

The bizarreness effect: evidence for the critical influence of retrieval processes

Lisa Geraci · Mark A. McDaniel · Tyler M. Miller ·
Matthew L. Hughes

Published online: 5 June 2013
© Psychonomic Society, Inc. 2013

Abstract People show better memory for bizarre sentences relative to common sentences, a finding referred to as the *bizarreness effect*. Interestingly, this effect is typically only obtained using a mixed-list design, in which participants study common and bizarre sentences in the same list. This bizarreness effect in mixed-list designs has been explained as the result of both enhanced encoding processes and efficient retrieval processes. The present experiment was designed to isolate the unique contributions of the retrieval context to the bizarreness effect. Participants studied common sentences in one room under one set of instructions, and bizarre sentences in another room under another set of instructions. At test, participants recalled the common and bizarre sentences either together or separately. The results showed that the bizarreness effect was only obtained when participants recalled the common and bizarre items together; no bizarreness advantage emerged when participants were required to recall the common and bizarre items separately. These results suggest that differential encoding processes are not necessary for explaining the bizarreness effect in memory. Rather, retrieval of the mixed-list context appears to be critical for obtaining the effect.

Keywords Memory · Distinctiveness · Bizarreness · Encoding · Retrieval

L. Geraci (✉) · M. L. Hughes
Department of Psychology, Texas A&M University,
College Station, TX 77843, USA
e-mail: lgeraci@tamu.edu

M. A. McDaniel
Department of Psychology, Washington University,
St. Louis, MO, USA

T. M. Miller
Department of Psychology, South Dakota State University,
Brookings, SD, USA

Introduction

The *bizarreness effect* refers to the finding that people have superior memory for bizarre sentences relative to common ones (see Einstein & McDaniel, 1987, and Worthen, 2006, for reviews). In the standard bizarreness paradigm investigated here, people study common and bizarre sentences that include target capitalized nouns. Later, participants are given a free recall task for the nouns. Typically, when the nouns are presented in a bizarre sentence, such as “The DOG rode the BICYCLE down the STREET,” they are recalled better than when they are presented in the common counterpart (“The DOG chased the BICYCLE down the STREET”). This bizarreness effect is a robust finding in recall that has been obtained across a variety of encoding tasks and delays.

Although the bizarreness effect is a robust and intuitive finding, it has one peculiarity: The bizarreness effect is usually only obtained if participants study (and recall) bizarre sentences in a list that also contains common sentences. The bizarreness recall advantage does not occur when participants are asked to study a list containing only common sentences or a list containing only bizarre sentences (see McDaniel, Einstein, DeLosh, May, & Brady, 1995, for a review). The finding that the bizarreness effect is obtained in mixed- but not between-list designs mimics the pattern of other types of distinctiveness effects. For example, the orthographic distinctiveness effect (superior memory for odd-looking words, such as *sphinx*, *subpoena*, *xerox*, *epitome*, etc., relative to common-looking words) is obtained in mixed-list designs (e.g., Geraci & Rajaram, 2002; Hunt & Elliott, 1980; Hunt & Toth, 1990), but not in pure-list designs (for recall; Hunt & Elliott, 1980; McDaniel, Cahill, Bugg, & Meadow, 2011). By definition, the isolation effect (or the von Restorff effect: von Restorff, 1933; see also Hunt, 1995) is a within-subjects effect, in that the isolate is distinct with respect to the other items in the study list. Other memory effects that have sometimes been attributed to distinctiveness are similarly obtained only in mixed-list designs (see McDaniel & Bugg, 2008, for a comprehensive summary).

