Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Multiple species of distinctiveness in memory? Comparing encoding versus statistical distinctiveness on recognition

Matthew R. Gretz & Mark J. Huff

To cite this article: Matthew R. Gretz & Mark J. Huff (2020): Multiple species of distinctiveness in memory? Comparing encoding versus statistical distinctiveness on recognition, Memory, DOI: 10.1080/09658211.2020.1803916

To link to this article: https://doi.org/10.1080/09658211.2020.1803916

Published online: 07 Sep 2020.

MEMORY
https://doi.org/10.1080/09658211.2020.1803916

Multiple species of distinctiveness in memory? Comparing encoding versus statistical distinctiveness on recognition

Matthew R. Gretz and Mark J. Huff

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

ABSTRACT
The distinctiveness effect refers to the memorial benefit of processing unique or item-specific features of a memory set relative to a non-distinctive control. Traditional distinctiveness effects are accounted for based on qualitative differences in how distinctive items are encoded and subsequently retrieved. This study evaluates whether a separate species of distinctiveness – statistical distinctiveness – may provide an additional benefit to memory beyond traditional task-based processes. Statistical distinctiveness refers to the relative frequency with which a specific memory item or set is processed. The current study examined the presence of statistical distinctiveness through a series of levels-of-processing mixed groups in which related lists were studied using two of the following three tasks to promote either shallow ("E" identification), neutral (reading silently), or deep/distinctive (pleasantness ratings) processing followed by a recognition test. Participants studied lists in which these tasks were used frequently (80% of lists), equally (50% of lists), or infrequently (20% of lists). No recognition advantage was found when tasks were completed infrequently versus frequently. Instead, recognition was greatest for the deeper/more distinctive task – a pattern consistent with an encoding but not a statistical distinctiveness account.

ARTICLE HISTORY
Received 24 December 2019
Accepted 24 July 2020

KEYWORDS
Distinctiveness; encoding distinctiveness; statistical distinctiveness; levels-of-processing; recognition

Distinctive information often receives a memorial benefit relative to information that is non-distinctive, a pattern referred to as the *distinctiveness effect*. Classically, the distinctiveness effect occurs when a specific group of items lies in contrast to an established cohesive context (Hunt, 2006; Surprenant & Neath, 2009). This context can be established perceptually, such as the presentation of a red-colored item in a context of blue-colored items, or conceptually, such as a number embedded in a string of letters. In both cases, items that violate a prevailing context are better remembered. Although distinctiveness effects have been broadly demonstrated over a variety of materials (see Burns, 2006; Hunt & Worthen, 2006, for reviews), a critical question is whether memorial benefits elicited by distinctiveness reflect a simple contrast between an event and the context in which it occurs, or whether a *degree* of contrast is required to produce memory benefits. The purpose of this study is to provide a comparison of two possible types of distinctiveness effects in recognition and gauge their potential memory benefits.

Task-based distinctiveness in memory

Distinctiveness effects can occur at both the item level and at the task level, such as when a study task encourages the encoding of distinctive/item-specific features (Hunt, 2006; Hunt & Lamb, 2001; Smith & Hunt, 1998). One such example is the generation effect or the memory advantage for study materials that are self-generated versus materials that are merely provided. In an early demonstration, Slamecka and Graf (1978) presented participants with related cue-target pairs for study that were either incomplete (e.g., rapid-f___), in which participants had to generate a related target word, or intact (e.g., rapid-fast). At test, recognition was greater following the study of generated than intact pairs. The act of generating words was argued to differentiate words in memory from intact words, making them more memorable. Generation has been argued to promote the processing of item-specific information, or the processing of distinctive or unique features of generated items (Huff & Bodner, 2013; Hunt & Einstein, 1981; McDaniel & Waddill, 1988).

Similar memory benefits have been reported with other distinctive-type tasks. For instance, producing (i.e., reading) items aloud (MacLeod et al., 2010; see too Gathercole & Conway, 1988 for review), viewing images of referents of to-be-remembered words (Israel & Schacter, 1997), rating words based on their relative pleasantness (Craik & Lockhart, 1972; Huff & Bodner, 2013), and rating words on their fitness relevance (Nairne et al., 2007), have all yielded distinctiveness-type memory benefits relative to control conditions. Thus, distinctiveness benefits occur using a wide array of tasks.

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

Though distinctive tasks are diverse, the magnitude of their effects appear to be sensitive to between- versus within-subject designs. Begg and Snider (1987) reported that the generation effect was diminished when generation versus reading was completed between- versus within-subjects – a pattern echoed by MacLeod et al. (2010) who reported a reduced production effect for pure lists, in which all study items were read aloud or silently compared to mixed lists. Further, Bodner and Taikh (2012) reported that the production effect was eliminated when a between-subject design was employed. Meta-analyses have since confirmed that generation and production do indeed produce memory benefits under between-subject conditions – though at considerably diminished rates. Relatively small effect sizes have been reported for between-subject generation ($g = 0.28$) and production ($g = 0.37$), which are approximately half the size found in within-subject designs (Bertsch et al., 2007; Fawcett, 2013, respectively). More moderate within-subject effects have been attributed to greater emphasis placed on distinctive versus non-distinctive items, due to an available comparison between the two item types. This comparison provides enhanced discrimination that is not available in a between-subject design when participants are exposed to a single item type. Another explanation, evaluated in this study, is that distinctive items are also statistically less frequent in within- than between-subject designs as they are typically only half of the studied items. The relative proportion of distinctive to non-distinctive items may be related to the magnitude of distinctiveness benefits. Thus, the statistical rarity of distinctive items or tasks appears to be related to the presence and magnitude of distinctiveness effects.

Statistical distinctiveness in memory

Given robust within-subject generation and production effects, a critical question is whether the relative *proportions* of distinctive generate/aloud items versus non-distinctive items are related to the magnitude of the memory benefit. Icht et al. (2014) argued that distinctiveness originates from two sources: Distinctiveness due to encoding processes at the task level, and distinctiveness due to the statistical frequency in which a task is utilised. Encoding distinctiveness refers to a specific mode of processing that can qualitatively affect encoding processes at study or monitoring processes at test, such as tasks that promote item-specific processing (Huff et al., 2015; Hunt & Einstein, 1981). In contrast, statistical distinctiveness refers to the relative distribution of distinctive versus non-distinctive information, a distribution that varies when comparing between- and within-subject designs. Importantly, Icht et al. argued that these two types of distinctiveness are not mutually exclusive: Distinctive tasks that would normally yield a memory improvement through distinctive processing may be ineffective if used frequently or even exaggerated if used infrequently.

To evaluate the contributions of encoding and statistical distinctiveness, Icht et al. (2014) used a production-effect paradigm in which proportions of study items that were read aloud versus read silently varied in mixed lists. Specifically, of the total number of list items, 20%, 50%, or 80% were read aloud with the remaining read silently. According to the encoding distinctiveness account, aloud items should be better remembered than silent items regardless of the frequency in which aloud versus silent items were studied due to distinctive processing of aloud items. In contrast, the statistical distinctiveness account affirms that items studied using a task that is employed less frequently should be better remembered than items studied using a frequent task regardless of task type. In Experiment 1, the authors reported a reliable production effect overall in recall – a pattern consistent with encoding distinctiveness – but the magnitude of the production effect interacted with the production proportion. Specifically, the production effect was greatest in the 20% condition in which aloud items were less frequent, but declined across the 50% and 80% conditions. Indeed, the 80% aloud condition produced a *reversed production effect*, such that correct recall was greater for silent than aloud items. Thus, production benefits were diminished and even reversed, when aloud items were more frequent. In a second experiment, which used recognition, a similar pattern was found though the reversed production effect in the 80% aloud condition was not reliable (though the production effect was eliminated), providing additional evidence that production is sensitive to the frequency of task presentation. These findings suggest that the relative task frequency affects later recall and recognition for items that were studied using a distinctive and non-distinctive study tasks.

