5,* 710–715. https://doi.org/10.3758/BF03208850.

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