Encoding and retrieval theories of the bizarreness effect

Two classes of theories have been offered to account for the mixed-list advantage for distinctiveness effects in general. These theories can be grouped into *encoding* and *retrieval* theories (see McDaniel & Geraci, 2006, for a review). According to encoding theories, the unusual item attracts additional processing at encoding that leads to superior memory performance for that item (Green, 1956; Jenkins & Postman, 1948). Additional attention paid to distinctive items may cause participants to rehearse these items (Rundus, 1971; but see Dunlosky, Hunt, & Clark, 2000), to elaborate on them (Waddill & McDaniel, 1998), to evaluate them as being different (Geraci & Manzano, 2010; Geraci & Rajaram, 2002, 2004), or simply to increase overall processing of the distinctive items relative to the common ones (e.g., Slamecka & Katsaiti, 1987; Watkins, LeCompte, & Kim, 2000; Wollen & Cox, 1981). The mixed-list design is critical in this respect. The idea is that the distinctive item must be placed in a context that causes the item to stand out and to require more processing than the common items in the list. With respect to the bizarreness effect in particular, some have suggested that this effect is driven by enhanced encoding processes due to the imagery requirements of the encoding task (Wollen & Margres, 1987). Unlike other distinctiveness paradigms, in a typical bizarreness paradigm, participants read sentences and are explicitly instructed to form mental images of the events described by the sentences. Thus, Wollen and Margres suggested that bizarre sentences (such as “The BISCUITS screamed when the OVEN jumped out the WINDOW”) are more difficult to image than common sentences (such as “The BISCUITS were visible through the OVEN WINDOW”). The increased effort required to image the bizarre sentences at encoding leads to the recall advantage for the nouns in those sentences in mixed-list designs, in part because people devote more processing to imaging bizarre sentences at the expense of imaging common sentences. In pure-list designs, bizarre sentences do not steal attentional resources from the common sentences.

A general alternative to the encoding theories is that the bizarreness effect (like distinctiveness effects in general) results from processes operating at the time of retrieval. According to one view, distinctive items contain unusual features that provide diagnostic information that aids memory retrieval (Hunt & McDaniel, 1993). Because unusual items typically contain features that are uncommon to the other to-be-retrieved items, they are more easily discriminated from other items that are generated during retrieval (Hunt & McDaniel, 1993). Another idea is that the unusual features of these items may guide access to the bizarre sentences during attempts to retrieve the study episode (Knoedler, Hellwig, & Neath, 1999; Peynircioğlu & Mungan, 1993;

Waddill & McDaniel, 1998). Some evidence has demonstrated the role that retrieval factors play in bizarreness effects in particular. For example, the bizarreness effect occurs more often in free recall than in cued recall tasks, in which the retrieval demands are more minimal (Riefer & LaMay, 1998; Riefer & Rouder, 1992). Thus, retrieval theories posit a primary role for retrieval processes in producing the mixed-list advantage for distinctive items. We note that these theories do not require that the common and bizarre items be processed in the same manner during encoding. They simply propose that the critical process that leads to the memory advantage occurs at the time of retrieval.

Disentangling encoding from retrieval processes

Though encoding and retrieval accounts of distinctiveness effects are conceptually discriminable, it has proven empirically difficult to distinguish between them. Indeed, past studies designed to provide support for encoding or retrieval processes have failed to yield unequivocal interpretations for other types of distinctiveness effects that may share some similarities with the bizarreness effect (e.g., Geraci & Rajaram, 2002, regarding orthographic distinctiveness; Malhotra & Dixit, 1982; McLaughlin, 1968; Smith & Hunt, 2000, regarding the isolation effect). The problem of disentangling encoding from retrieval processes is particularly troublesome for understanding the mixed-list advantage for bizarre sentences. Currently, both encoding and retrieval theories offer explanations for why the bizarreness effect occurs in mixed- but not pure-list designs. It has been difficult to empirically discriminate between these theories because the encoding and retrieval contexts are confounded in nearly all studies of the bizarreness effect. That is, in a mixed-list design, common and bizarre sentences are not only studied together in a single list, they are also retrieved together. In a pure-list design, common and bizarre sentences are studied in separate lists, and then they are also recalled separately. Thus, the standard methodology has not allowed researchers to determine the separate contributions of encoding and retrieval processes to the mixed-list advantage.