Similar statistical distinctiveness patterns have been reported by Bodner et al. (2016; Experiment 2). In their experiment, proportions of items that were produced (via typing on a keyboard) versus unproduced (read silently) were completed by either pure groups (i.e., 0% or 100% typed) or mixed groups (20%, 50%, or 80% typed). As a means of equating encoding for production versus non-production words, encoding duration was three times longer for non-production than production words. On a final recognition test, a between-subject production effect was absent (possibly due to differences in the encoding duration), as was a within-subject production effect in the 50% and 80% production groups. However, a production effect was found in the 20% production group when typed items were statistically rare and despite non-production items receiving considerably more encoding time. Therefore, at least with production, the relative frequency with which production is completed appears to moderate the memory benefit.

Although statistical distinctiveness benefits are often interpreted as infrequent items procuring additional processing at study, it is also possible that frequent items may be contributing to the pattern due to processes such as cue overload (Surprenant & Neath, 2009, for review; Parkin, 1980). The cue-overload principle states that cues established at encoding are less beneficial at retrieval when the same cue is employed for many versus few items. For instance, Keppel and Underwood (1962) reported that recall of word lists that are taken from the same category decreased across successive study/test trials – a pattern of proactive interference. The authors reasoned that interference effects may reflect an overuse of the category retrieval cue across lists, minimising the cue effectiveness. The benefits of a distinctive study task may similarly yield diminishing returns due to the inability of repeated tasks to provide distinguishable retrieval cues at test, as in Icht et al. (2014) and Bodner et al. (2016). Of course, both processes may operate concurrently as infrequent items may benefit from enhanced encoding while frequent items may fall victim to cue overload.

Distinctive processing on memory errors

Despite the vast benefits of distinctive processing on correct memory, it is equally important to evaluate its effects on overall memory accuracy, which includes memory errors. A common method for evaluating memory errors is through the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In this paradigm, participants study lists of associates (e.g., *bed, rest, tired, dream, slumber,* etc.) which all converge upon a single, non-presented critical lure (e.g., *sleep*). At test, participants often incorrectly remember that the critical lure was studied, a phenomenon termed the *DRM illusion*. The DRM illusion is robust: False recall often approaches recall rates of studied items from the middle serial list position, and false recognition has been shown to meet or exceed hit rates (see Gallo, 2006; 2010, for reviews; Lampinen et al., 1999). Given these elevated rates, researchers have found several methods in order to reduce (but not eliminate) the illusion including increasing study repetitions (Benjamin, 2001; McDermott, 1996), and by presenting warnings about the DRM illusion prior to study and/or test (Gallo et al., 2001; McCabe & Smith, 2002).

Relevant to the present study, reduction in the DRM illusion has also been found following the processing of distinctive features of list items at study. In one example, Israel and Schacter (1997) presented participants with DRM lists that were either presented as words in isolation, or as words that were accompanied by a picture of each word's referent. At test, false recall of critical lures was reduced when lists were studied with pictures versus lists studied as words, presumably because pictures provide distinctive retrieval cues for each word that can assist test-based monitoring. A similar reduction has been found using a variety of other distinctive manipulations including studying lists of words in a unique (vs. same) font type (Arndt & Reder, 2003), generating words from anagrams (Gunter et al., 2007; McCabe & Smith, 2006), creating mental images of list words (Foley et al., 2006; Oliver et al., 2016; Robin, 2010), and critically, processing the unique or distinctive features of study words via item-specific processing (Huff & Bodner, 2013; McCabe et al., 2004; Smith & Hunt, 1998). These benefits are particularly noteworthy because they often induce a mirror effect benefit to overall memory accuracy (Glanzer & Adams, 1990) – an increase in correct memory coupled with a decrease in false memory versus a nondistinctive or processing-neutral task.

Reductions in the DRM illusion can be attributed to the encoding strategies used at study and/or the monitoring strategies used at test. Tasks that require item-specific processing disrupt the thematic consistency or associative strength between items, resulting in impoverished relational encoding (Hege & Dodson, 2004). Alternatively, participants may employ a distinctiveness heuristic – a test-based monitoring strategy in which recollections of distinctive details at the time of test can serve as diagnostic evidence that a studied item was either studied or not (Gallo, 2004; 2010). Since DRM critical lures are not paired with distinctive details because they were not studied, the distinctiveness heuristic aids in screening critical lures from being reported. Impoverished relational encoding and the distinctiveness heuristic have been argued to be competing explanations for the reduction in the DRM illusion following distinctive processing (e.g., Dodson & Schacter, 2001; Hege & Dodson). However, evidence has shown that both can operate in tandem (Huff & Aschenbrenner, 2018; Huff & Bodner, 2013).

The current study

Given the benefits of distinctive processing at the task level on increasing correct recognition and reducing the DRM illusion, the purpose of the current study was to examine whether statistical distinctiveness may show comparable benefits on overall memory accuracy. A standard task that has been shown to successfully promote distinctive processing is pleasantness ratings, which is also considered a deep-processing task according to the levels-of-processing framework (LOP; Craik & Lockhart, 1972). Pleasantness ratings have been shown to simultaneously increase correct recognition and decrease false recognition in the DRM paradigm (Huff & Bodner, 2013; Hunt et al., 2011). However, it remains unclear whether distinctive processing tasks may be particularly potent when completed relatively infrequently.

The current study used four experiments to gauge the effects of task and statistical distinctiveness on recognition in the DRM paradigm. In Experiments 1–3, we focused on statistical distinctiveness effects by having participant study subsets of lists that were either encoded with a distinctive or non-distinctive task using a set of *mixed groups*. In accord with previous literature (Bodner et al., 2016; Icht et al., 2014), participants studied a set of 10 DRM lists, in which either 20%, 50%, or 80% of those lists were studied using a distinctive task with the remaining lists using a non-distinctive task. In Experiment 1, we compared pleasantness ratings (PR), a standard distinctive/deep-encoding task to silent reading. In Experiment 2, we compared pleasantness ratings to an "e" letter detection task (E-Task), a standard non-distinctive/shallow-processing task. In Experiment 3, we compared the read task to the E-Task to determine whether statistical distinctiveness effects would emerge when two weaker encoding tasks were compared. In Experiment 4, to affirm our task processing claims, we compared a set of *pure groups* in which 100% of the DRM lists were studied using either the PR, Read, or E-Tasks. Finally, we included a series of secondary cross-experimental analyses to compare correct and false recognition on mixed groups to pure groups as another test of statistical distinctiveness given pure groups complete a study task at 100% frequency. Thus, our study addressed a novel question on whether distinctive task effects on recognition accuracy were moderated by the relative frequency with which a distinctive task is deployed.

Experiment 1: PR vs. read tasks

Experiment 1 assessed the memory contributions of PR and read tasks when both were utilised at different frequencies. Correct recognition was expected to be greater overall following the PR than the read task, consistent with encoding distinctiveness. However, based on statistical distinctiveness effects reported by Bodner et al. (2016) and Icht et al. (2014), we expected that correct recognition would be greatest for 20% PR lists in which the PR task is completed less frequently relative to the 50% and 80% PR lists. A similar prediction was made for read lists in which 20% read lists were expected to produce greater correct recognition relative to 50% and 80% read lists. For false recognition, we expected PR lists to reduce false recognition overall relative to read lists (cf. Huff & Bodner, 2013), and particularly for 20% PR lists over 50% and 80% PR lists due to statistical distinctiveness, a pattern that was expected to be echoed by the read lists. Thus, we anticipated that our experiment would provide support for both encoding and statistical distinctiveness consistent with Icht et al.'s surmise that both distinctiveness types are not mutually exclusive.