McDaniel, Dornburg, and Guynn (2005) attempted to overcome this encoding–retrieval confound by having participants study pure lists containing either all common or all bizarre sentences, and then having participants either recall the lists either together or separately. The common and bizarre sentences were presented on different colored papers (green and purple) in two different fonts (italicized and bold). In the separate-recall condition, participants were instructed first to recall the sentences presented in italicized print on the green paper, and after completing this task they were instructed to recall the sentences presented in bold print on the purple paper (these orders were counterbalanced). The assumption was that in this separate-recall condition, potential retrieval dynamics

relying on distinctive features would be eliminated. In the combined-recall condition, participants were asked to recall all of the sentences. In this condition, distinctive retrieval dynamics, if operative in the bizarreness effect, could be manifested. A significant bizarreness effect was obtained in the combined-recall condition (bizarre = .47, common = .31), even though the items had not been studied together, but no significant effect emerged in the separate-recall conditions (bizarre = .38, common = .31), in which participants had both studied and retrieved the common and bizarreness sentences separately. In a second experiment, a mixed list was used at encoding, in which the first half of the list contained one type of sentence and the second half contained the other type. At retrieval, participants were asked to recall either the entire list or each half of the list successively (with each half corresponding to either the common or the bizarre sentences). Again, a significant bizarreness effect was obtained in the combined (whole-list) recall condition (bizarre = .44, common = .31), but not in the separate (half-list) recall condition (bizarre = .38, common = .30). The results of both experiments were interpreted as evidence that the bizarreness effect is a consequence of retrieving (and not of encoding) the common and bizarre sentences together.

However, the mean performances provided above clearly show that there was a memory advantage for bizarre items in both experiments when the common and bizarre sentences were recalled separately (the separate-retrieval conditions). Moreover, McDaniel et al. (2005) found no significant interaction between sentence type (common and bizarre) and retrieval condition (combined or separate) in either experiment, also suggesting that the bizarreness effect was not differentially influenced by the different retrieval conditions (see also Mulligan & Peterson, 2008, and Peterson & Mulligan, 2010, for failures to find evidence for a retrieval effect in perceptual interference, generation, and enactment effects using McDaniel et al.'s, 2005, encoding/retrieval separation paradigm). Thus, it is not clear that processes that rely on a contrastive retrieval context (through which bizarre items become functionally distinct) play a primary role in mediating the mixed-list bizarreness effect.

Alternatively, it could be that retrieval processes do play a critical role in producing a bizarreness advantage. Because the common and bizarre sentences were studied during the same study episode in the McDaniel et al. (2005) study, it may have been difficult for participants in the separate-recall condition to exclude the other-list items from memory, thereby functionally producing a retrieval context in which both common and bizarre items were somewhat intermixed in a retrieval set.

The present experiment was designed to gain clear evidence for the selective role of retrieval in mediating the bizarreness effect in memory. To do so, we created two separate study episodes using different rooms, cover stories, and

encoding tasks. Participants either recalled all of the items together (regardless of the episode in which they had been studied) or recalled them separately (from just one study episode at a time). If the bizarreness effect depends on processes operative at retrieval, then the effect should be obtained in the together-recall condition, but not in the separate-recall condition (i.e., we should find a significant interaction between sentence type and recall condition, along with virtually identical recall of bizarre and common items in the separate-recall condition). It is worth noting that not only were the tests of the bizarreness effect conducted within subjects, but also a large number of participants were tested, so that power would be high to detect a bizarreness effect if it were present (i.e., in the separate-recall condition). Alternatively, if encoding processes are paramount in bizarreness effects and there is no selective role for retrieval in producing this mixed-list advantage, the recall condition should have no influence on the pattern of effects, and we should find no significant bizarreness effect in either recall condition.