Method

Participants

Seventy-three undergraduates (88% female, $M_{age} = 19.60$, $SD_{age} = 3.41$) from The University of Southern Mississippi were recruited for participation and compensated with partial course credit. All had normal or corrected-to-normal vision.

Materials

The 20 DRM lists that produced the highest rates of false recognition in Stadler, Roediger, & McDermott (1999) were used. These lists were divided into 2 sets of 10 lists and matched on backward-associative strength (BAS), a metric of association from the study lists to the critical lure which has shown to be a strong predictor of later false recognition (Roediger et al., 2001). The DRM lists in each set were once randomised and presented in the same order to all participants. Across participants, one set was studied, and the other was not and served as the control set. The studied versus non-studied set was counterbalanced across participants. Each list contained 15 total study words, that were presented in random order at study. An 80-item recognition test was then created which comprised of 30 list items (from positions 2, 8, and 10 in each studied list), 30 list item controls (taken from the same positions in non-studied lists), 10 critical lures from studied lists, and 10 critical lures from non-studied lists (see Huff & Bodner, 2013, for a similar list construction). Recognition tests were presented in a newly randomised order for each participant.

Procedure

The study was conducted using a computer running E-Prime 3.0 software (Psychology Software Tools, 2016, Pittsburg, PA). Participants were tested individually. Following informed consent, participants were randomly assigned to 1 of 3 mixed groups (Table 1). Participants were asked to use two tasks to study list items that were presented individually. In the PR task, participants were asked to rate each word for its pleasantness on a 5-point scale (1 = extremely unpleasant; 5 = extremely pleasant). In the Read task, participants were asked to read each word silently. Participants indicated their PR responses by pressing 1–5 keys on a keyboard for PR lists which advanced to the next list item and pressed the spacebar to advance to the next list item on Read lists. All participants were provided with practice examples for each task from the "fruit" DRM list, and then completed the study trials. In the 20% PR versus 80% Read group, participants studied 2 lists using the PR task and 8 lists using the Read task. In the 50% PR versus 50% Read group, participants studied 5 lists using the PR task and 5 lists using the Read task. In

Table 1. Study Conditions and Task Distributions in Mixed and Pure Groups in Experiments 1–4.

Experiment/List Type	Experiments 1–3: Mixed Groups		
Distribution	80%/20%	50%/50%	20%/80%
Experiment 1: Deep/Neutral	PR/Read	PR/Read	PR/Read
Experiment 2: Deep/Shallow	PR/E-Task	PR/E-Task	PR/E-Task
Experiment 3: Neutral/Shallow	Read/E-Task	Read/E-Task	Read/E-Task
	Experiment 4: Pure Groups		
Proportion	100%	100%	100%
List Type	PR	Read	E-Task

the 80% PR versus 20% Read group, participants studied 8 lists using the PR task and 2 lists using the read task. During study, an instruction screen appeared prior to each list to inform participants which study task they would be using for that list. In each mixed group, the task orderings for each set of 10 lists were semi-randomised to create five separate orders of study tasks which were counterbalanced across each of the two list sets. The orderings were created such that tasks were not blocked together (i.e., 2 PR lists then 8 Read lists in the 20% PR/80% Read condition).

Across conditions, lists were studied back-to-back and the recognition test was completed immediately after study of the final list. The recognition test was an old/new recognition test in which participants were instructed to press an "old" labelled key on a keyboard for each item that was studied, and a "new" labelled key for each word that was not studied. Participants were instructed to respond quickly and accurately and were not informed about the presence of critical lures on the recognition test. Following the recognition test, participants were fully debriefed regarding the experiment and were awarded participation credit.

Results

For all results reported, a $p < .05$ level of significance was used unless noted otherwise. Measures of effect size were calculated by using partial-eta squared ($\eta_p{}^2$) for analyses of variance (ANOVAs) and Cohen's d for t-tests for all significant and marginal effects. Proportions of correct and false recognition as a function of task type and proportion are reported in Table 2.[1] All nonsignificant comparisons were further tested using a Bayesian estimate of the strength of evidence supporting the null hypothesis (Masson, 2011; Wagenmakers, 2007). This analysis compares two models: One which assumes an effect and another which assumes a null effect. This Bayesian analysis yields the p-value p_{BIC} (Bayesian Information Criterion) which is an estimate for the probability that the null hypothesis is retained. This estimate is sensitive to sample size and can act as a power analysis to increase confidence in the null. Therefore, we incorporate a p_{BIC} analysis to supplement null effects from standard null-hypothesis significance testing.

Correct recognition

Statistical distinctiveness is contingent upon the interaction between list type combinations and their proportional manipulations. In order to assess statistical distinctiveness within task combinations, the first set of analyses examined individual task types when they were used to study either 20%, 50%, or 80% of the lists (i.e., comparing groups across rows in Table 2). A 2(Task Type: PR vs. Read) × 3(Proportion: 20/80 vs. 50/50 vs. 80/20) mixed ANOVA was used to examine proportions of correct

Table 2. Mean (95% CI) Proportion of "Old" Responses for Studied List Items, List Item Controls, Critical Lures, and Critical Lure Controls as a Function of Task Type and Task Proportion for Mixed Groups (Experiments 1–3) and Pure Groups (Experiment 4).

	Mixed Groups								
Task Type/Task Proportion		80/20			50/50			20/80	
Experiment 1: PR vs. Read	80% PR		20% Read	50% PR		50% Read	20% PR		80% Read
N		24			24			25	
List items	.90 (.04)		.63 (.11)	.94 (.03)		.72 (.09)	.97 (.03)		.66 (.07)
List item controls		.13 (.03)			.10 (.03)			.17 (.04)	
Critical lures	.67 (.10)		.69 (.16)	.70 (.11)		.65 (.10)	.77 (.14)		.67 (.11)
Critical lure controls		.11 (.04)			.13 (.04)			.23 (.07)	
Experiment 2: PR vs. E	80% PR		20% E	50% PR		50% E	20% PR		80% E
N		24			24			24	
List items	.89 (.03)		.53 (.11)	.95 (.02)		.68 (.08)	.89 (.06)		.63 (.07)
List item controls		.16 (.03)			.15 (.04)			.22 (.05)	
Critical lures	.73 (.10)		.38 (.16)	.71 (.11)		.51 (.12)	.69 (.14)		.52 (.09)
Critical lure controls		.24 (.07)			.25 (.07)			.36 (.08)	
Experiment 3: Read vs. E	80% Read		20% E	50% Read		50% E	20% Read		80% E
N		24			24			24	
List items	.72 (.05)		.69 (.09)	.70 (.08)		.75 (.06)	.76 (.12)		.65 (.08)
List item controls		.26 (.06)			.30 (.08)			.29 (.06)	
Critical lures	.71 (.10)		.58 (.17)	.69 (.12)		.63 (.10)	.67 (.15)		.60 (.10)
Critical lure controls		.28 (.08)			.38 (.10)			.40 (.09)	

	Pure Groups		
Task Type	PR	Read	E
Experiment 4:			
N	26	26	24
List items	.95 (.02)	.78 (.05)	.66 (.09)
List item controls	.15 (.04)	.12 (.05)	.39 (.06)
Critical lures	.69 (.08)	.77 (.08)	.65 (.09)
Critical lure controls	.17 (.06)	.20 (.05)	.43 (.09)

Notes: PR: Pleasantness Ratings; Read: Silent Reading; E: "E"-Judgment Task.

recognition of studied list items across tasks and task distributions.[2] A significant main effect of Task Type was found, $F(1, 70) = 85.52$, $MSE = .03$, $\eta_p^2 = .55$, which indicated greater correct recognition for items studied using PR than Read tasks (.94 vs. .67). However, in contrast to the statistical distinctiveness account, both the main effect of Proportion and the interaction were not significant, $F(2, 70) = 1.83$, $MSE = .03$, $p = .17$, $p_{BIC} = .92$, and $F < 1$, $p_{BIC} = .97$, respectively.

False recognition

A second 2(Task Type) × 3(Proportion) ANOVA on proportions of false recognition of critical lures yielded null effects of Task Type, $F(1, 70) = 1.10$, $MSE = .08$, $p = .30$, $p_{BIC} = .83$, Proportion, $F < 1$, $p_{BIC} = .98$, and a non-significant interaction. $F < 1$, $p_{BIC} = .97$. Thus, PR ratings failed to reduce false recognition overall, and false recognition was not sensitive to task proportions as predicted by statistical distinctiveness.

Discussion

Although no evidence was found for statistical distinctiveness, clear encoding differences between PR and reading tasks were found across proportion manipulations. Namely, correct recognition was greater for PR lists compared to read lists consistent with encoding distinctiveness. Such encoding differences did not extend to false recognition, which was equivalent across tasks and proportions. The finding that PR lists did not produce a reduction in false recognition is inconsistent with patterns found in previous studies (e.g., Huff & Bodner, 2013), though the present experiment deviates by using a within versus between-subject design. Recently, it has been shown that within-subject designs may show carry-over effects of processing on false recognition, which can eliminate differences between individual tasks (Huff et al., in press).

One possibility for the lack of statistical distinctiveness in the present study may be due to the relative differences in encoding processes between the two task types. Though PR and Read tasks often produce large recognition differences (Huff et al., 2015), task differences may need to be exaggerated to be sensitive to statistical effects. To test this possibility, in Experiment 2, we compared PR ratings to a shallow LOP E-Task. We reasoned that the magnitude of the task effects on recognition would be even greater with the PR/E-Task comparison.

Experiment 2: PR vs. E tasks

The purpose of Experiment 2 was to provide another test of statistical distinctiveness by comparing two tasks that were expected to produce even greater differences in recognition: A PR task, a deep/distinctive processing task, and "E" judgments, a shallow task. We therefore expected that correct recognition would be greater overall for lists studied using the PR task over the E-Task, consistent with encoding distinctiveness (and LOP), but would also be sensitive to the proportion of lists that are completed using these tasks. With a more powerful encoding distinctiveness manipulation across tasks, we now expected a statistical distinctiveness pattern would emerge in which correct recognition would be greatest for 20% PR lists relative to the 50% and 80% PR lists. Similarly, for E-Task lists, 20% E-Task lists were expected to produce greater correct recognition relative to 50% and 80% lists. In contrast to the predictions for Experiment 1, we expected that overall false recognition would be greater for PR lists relative to E-Task lists. Previous research has shown that shallow LOP tasks often reduce false recognition relative to deep LOP tasks (Thapar & McDermott, 2001; Toglia et al., 1999) due to shallow tasks failing to encode sufficient associative/thematic information about study lists to produce a robust DRM illusion. Despite these task effects, we still anticipated sensitivity to task proportions, with 20% PR lists and 20% E-Task lists producing lower false recognition relative to their respective 50% and 80% lists due to statistical distinctiveness.

Method

Participants

A separate sample of undergraduates ($N = 72$; 83% female, $M_{age} = 19.78$, $SD_{age} = 2.62$) from The University of Southern Mississippi were recruited for participation. Participants were compensated with partial course credit and had a normal or corrected-to-normal vision. Participants were randomly assigned to one of the 3 PR/E-Task mixed groups (Table 1).

Materials and Procedure

The same materials used in the Experiment 1 were used in Experiment 2. Procedurally, the only differences between these experiments were the study task combinations used in the mixed groups. In the E-Task, participants were asked whether each list word contained the vowel "e" by responding "yes" or "no" on a labelled keyboard, and the PR task made pleasantness ratings as in Experiment 1. In the 80% PR versus 20% E group, participants studied 8 lists using the PR task and 2 using the E-Task. In the 50% PR versus 50% E-Task group, participants studied 5 lists using the PR task and 5 using the E-Task. In the 20% PR versus 80% E-Task group, participants studied 2 lists using the PR task and 8 using the E-Task. PR and E-Tasks were counterbalanced as in Experiment 1.

Results

Proportions of correct and false recognition as a function of task type and proportion are reported in Table 2.

Correct Recognition

Analyses were conducted as in Experiment 1. A 2(Task Type: PR vs. E) × 3(Proportion: 20/80 vs. 50/50 vs. 80/80) mixed ANOVA found that correct recognition was greater for PR than E-Task lists (.91 vs. .62), $F(1, 69) = 129.41$, $MSE = .02$, $\eta_p^2 = 0.65$, consistent with encoding distinctiveness. Unlike Experiment 1 however, a main effect of Proportion also emerged, $F(2, 69) = 3.65$, $MSE = .04$, $\eta_p^2 = .10$. Follow-up tests indicated that correct recognition was greater when collapsed across task types in the 50/50 than the 80/20 group (.82 vs. .72), $t(46) = 2.68$, $SEM = .04$, $d = 0.77$, but equivalent to the 20/80 group (.82 vs. .76), $t(46) = 1.59$, $SEM = .04$, $p = .12$, $p_{BIC} = .65$. Recognition did not differ between the 20/80 and 80/20 groups (.76 vs. .72), $t(46) = 1.14$, $SEM = .04$, $p = .26$, $p_{BIC} = .78$. The List Type × Proportion interaction was not reliable, $F(2, 69) = 1.55$, $MSE = .02$, $p = .22$, $p_{BIC} = .94$, yielding no evidence for statistical distinctiveness.

False Recognition

False recognition was then analysed to examine the contribution of statistical distinctiveness on recognition errors. A 2(Task Type) × 3(Proportion) ANOVA showed a main effect of List Type, $F(1, 69) = 34.68$, $MSE = .06$, $\eta_p^2 = .33$, indicating that false recognition was greater for the PR than E-Task lists (.71 vs. .46), but both the main effect of Proportion and the interaction were not reliable, $F < 1$, $p_{BIC} = .98$, and $F(2, 69) = 2.05$, $MSE = .06$, $p = .14$, $p_{BIC} = .90$, respectively. Again, task frequency did not moderate false recognition.

Discussion

Similar to Experiment 1, no evidence was found for statistical distinctiveness when comparing PR and E-Tasks in correct or false recognition. Correct recognition was greater overall for PR lists relative to E-Task lists, consistent with encoding distinctiveness and interestingly, was greater in the 50/50 group. We discuss this pattern in the General Discussion but note here that this pattern suggests an advantage for an equal distribution of study tasks relative to an unbalanced distribution. Further, lists studied using the PR task were at ceiling (.89 or greater across conditions) which may have restricted differences across task distributions. False recognition was greatest for PR lists over read lists consistent with previous literature (Thapar & McDermott, 2001; Toglia et al., 1999), but was not sensitive to task frequency.

Experiment 3: read vs. E tasks

Given the null effects of statistical distinctiveness in the previous experiments, possibly due to correct recognition at ceiling on PR lists, Experiment 3 sought to provide another test of statistical distinctiveness by excluding the PR task and comparing two study tasks that do not produce exceptional correct recognition. To remain consistent with previous experiments, Experiment 3 compared different distributions of task lists by comparing recognition following the Read task to the E-Task. We expected that the Read task would operate as a “middle” LOP task and the E-Task as a shallow task. When compared to each other in the same context, we anticipated that the Read task would be the relatively more distinctive task relative to the E-Task. Based on this logic, we predicted that the Read task would produce an encoding distinctiveness benefit over the E-Task, even though recognition would be lower overall. Additionally, we expected that with lower recognition rates, correct recognition would be greatest for tasks that constitute 20% of study lists, followed by 50% and 80% of study lists, consistent with a statistical distinctiveness benefit. For false recognition, given E-Task lists produced low rates of false recognition in Experiment 2, we similarly anticipated that false recognition following the E-Task would be lower relative to Read lists. Additionally, we expected that 20% Read lists and 20% E-Task lists would produce lower false recognition relative to their respective 50% and 80% lists, consistent with statistical distinctiveness.