Method

Participants

A total of 192 undergraduate Texas A&M University students participated in exchange for course credit. Participants were tested in groups of one to four.

Design

A 2×2 mixed design was used in which sentence type (bizarre or common) served as the within-subjects variable and recall condition (together or separate) served as the between-subjects variable (with 96 participants assigned to each recall condition).

Materials

The participants studied 48 items in 16 sentences (eight common and eight bizarre). We used a second list of 16 sentences (eight common and eight bizarre) and counterbalanced which list of 16 sentences participants studied. Target nouns were presented in all capital letters in the eight common sentences (e.g., “The CAT knocked over the COFFEE on the SHELF.”) and eight bizarre sentences (e.g., “The MAID licked the AMMONIA off the TABLE.”).

Procedure

To make the two episodic contexts as different from each other as possible, the two types of sentences (common and bizarre)

were studied in different rooms, using different encoding and rating instructions (participants either read and rated the sentences for reading comprehension or read and rated the sentences for vividness). We were not interested in the effects of the specific room or cover story on performance. Rather, these variables were simply included to make the two encoding episodes different from each other. The two rooms (Room A and Room B) had different interiors and were located on different floors of the Psychology building. Room A was a bright and newly refurbished room on the fourth floor that contained four individual computer desks arranged along a wall, separated by dividers. Room B was a darker, less updated room on the main (i.e., second) floor that contained four individual computer desks, arranged on separate walls. In Room B, participants sat with their backs to each other so that they could not see the other participants. Participants began the experiment in either Room A or Room B, depending on the condition. Upon entering the experiment, participants were told that they would be given a series of cognitive tests within the hour. The first test was introduced as either a reading comprehension or an imagery test, depending on the condition. We did not expect the type of rating to interact with memory for common and bizarre sentences [and indeed it did not; $F(1, 190) = 2.41$, $MSE = .08$, $\eta^2_p = .01$]. We expected a bizarreness advantage to emerge in the together-recall condition with both types of ratings. Note that some rating tasks (e.g., rating the degree to which the relationship between the critical nouns in sentences is unusual) eliminate the bizarreness effect (e.g., McDaniel & Einstein, 1986, Exp. 2). Here, our interest was in using encoding instructions that would simply allow us to distinguish the two study episodes from each other. For the reading comprehension condition, participants silently read the sentences, each containing three target nouns, one at a time. They indicated how well they comprehended the sentence using a 5-point scale (1 indicated *very good comprehension*, and 5 indicated *very poor comprehension*). For the imagery condition, participants silently read the sentences one at a time and were asked to rate the vividness of their mental image of the events described by each sentences using a 5-point scale (1 indicated *very vivid*, and 5 indicated *not very vivid*). They entered their ratings using the designated keys on the keyboard. Sentences were presented on the computer using the Cedrus SuperLab program, and participants were allowed to pace their reading and rating of the sentences. During this phase, the participants studied either common or bizarre sentences.

After studying the first set of sentences (either common or bizarre) in the first room (either Room A or Room B), participants were asked to follow the experimenter to a second room located on a different floor of the Psychology building for another cognitive test. During this phase, participants studied the other set of sentences, also on the computer. They read this set of sentences under the guise that this was a different type of

test. After studying and rating this set of sentences, participants were led to a third room for the recall test. This third room was also dissimilar from Rooms A and B; here, participants sat in a small waiting area with four seats and were given clipboards to hold.

In the together-recall condition, participants were asked to recall as many capitalized words as they could from the sentences that they had seen earlier from either the reading comprehension or the imagery experiments. They were reminded that they had read some sentences in the room on the second floor of the building, where they had taken a particular type of test, and other sentences in the room on the fourth floor, where they had taken a different test. The experimenter emphasized that the participants should try to recall words from any of the sentences presented in either of the rooms.