Method

Participants

A separate sample of University of Southern Mississippi undergraduates ($N = 72$; 75% female, $M_{age} = 20.62$, $SD_{age} = 5.75$) completed the experiment for partial course credit and had normal or corrected-to-normal vision. They were randomly assigned to one of 3 mixed groups (Table 1).

Materials and procedure

Study materials were identical to those used in Experiment 1 with the exception that participants were instructed to use either the Read task or the E-Task. Instructions for these two tasks were identical to those used in previous experiments.

Results

Proportions of correct and false recognition as a function of task type and proportion are reported in Table 2.

Correct recognition

The same mixed ANOVA was again used to compare Read and E-Tasks across different study proportions. Main effects of Task Type, $F(1, 69) = 1.30$, $MSE = .03$, $p = .26$, $p_{BIC} = .85$, and Proportion, $F < 1$, were not reliable, but a marginal interaction was found, $F(2, 69) = 2.81$, $MSE = .03$, $p = .07$, $\eta_p^2 = .08$, $p_{BIC} = .81$. Follow-up tests revealed a marginal increase in correct recognition for 20% Read lists over

80% E-Task lists (.76 vs. .65), $t(23) = 1.92$, $SEM = .06$, $p = .07$, $d = 0.45$, $p_{BIC} = .80$, however, this pattern was not found in either the 50/50 condition (.70 vs. .75), $t(23) = 1.44$, $SEM = .03$, $p = .17$, $p_{BIC} = .90$, nor the 80/20 condition (.72 vs. .69), $t < 1$, $p_{BIC} = .95$.

False recognition

The same ANOVA was also conducted on proportions of false recognition. A marginal effect of Task Type was found, $F(1, 69) = 3.65$, $MSE = .07$, $p = .06$, $\eta_p^2 = .05$, $p_{BIC} = .57$, in which false recognition was greater for Read than E-Task lists (.69 vs. .61), however the main effect of Proportion and the interaction, were not reliable, Fs < 1, p_{BIC}s > .98. Thus, mixed groups did not show evidence for statistical distinctiveness in false recognition.

Discussion

When examining correct recognition for Read and E-Task lists, a marginal interaction was found in which correct recognition was greatest for the 20% Read lists over 80% E-Task lists. Of course, we do not make any strong claims based on this marginal pattern and want to emphasise that although a trend was found between these list types, the pattern predicted by statistical distinctiveness in the 20% E-Task/80% Read task group did not emerge. Thus, the marginal benefit found for the 20% Read lists over the 80% E-Task lists may simply reflect a LOP effect versus a statistical distinctiveness pattern. For false recognition, a marginal reduction was found for E-Task lists over Read lists, a trend consistent with predictions, but again, no interaction with proportion was found. Thus, despite off-ceiling correct recognition for both task lists, there was little support for statistical distinctiveness.

Experiment 4: pure group comparisons

Our previous experiments have operated under the assumptions that the PR task serves as a distinctive/deep task, the E-Task asa shallow task, and the Read task as a neutral task. We therefore tested a set of pure groups in Experiment 4 to confirm the graded depth of processing within the LOP framework. For correct recognition, a LOP effect was expected across the pure groups, in which recognition would be greatest following the PR task, followed by the Read task, and then the E-Task. For false recognition, the PR task and the E-Task were expected to reduce false recognition relative to the Read task, consistent with both Huff and Bodner (2013) and Toglia et al. (1999). The reduction in false recognition in the PR task is expected to reflect the use of item-specific/distinctive processing, and the reduction in the E-Task is expected to be due to poor encoding of semantic information associated with studied list items.

Method

Participants

An additional sample of University of Southern Mississippi undergraduates ($N = 76$; 90% female, $M_{age} = 20.31$, $SD_{age} = 5.34$) participated for partial course credit. Participants were randomly assigned to either the PR, Read, or E-Task pure groups (Table 1).

Materials and procedure

The same materials and study instructions were identical to those used in previous experiments. The only difference was that participants used a single study task for all lists.

Results

Proportions of correct and false recognition for each of the three pure groups are reported in Table 2.

Correct recognition

Analyses were first conducted on correct recognition from the pure groups to verify LOP/distinctiveness effects on recognition. A one-way ANOVA revealed a significant difference across groups, $F(2, 73) = 25.90$, $MSE = .02$, $\eta_p^2 = .07$, which indicated that correct recognition was greater in the PR group than both the Read (.95 vs. .78), $t(50) = 6.13$, $SEM = .03$, $d = 1.73$, and E-Task groups (.95 vs. .66), $t(48) = 6.88$, $SEM = .04$, $d = 1.99$. Correct recognition was also greater in the Read than the E-Task group (.78 vs. .66), $t(48) = 2.46$, $SEM = .05$, $d = 0.71$. Thus, PR, Read, and E-Tasks showed patterns consistent with LOP and encoding distinctiveness.

False recognition

Analyses were then conducted to examine false recognition for non-presented critical lures. False recognition was not found to differ across conditions, $F(2, 73) = 2.02$, $MSE = .05$, $p = .14$, $p_{BIC} = .91$, though the pattern was in the expected direction: False recognition for critical lures in the Read group ($M = .77$) was numerically greater than the PR group ($M = .69$) and the E-Task group ($M = .65$).

Discussion

The pure group comparisons found evidence supporting a LOP/encoding distinctiveness pattern in which correct recognition was greatest for the PR group, followed by the Read group, and the E-Task group. This pattern supports our previous assumptions regarding the depth of encoding manipulations in the previous experiments. For false recognition, the Read group numerically produced the greatest false recognition rate relative to the PR and E-Task groups. These comparisons were not statistically reliable, though they were in the expected direction given patterns reported in previous studies (Huff & Bodner, 2013; Thapar & McDermott, 2001).

Prior to providing a detailed discussion of our findings, we next report a series of secondary cross-experimental comparisons of correct and false recognition. Given the consistency found across the three within-group experiments (Experiments 1–3) regarding the proportional manipulations across tasks, our first cross-experimental comparison evaluates these similarities. The cross-experimental comparison also provides us with additional statistical power to detect potential statistical distinctiveness patterns that were absent in the mixed group experiments. Following comparisons across mixed groups, we then provide a final test of statistical distinctiveness by comparing the pure groups in Experiment 4 to the 20% mixed lists in Experiments 1-3. We home in on this latter comparison because the pure groups maximise the frequency with which a task is completed (100%), which may provide a more sensitive test of statistical distinctiveness versus the 80% comparisons conducted in the mixed-group experiments.

Mixed-list cross-experimental comparison: experiments 1–3

Correct recognition

To determine the consistency of the proportion manipulation across experiments, we first conducted a 3(Experiment: 1 (PR/Read) vs. 2 (PR/E) vs. 3 (Read/E)) × 3 (Proportion: 20/80 vs. 50/50 vs. 80/20) × 2 (Task Type: Deep vs. Shallow) mixed ANOVA in which Experiment and Proportion were between-subject variables and Task Type was within. For this analysis, the deep versus shallow levels for the task type variable reflected a relative difference between the two tasks used in the three experiments (i.e., experiment tasks that were relatively deep vs. shallow).

A main effect of Task Type was found, $F(1, 208) = 153.58$, $MSE = .03$, $\eta_p^2 = .43$, which indicated greater correct recognition following deep than shallow tasks (.86 vs. .66). An effect of Proportion was also found, $F(2, 208) = 3.28$, $MSE = .04$, $\eta_p^2 = .03$, in which the 20% Deep/80% Shallow tasks did not differ from either the 50%/50% tasks or the 80% Deep/20% Shallow tasks (.76 vs. .79, $t(143) = 1.21$, $SEM = .02$, $p = .23$, $p_{\mathrm{BIC}} = .85$, and .76 vs. .79, $t(143) = 1.23$, $SEM = .02$, $p = .22$, $p_{\mathrm{BIC}} = .84$, respectively), however correct recognition was greater in the 50%/50% grouping than the 80% Deep/20% Shallow grouping (.79 vs. .73), $t(142) = 2.63$, $SEM = .02$, $d = .44$. An effect of Experiment was found in which correct recognition was marginally greater for tasks in Experiment 1 (PR/Read tasks) than in Experiment 2 (PR/E-Tasks; .80 vs. .76), $t(143) = 1.84$, $SEM = .02$, $p = .07$, $d = 0.31$, $p_{\mathrm{BIC}} = .80$, greater in Experiment 1 than in Experiment 3 (Read/E-Tasks; .80 vs. .71), $t(143) = 3.82$, $SEM = .03$, $d = 0.63$, and greater in Experiment 2 than Experiment 3 (.76 vs. .71), $t(142) = 2.63$, $SEM = .03$, $d = 0.34$. The relatively deeper task combinations therefore generally produced greater correct recognition than shallower task combinations – a pattern also consistent with encoding distinctiveness. The effect of Experiment was only found to interact with Task Type, $F(2, 208) = 27.14$, $MSE = .03$, $\eta_p^2 = .21$, in which the recognition improvement for the relatively deep over shallow task was reliable for both Experiment 1 (.94 vs. .69), $t(72) = 9.31$, $SEM = .03$, $d = 1.55$, and Experiment 2 (.91 vs. .62), $t(71) = 11.29$, $SEM = .03$, $d = 1.68$, but not Experiment 3 (.73 vs. .70), $t(71) = 1.11$, $SEM = .03$, $p = .27$, $p_{\mathrm{BIC}} = .88$, in which the PR task was not included. All other interactions, including the three-way interaction, were not reliable, all Fs < 2.37, ps > .09, p_{BIC}s > .94, demonstrating that the lack of statistical distinctiveness was consistent across experiments which utilised different tasks.

False recognition

Cross-experimental analyses were similarly conducted on false recognition using the same ANOVA. An effect of Task Type was found, $F(1, 208) = 24.20$, $MSE = .07$, $\eta_p^2 = .10$, in which false recognition was greater for relatively deeper than shallow tasks across experiments (.71 vs. .58). The effect of Proportion was not significant, $F < 1$, but an effect of Experiment was found, $F(2, 208) = 3.12$, $MSE = .12$, $\eta_p^2 = .03$, in which false recognition was greater in the relatively deeper task combinations in Experiment 1 relative to Experiment 2 (.69 vs. .59), $t(143) = 2.57$, $SEM = .04$, $d = 0.43$, but did not differ from Experiment 3 (.69 vs. .65), $t(143) = 1.13$, $SEM = .04$, $p = .26$, $p_{\mathrm{BIC}} = .86$. False recognition did not differ between Experiment 2 and 3 (.59 vs. .65), $t(142) = 1.37$, $SEM = .04$, $p = .17$, $p_{\mathrm{BIC}} = .82$. A reliable Task Type × Experiment interaction was found, $F(2, 208) = 5.46$, $MSE = .07$, $\eta_p^2 = .05$, which reflected generally lower false recognition for E-Task lists. Specifically, false recognition did not differ between PR and Read lists in Experiment 1 (.72 vs. .68), $t(72) = 1.08$, $SEM = .05$, $p = .29$, $p_{\mathrm{BIC}} = .83$, but was greater for the deeper PR lists relative to E-Task lists in Experiment 2 (.71 vs. .47), $t(71) = 5.80$, $SEM = .04$, $d = 0.80$, and marginally greater for Read lists over E-Task lists in Experiment 3 (.68 vs. .60), $t(71) = 1.93$, $SEM = .04$, $p = .06$, $d = 0.27$, $p_{\mathrm{BIC}} = .57$. Importantly, all other interactions, including the three-way interaction with Experiment were not reliable, Fs < 1.17, ps > .32, p_{BIC}s >.97, again demonstrating no evidence for statistical distinctiveness across experiments.

Cross-experimental comparisons between mixed and pure groups: experiments 1–4

Given the lack of statistical distinctiveness effects in correct and false recognition in the mixed groups, our next set of analyses examined whether statistical distinctiveness effects emerged when 20% task lists in Experiments 1–3 were compared to their corresponding pure group in Experiment 4. Since pure groups maximise the frequency of a given task, we reasoned that the comparison between 20% task lists and the 100% pure groups would be most sensitive towards detecting statistical distinctiveness effects.

Correct recognition

20% E-Task Mixed Lists vs. Pure E-Task Group. In the PR/E task combination, where the E-Task was used for 20% of lists, a marginal reduction was found for 20% lists relative to the pure E-Task group (.54 vs. .65), $t(46) = 1.74$, $SEM = .07$, $p = .09$, $d = 0.51$, $p_{BIC} = .60$. In the Read/E task combination, where the E-Task was used for 20% of lists, no difference in recognition was found between the 20% E-Task listsand the pure E-Task group (.70 vs. .65), $t < 1$, $p_{BIC} = .85$.

20% Read Mixed Lists vs. Pure Read Group. In the PR/Read task combination, where Reading was used for 20% of lists, a significant *reversed* statistical distinctiveness effect was found in which correct recognition was lower in the 20% Read lists versus the Read pure group (.63 vs. .78), $t(48) = 2.42$, $SEM = .06$, $d = 0.68$. This pattern did not occur however when comparing the 20% Read lists in the Read/E combination, as there was no difference relative to the pure Read group (.78 vs. .78), $t < 1$, $p_{BIC} = .87$.

20% PR Mixed Lists vs. PR Pure Group. In the PR/E task combination, where PR was used for 20% of lists, a reversed statistical distinctiveness pattern was again found (.89 vs. .95), $t(48) = 1.98$, $SEM = .03$, $d = 0.55$, for the 20% PR lists and PR pure group, respectively. In the PR/Read task combination, correct recognition was equivalent between the 20% PR listsand the pure group (.97 vs. .95), $t < 1$, $p_{BIC} = .87$. Collectively, across list types, comparisons between the 20% mixed lists and their corresponding pure group revealed no evidence of statistical distinctiveness and indeed, some comparisons showed a reversed statistical distinctiveness pattern.

False recognition

20% E-Task Mixed Lists vs. Pure E-Task Group. In the PR/E task combination, where the E-Task was used for 20% of lists, analyses revealed a significant reduction in false recognition for 20% E-Task lists relative to the pure E-Task group (.39 vs. .64), $t(47) = 2.78$, $SEM = .09$, $d = 0.80$, a pattern consistent with PR lists reported above. In the Read/E-Task combination, where the E-Task was used for 20% of list items, no differences were also found between 20% E-Task list items and the E-Task pure group (.58 vs. .65), $t < 1$, $p_{BIC} = .82$.

20% Read Mixed Lists vs. Pure Read Group. In the PR/Read task combination, where Reading was used for 20% of lists, no differences were found in recognition between the 20% Read lists and the Read pure group (.69 vs. .77), $t < 1$, $p_{BIC} = .82$, as was the case when comparing 20% Read lists in the Read/E combination (.67 vs. .77), $t(48) = 1.19$, $SEM = .09$, $p = .24$, $p_{BIC} = .77$.

20% PR Mixed Lists vs. PR Pure Group. In the PR/E task combination, where PR was used for 20% of list items, analyses revealed no differences in recognition for 20% PR lists compared to the PR pure group (.69 vs. .69), $t < 1$, $p_{BIC} = .87$, as was the case in the PR/Read task combination, where PR was used for 20% of lists, (.97 vs. .95), $t(49) = 1.11$, $SEM = .08$, $p = .27$, $p_{BIC} = .79$. In summary, statistical manipulations yielded minimal effects on false recognition with the exception of the 20% E-Task lists. When taken together with correct recognition, the comparison of 20% lists to the pure groups, which is a comparison that is more sensitive towards statistical distinctiveness, again did not produce consistent evidence for this process.