In the separate-recall condition, participants were asked to recall as many of the capitalized words from the first set of sentences studied (and were reminded of the room and the type of task that they had engaged in, similar to the together-recall condition). Participants were warned that they should only recall words presented in that specific room. After completing the first recall test, participants were asked to perform a second memory test for the second set of sentences that they had studied. Thus, in both recall conditions participants took a free recall test. The only difference was whether they were told to recall from both study episodes or from one study episode only.

We counterbalanced the rooms in which the common and bizarre items were studied (Room A vs. Room B), the order of the rooms in which the common and bizarre items were studied, the order of the common and bizarre lists, the cover story associated with each study list (imagery task vs. reading comprehension task), and the sentence type that the nouns appeared in (common or bizarre)

Results

The mean proportions of common and bizarre items recalled in the together- and separate-recall conditions appear in Fig. 1. Significance was set at $p < .05$ for all analyses. A 2×2 mixed analysis of variance (ANOVA) was used to analyze the data, with Sentence Type (bizarre or common) as a within-subjects factor and Recall Condition (together or separate) as a between-subjects factor. We found no difference in total free recall performance as a function of recall condition ($F < 1$). Bizarre items were recalled better than common items in general, $F(1, 190) = 10.92$, $MSE = 0.35$, $\eta^2_p = .05$. This main effect was qualified by a significant interaction between recall condition and sentence type, $F(1, 190) = 5.76$, $MSE = 0.18$, $\eta^2_p = .03$ (see Fig. 1). Planned comparisons showed that participants recalled significantly more words from the bizarre sentences ($M = .38$, $SD = .19$) than from the common

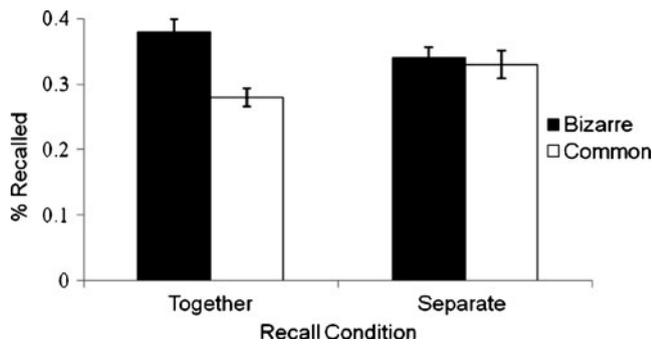


Fig. 1 Mean proportional recall for words from common and bizarre sentences in the together-recall and separate-recall conditions

sentences ($M = .28$, $SD = .14$) in the together-recall condition, $F(1, 190) = 18.17$, $MSE = .52$, $\eta_p^2 = .09$, but not in the separate-recall condition ($F < 1$; bizarre recall = $.34$, $SD = .17$; common recall = $.33$, $SD = .21$). Note that no other comparisons were significantly different.

To further examine the absence of a bizarreness effect in the separate-recall condition, we performed a Bayesian analysis, developed by Wagenmakers (2007; see also Masson, 2011). For this analysis, the null hypothesis (no bizarreness effect) and the alternative hypothesis (a bizarreness effect) were set up as competing models, and Bayesian information criterion (BIC) values were used to estimate a Bayes factor and to generate the posterior probability for each hypothesis. This analysis showed that in the separate-recall condition, the probability of the null (no bizarreness effect) model, $p\text{BIC}(H_0|D)$, was $.91$ (i.e., the probability of an effect was $.09$). In other words, there was a 91 % chance that the null hypothesis was correct—considered strong evidence for the null hypothesis, following guidelines presented by Raftery (1995). By contrast, in the together-recall condition, the probability of the null effect was 0 % (probability of an effect was then 100 %). Thus, using both the standard method of null-hypothesis testing and the Bayesian approach, there is clear evidence of a bizarreness effect in the together-recall condition, but not in the separate-recall condition.