General discussion

The purpose of our study was to evaluate contributions of two types of distinctiveness on recognition. Encoding distinctiveness refers to the memorial benefits that originate from distinctive processing fostered by the encoding task itself. Statistical distinctiveness refers to the benefits found when encoding tasks are utilised infrequently. Statistical distinctiveness on correct and false recognition was examined in Experiments 1–3 by manipulating the frequency with which participants utilised a given study task in a set of mixed groups. Participants studied lists of words using one task for 20%, 50%, or 80% of lists and another task for the remaining lists, making the task used for 20% of the lists statistically distinctive. In Experiment 4, encoding distinctiveness on correct and false recognition was examined using a set of pure groups by comparing a distinctive/deep task (Pleasantness Ratings) to a relative neutral task (silent reading), and a non-distinctive shallow task ("E" identification).

Across experiments, consistent evidence for encoding distinctiveness was found in correct recognition. In the Experiment 4 pure groups, a standard LOP effect emerged in which the deeper/distinctive PR task led to greater correct recognition than the neutral Read task and the shallow E-Task. Of note, similar encoding patterns were detected in mixed groups used in Experiments 1-3, in which relatively deeper task combinations (e.g., PR/Read) led to an increase in correct recognition relative to more neutral or shallow task combinations (i.e., PR/E-Task and Read/E-Task) – a novel result. The combined task effects of pure and mixed groups confirm powerful benefits of encoding distinctiveness as the distinctive PR task, and any mixed group combination that included this task, increased correct recognition.

Statistical distinctiveness, however, was not in evidence. In Experiments 1-3, correct recognition in the statistically rare 20% task lists did not produce a recognition benefit relative to the more frequent 50% or 80% task lists and this pattern was generally stable across experiments. The only exception to this pattern occurred in Experiment 3 in which 20% Read lists marginally improved correct recognition relative to 80% E-Task lists. Of note, this pattern did not hold in the reversed mixed group (20% E-Task/80% Read) in which E-Task lists were infrequent. Moreover, when statistical frequency was evaluated relative to pure groups which maximize task use, no recognition benefit for the infrequent 20% lists emerged. These patterns are inconsistent with those reported by Icht et al. (2014) and Bodner et al. (2016), whereby recall and recognition benefits in a production-effect paradigm were found for silent and aloud items that were studied infrequently.

Encoding and statistical distinctiveness effects were also examined for false recognition of DRM critical lures. We anticipated that, given distinctive encoding tasks often reduce the DRM illusion relative to a non-distinctive control (e.g., Huff et al., 2015), conditions that promoted statistical distinctiveness would similarly reduce false recognition of lures, improving overall memory accuracy. As was found in correct recognition, statistical distinctiveness produced no effect on false recognition, either when comparing frequency rates in mixed groups or when comparing the rarer 20% lists to the pure groups in the cross-experimental comparisons. Although examining statistical distinctiveness effects on false recognition has not been explored previously, the benefits of statistical distinctiveness reported by Bodner et al. (2016) and Icht et al. (2014) should theoretically have translated to reductions in false recognition for 20% lists. However, since no evidence for statistical distinctiveness in correct recognition was found, it is unsurprising that such proportional influences did not extend to false recognition.

When examining false recognition rates on pure groups in Experiment 4, false recognition was only numerically reduced for the PR and E-Task relative to the Read task. These patterns are at least directionally consistent with previous work showing that shallow LOP tasks can produce a reduction in the DRM illusion (Thapar & McDermott, 2001), likely due to shallow processing disrupting the amount of encoded associative/thematic information from the list items needed to produce the illusion, and other studies that have shown the PR task reduces the DRM illusion relative to a read control (Huff & Bodner, 2013).

The hunt for statistical distinctiveness

Across both PR/Read and PR/E combinations in Experiments 1 and 2, our analyses did not reveal an interaction between task type and task proportion, suggesting that statistical manipulations do not influence correct recognition. For these task combinations, PR lists consistently performed at ceiling, with no change in performance regardless of the frequency in which the task was completed. Although the tasks used in this study exhibited a standard LOP effect, the PR task produced ceiling performance which was immune to our proportional manipulations. Separately, the Read/E group in Experiment 3, which did not include the PR task, appeared to be more sensitive towards the proportional manipulation. Correct recognition was marginally greater for 20% Read lists over 80% E-Task lists, but this pattern was not found in the reversed distribution. These findings are somewhat consistent with Icht et al. (2014; Experiment 2), in which a "dilution" of the production effect was found for the 80% aloud lists versus the 20% silent lists, though silent lists did not exceed aloud lists.

The finding of some sensitivity towards proportional effects when the PR task was absent in mixed groups may suggest that statistical distinctiveness may be more likely to occur when two task types are relatively equivalent in encoding strength. To date, statistical distinctiveness patterns have only been reported in studies that have utilised a production paradigm. The relative memory difference between aloud and silent items is typically moderate (Fawcett, 2013), which may increase the likelihood that statistical distinctiveness effects are detected. Thus, the relative difference in encoding strength between the two tasks used in mixed groups may moderate whether statistical distinctiveness benefits are detected.

Additionally, our mixed-group analyses yielded an interesting task distribution effect based on whether study tasks were used on an equal versus unequal number of study lists. In our cross-experimental comparison, correct recognition in the balanced 50/50 groups was reliably greater than the unbalanced 80/20 groups, and numerically greater than the 20/80 groups. It is unclear exactly why this pattern emerged, but one possibility may be due to the 80/20 and 20/80 groups having to complete one task repetitively relative to the other. When using a single task repetitively, participants may have "loafed" their use of that task which may have rendered task encoding less effective than alternating between tasks in the 50% group (see Huff et al., 2016, for a similar repetitive task-loafing discussion). Consistent with a loafing pattern based on repetition, if one compares the 80% lists to the 50% lists across experiments in Table 2, correct recognition for the 80% lists is 4-10% *lower* than the task matched 50% lists (the one exception being the 80% Read lists versus the 50% Read lists in Experiment 3). Alternating between two tasks equally may have reduced loafing as tasks were not repeated and participants were more likely to remain engaged across lists. Of course, we only speculate on the processes behind this pattern, but these findings suggest that equivalent distributions of tasks may produce an overall recognition benefit relative to task being used unequally which could encourage loafing. Future research on this effect is clearly needed, but it may have implications in educational settings, such as having learners distribute separate study tasks evenly versus favouring one over another.

Although our data did not support a statistical distinctiveness pattern, our methods were designed to mimic those used in other studies that have shown statistical distinctiveness effects (Bodner et al., 2016; Icht et al., 2014), but with some exceptions. First, as noted above, the present study utilised the LOP framework to examine statistical distinctiveness, whereas other studies that have shown statistical distinctiveness effects used a production task. It is possible that the type of task used may be sensitive to the proportional manipulation needed to elicit statistical effects, and the inclusion of the PR task, which was quite powerful, may have masked the effect. Further, the present study used DRM word lists as study materials (vs. unrelated word lists used in Bodner et al. and Icht et al.) to gauge statistical distinctiveness effects on correct and false recognition. The strong association between items in DRM lists may have muted task-proportion effects in recognition as participants may have been more reliant upon thematic cues at test, rather than the task type and task frequency to encode lists.

Finally, though the task proportions used in the present study were similar to statistical proportions used by Bodner et al. (2016) and Icht et al. (2014), the previous studies manipulated statistical distinctiveness at the item level, rather than at the list level. At the item level, participants were presented with a single list of items in which aloud versus silent tasks were cued for individual items. At the list level, however, a single task was used for an entire DRM list, in which participants used the same task repeatedly over several items. It is possible that participants may be treating the individual DRM lists, and the tasks used to study these lists, as separate "memory events" versus a single memory event, which would be the case for a single study list used in previous studies. The frequency of tasks used across memory events may therefore fail to produce a statistical distinctiveness pattern. Instead, the frequency in which a task is used *within* a single memory event, such a single study list, will only produce statistical distinctiveness.