Note that one can also score the data in terms of the number of sentences of each type that are recalled, in which a sentence counts as being recalled if any one word from that sentence is recalled. By this scoring method, if one word, two words, or all three words from a sentence are recalled, then the sentence is scored as being recalled and is given a 1. If no words are recalled, it is scored as not being recalled and is given a 0. This scoring method uses a slightly blunted scoring range (a 0 vs. a 1) relative to the word-scoring method, in which scores range from 0, *no words recalled*, to 3, *all words recalled*. Nonetheless, one can examine recall using this sentence-level method of scoring. Using this sentence level of analysis, the 2 (sentence type) \times 2 (recall condition) ANOVA showed a main effect of sentence type, $F(1, 190) = 18.46$, $MSE = 0.05$, $\eta_p^2 = .09$, showing that people recalled more bizarre sentences

than common sentences. We found no overall effect of recall condition on memory performance, $F(1, 190) < 1$. The interaction between recall condition and sentence type approached significance ($p = .09$), $F(1, 190) = 2.93$, $MSE = 0.05$, $\eta_p^2 = .02$. For the together-recall condition, recall for bizarre sentences was $.54$ ($SD = .19$) and recall for common sentences was $.41$ ($SD = .18$). For the separate-recall condition, the recall for bizarre sentences was $.49$ ($SD = .19$) and the recall for common sentences was $.44$ ($SD = .22$). We again conducted Bayesian analysis on the sentence level to examine the probability at which a bizarreness effect occurred in each recall condition. In the separate-recall condition using the sentence scoring, the probability of an effect was 33 %, not considered to constitute even weak evidence for an effect, whereas in the together-recall condition, the probability of an effect was 100 %, which constitutes very strong evidence for an effect. Thus, using both the word and sentence scoring methods for both standard significance testing and Bayesian analyses, very good evidence emerged that a bizarreness effect was obtained when the items were recalled together and not when they were recalled separately.

Nonetheless, one might still wonder whether encoding processes moderate the effect in the together-recall condition. For example, one might suggest that the bizarreness effect is larger when common items were studied before bizarre items, which could be consistent with the idea that one must first encode common items for them to set up a background against which bizarre items become distinctive. Returning to our main analyses of words recalled, we examined the effect of study order (common list first vs. bizarre list first) on memory for bizarre and common items in the together-recall condition. The results showed a significant main effect of sentence type, $F(1, 93) = 23.81$, $MSE = .04$, $\eta_p^2 = .20$, but no main effect of order, $F(1, 93) < 1$. However, a significant interaction was observed, $F(1, 92) = 21.98$, $MSE = .04$, $\eta_p^2 = .19$: Recall was better for bizarre items when common items had been studied first, as compared to when bizarre items had been studied first ($M_s = .44$ vs. $.33$, respectively), and planned comparisons confirmed this difference, $t(94) = 2.75$, $SE = .03$, $d = 0.46$, which could be taken as evidence that the bizarreness advantage in memory was partially driven by encoding common items prior to bizarre items—that is, evidence for the role of encoding in mediating the bizarreness advantage in memory. However, the data were consistent with the simple idea that more recently encountered information is better remembered. In fact, recall was better for common items when bizarre items had been studied first than when common items had been studied first ($M_s = .32$ vs. $.23$), $t(94) = 3.40$, $SE = .02$, $d = 0.69$. Indeed, the delay between Study List 1 and test was approximately 5 min longer than the delay between Study List 2 and test. Taken as a whole, the order effects on both types of items were most consistent with a simple recency account rather than a selective encoding advantage for bizarre items.

One might also wonder whether a similar type of evidence might emerge for the role of encoding processes in the separate-recall conditions, such that bizarre items would have a memory advantage in separate recall when the common items were studied first. The results showed a significant interaction between order and item type on recall, $F(1, 94) = 14.36$, $MSE = .03$, $\eta_p^2 = .13$, but no main effect of item type or order ($F_s < 1$). Bizarre items were remembered better when they were studied after common items, as compared to when they were studied before common items, $t(94) = 3.73$, $SE = .02$, $d = 0.77$. But again, common items were also remembered better when they were studied after bizarre items than when they were studied before bizarre items, $t(94) = 1.76$, $SE = .03$, $d = 0.36$, although this comparison did not reach significance ($p = .08$). Again, the pattern appears to be most consistent with a simple recency effect: Items that are studied last are remembered better than items that are studied earlier. Taken together, the order analyses do not explain the effect of the recall condition (together or separate) on the presence of the bizarreness advantage in memory. What the results do show is that the bizarreness effect was obtained only in the together-recall condition, thus superseding any potential order effects. However, in the separate-recall condition, a bizarreness effect was not obtained. Thus, it appears that the act of retrieving common and bizarre items together is critical for obtaining a bizarreness advantage in memory.

Discussion

The finding of a bizarreness effect in the together- and not the separate-recall conditions demonstrates a significant role for retrieval factors in obtaining the bizarreness effect. These results suggest that, for a bizarreness advantage to occur in memory, common and bizarre items must be recalled together. This study is the first to demonstrate the selective role of retrieval in mediating the bizarreness effect. As such, the present results extend in two ways the findings reported by McDaniel et al. (2005) reviewed at the outset. First, the bizarreness effect interacted with recall condition, and second, the effect was eliminated in the separate-recall condition, only emerging in the together-recall condition. Note that the latter finding is unusual, because bizarreness effects are typically not reported when bizarre and common sentences are studied in pure lists (as in the present experiment; see McDaniel et al., 2005; McDaniel & Einstein, 1986), and indeed, in pure-list designs common items can be recalled better than bizarre items (see McDaniel & Bugg, 2008). However, in the standard paradigm, when bizarre and common items are studied in pure lists, they are also recalled separately. In contrast, in the present paradigm, the together-recall condition required that the pure lists of bizarre and

common items be considered in the same retrieval set. In this situation, the emergence of a significant bizarreness effect suggests that the effect depends on retrieving the common and bizarre sentences together. The effect does not appear to depend on enhanced encoding, because we found no memory advantage for bizarre sentences when recall was separate.

Although we can conclude that retrieval processes play a critical role in obtaining a bizarreness effect with mixed-list designs, we cannot specify exactly how this occurs. Several possibilities, however, can be ruled out by the present experiment. One possibility is that the bizarreness effect occurs when common items are retrieved first. However, the data are most consistent with a simple recency effect, as the sentences studied last were remembered better, regardless of whether they contained common or bizarre items. In addition, despite the fact that order contributed to the recall levels in both the together- and separate-recall conditions, an overall advantage of bizarreness only occurred in the together-recall condition.

Another possibility is that the bizarreness effect occurs in together recall because bizarre items are retrieved before common items, resulting in more output interference for common items. However, when we compared the average output order for common and bizarre items in the together-recall condition, the results showed almost identical average rank order recall for common (9.12) and bizarre (9.57) items, $t(91) = 1.01$ (the missing data are from participants who did not recall any words of one type of sentence). Thus, we found no evidence that the bizarreness effect occurred at retrieval because common items were retrieved later, and therefore suffered more output interference.

A third possibility is that bizarre items steal resources from common items when they are recalled together—not in the form of output interference, necessarily, but in terms of the amount of attention or processing given to common items when bizarre items are recalled alongside them. If this hypothesis is correct, then one would expect memory for common items to suffer significantly in the together-recall condition relative to the separate-recall condition. Some evidence for this suggestion did emerge, as participants recalled fewer common items in the together-recall condition (.28) than in separate-recall condition (.33), $t(190) = 1.91$, $SE = .03$, although this difference was just shy of significance ($p = .06$). We also found a numerical difference between the proportions of bizarre items recalled in the together-recall condition (.38) and the separate-recall condition (.34), $t(190) = 1.48$, $SE = .03$, although this difference was not significant. Identifying the specific retrieval mechanism involved awaits additional research. For now, the key contribution of the present study is that it is the first to unequivocally demonstrate that retrieval processes play a selective role in producing the bizarreness effect in memory.