An additional factor which may affect the sensitivity for detecting statistical distinctiveness is the type of memory test completed. Icht et al. (2014) reported a diminished statistical distinctiveness pattern on recognition versus free recall. Free recall is generally believed to be more sensitive to recollection which requires greater use of controlled processing. In contrast, recognition is a discrimination memory task which allows for greater contributions from familiarity-based processes (Mandler, 1980; Yonelinas, 2002). It is possible that recollection-based processes may be more sensitive to statistical distinctiveness which may be underestimated in recognition due to familiarity. Thus, smaller distinctiveness pattern reported by Icht et al. (2014) and our lack of support for statistical distinctiveness using recognition indicate that test type may be an important moderator.

Finally, we acknowledge that our use of the LOP framework to examine the effects of encoding and statistical distinctiveness assumes that the deep PR task qualifies as a distinctive task. Deep processing tasks can occur from either item-specific or relational processing (e.g., relational generation; Huff & Bodner, 2013), whereas distinctive processing only utilises the former. The PR task qualifies as both a deep task, facilitating elaborative processing of items, and as a distinctive task, by inducing item-specific processing (Hunt & Einstein, 1981). In contrast, relational processing preserves the thematic consistency across list stimuli but dilutes distinctive processing at the item level. Therefore, relational processing includes deep semantic processing, but not distinctive processing. We note that the PR task is a standard item-specific task meant to induce distinctive encoding (Hodge & Otani, 1996; Hunt et al., 2011; Hunt & Smith, 2014). Thus, we used the LOP framework to provide reasonable non-distinctive comparison tasks (Read and the E-Task), but our reference to a deep processing task refers to item-specific/distinctive encoding.

Conclusion

The purpose of the current study was to assess the contributions of two possible species of distinctiveness: Encoding and statistical distinctiveness on recognition memory. Three encoding tasks were used (Pleasantness Ratings, Reading, or "E" Identification) which varied in their use of distinctive processing and were manipulated to occur relatively frequently when studying a series of lists (100% in the pure groups or 80% in the mixed groups), or relatively infrequently (20% in the mixed groups). Although distinctive encoding produced large correct recognition benefits, no evidence for statistical distinctiveness was found when any of the three tasks were completed infrequently. The absence of statistical distinctiveness may be due to the inclusion of the powerful pleasantness rating task, which may override the benefits of statistical distinctiveness, the use of strongly related lists, a recognition (vs. recall) test, or some combination thereof. We suggest that it is important for theoretical reasons to determine whether encoding and statistical distinctiveness are indeed separate species and factors that affect their magnitude.

Disclosure of interest

This project was used to satisfy the Master's thesis requirements for the first author. We thank Jessica Runnels and Kayla Carr for assistance in data collection. Participant-level data for all experiments reported are available via our OSF page: osf.io/k73r4.

Notes

1. Due to experimenter error, the practice "fruit" list was mistakenly included as a studied list in one of the counterbalanced versions in all experiments. Due to this repetition, study items from the fruit list and the "fruit" critical lure were removed from the analyses in this version. This counterbalance was represented equivalently across the three mixed groups.
2. In addition to the analyses using raw recognition rates for studied list items and critical lures, we further conducted a signal-detection analysis (Wickens 2002) on d' values for these items using Macmillan and Creelman's (1991) $1/2n$ correction for all experiments. In the mixed-list experiments (Experiments 1-3), only one false alarm rate was available for list item and critical lure controls which was subtracted equally from both task proportions. Given this equivalent

subtraction, the analyses closely matched those on raw recognition rates. In Experiment 4, d' values did differ from raw recognition analyses due to differences in false alarm rates across groups (task differences were generally larger), however the task patterns were directionally the same. Therefore, to avoid redundancy and remain consistent across experiments, our analyses only report raw recognition rates.

Disclosure statement

No potential conflict of interest was reported by the author(s).

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

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 941–947. https://doi.org/10.1037/0278-7393.27.4.941

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., 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/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 93–98. https://doi.org/10.1037/cep0000082

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1711–1719. https://doi.org/10.1037/a0028466

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 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*(1), 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), 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

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and 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*. 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

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*(3), 579–586. https://doi.org/10.3758/BF03196194

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*(2), 110–119. https://doi.org/10.3758/BF03213478

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 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*(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

Hodge, M. H., & Otani, H. (1996). Beyond category sorting and pleasantness rating: Inducing relational and item-specific processing. *Memory & Cognition*, *24*(1), 110–115. https://doi.org/10.3458/BF03197277

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*(8), 1287–1301. https://doi.org/10.3758/s13421-018-0837-1

Huff, M. J., Balota, D. A., & Hutchison, K. A. (2016). The costs and benefits of testing and guessing on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1559–1572. https://doi.org/10.1037/xlm0000269

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*(4), 1246–1256. https://doi.org/10.1037/a0031338

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

Huff, M. J., Bodner, G. E., & Gretz, M. R. (in press). Performing item-specific vs. relational generation at encoding on a subset of DRM lists: Benefits, costs, and spillovers. *Psychological Research*. https://doi.org/10.1007/s00426-019-01241-y.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt, J. B. Worthen, R. R. Hunt, & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York, NY: 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., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1359–1366. https://doi.org/10.1037/0278-7393.27.6.1359

Hunt, R. R., & Smith, R. E. (2014). How distinctive processing enhances hits and reduces false alarms. *Journal of Memory and Language*, *75*, 45–57. https://doi.org/10.1016/j.jml.2014.04.007

Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language*, *65*(4), 378–389. https://doi.org/10.1016/j.jml.2011.06.003

Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory*. Oxford University Press.

Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology*, *5*, 1–7. https://doi.org/10.3389/fpsyg.2014.00886

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

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*(3), 153–161. https://doi.org/10.1016/S0022-5371(62)80023-1

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*, *6*(1), 130–135. https://doi.org/10.3758/BF03210820

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York, NJ: Cambridge University Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavioral Research Methods*, *44*(3), 1–23. https://doi.org/10.3758/s13428-011-0124-6

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. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*(7), 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 (Hove, England)*, *14*(5), 570–583. https://doi.org/10.1080/09658210600624564

McDaniel, M. A., & Waddill, P. J. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, *27*(5), 521–536. https://doi.org/10.1016/0749-596X(88)90023-X

McDermott, K. B. (1996). The Persistence of false Memories in list recall. *Journal of Memory and Language*, *35*(2), 212–230. https://doi.org/10.1006/jmla.1996.0012

Nairne, J. S., Thompson, S. R., & Pandierada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263

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/10.1080/00221309.2015.1110558

Parkin, A. J. (1980). Levels of processing and the Cue overload principle. *Quarterly Journal of Experimental Psychology*, *32*(3), 427–434. https://doi.org/10.1080/14640748008401836

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). https://www.pstnet.com

Robin, F. (2010). Imagery and memory illusions. *Phenomenology and the Cognitive Sciences*, *9*(2), 253–262. https://doi.org/10.1007/s11097-010-9159-x

Roediger, H. L., & 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., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning Memory*, *4*(6), 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*(4), 710–715. https://doi.org/10.3758/BF03208850

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*(3), 494–500. https://doi.org/10.3758/BF03211543

Surprenant, A., & Neath, I. (2009). *Principles of memory*. Psychology Press.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*(3), 424–432. https://doi.org/10.3758/BF03196393

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory (Hove, England)*, *7*(2), 233–256. https://doi.org/10.1080/741944069

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wickens, T. D. (2002). *Elementary signal detection theory*. New York, NJ: 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*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864