The present results do not preclude the possibility that bizarre sentences may also stimulate additional processes at encoding when common sentences are intermixed with bizarre sentences during study. It should also be noted that according to the strong theoretical position that encoding and retrieval processes are always interactive (cf. Tulving, 1983), it is theoretically possible that bizarre sentences may have stimulated additional processing at encoding relative to common sentences even in the present paradigm, but that these differences were only functional in recall when the bizarre and common items were recalled together. However, an advocate of this position would need to articulate why such encoding differences were not revealed when retrieval was separate. Accordingly, the more parsimonious implication of the present results is that differential encoding processes are not critical for obtaining the bizarreness effect, and that such processes may not even be necessary for the bizarreness effects found in mixed-list designs. As we will develop below, some theoretical work has converged on the idea that retrieval processes appear to be a sufficient mechanism for understanding bizarreness effects in memory.

Our interpretation is generally consistent with several theories of distinctiveness that emphasize the role of retrieval processes and the relative retrieval advantage of distinctive items that are recalled alongside more common items (Bruce & Gaines, 1976; Hunt & McDaniel, 1993). For example, the present findings are consistent with studies showing a critical role for retrieval processes in the appearance of other distinctiveness effects (the orthographic distinctiveness effect, the isolation effect, and the picture superiority effect), which shows that these effects are influenced by the type of processes reinstated at retrieval, such as when using different types of implicit memory tests (Geraci & Rajaram, 2002, 2004; Hamilton & Geraci, 2006, respectively). The present findings could also be consistent with Smith and Hunt's (2000) suggestion that distinctiveness effects emerge in memory because the recall task reinstates the context in which common items were studied, which then serves to provide a background against which the unusual items appear distinct during retrieval.

The lack of a bizarreness effect in the separate-recall condition is not, however, accommodated well by some formal distinctiveness models of memory. These models suggest that free recall (retrieval) is a discrimination problem, such that an item is retrieved most easily when that item is distinctive relative to other traces in memory (see, e.g., Brown, Neath, & Chater, 2007). Global models assume that all items in memory influence the calculation of the distinctiveness of each item (see Neath & Brown, 2007). According to this kind of model, bizarre items are distinctive, and thus are recalled well relative to common items, regardless of recall condition. Local-distinctiveness models also do not accommodate the present pattern of results, because these models assert that

only the nearest neighbors in the target set influence the distinctiveness of an item, which would presumably include other bizarre sentences. According to this theory, the distinctiveness of a bizarre item would be relatively low, and its effects should be equivalent across the separate- and the together-recall conditions (both of which have all of the bizarre items in the retrieval set). Yet, the bizarreness effect emerged in the together- but not in the separate-recall condition.

By contrast, the present pattern of results dovetails nicely with intermediate models that assume that distinctiveness is determined by the set of relevant items bounded by the retrieval request (in this case, the set of items presented in particular contexts during the experiment). This entire set of items would then determine the distinctiveness of an item (by considering all items in the retrieval set, this approach does not appear to be captured by the most recent local distinctiveness model, SIMPLE [Brown et al., 2007], although particular parameter settings for the breadth of the neighborhoods in which candidate items are considered might possibly accommodate the present results). According to this more intermediate distinctiveness model, bizarre items become distinctive (and therefore well-remembered) when they are considered in the context of a retrieval set of common and bizarre items (the together-recall condition) but not when they are considered in a retrieval set of other bizarre items (the separate-recall condition). The present findings are consistent with this kind of distinctiveness memory model, which defines distinctiveness in terms of the retrieval set, with distinctiveness in turn determining free recall (cf. Brown et al., 2007). The idea that memory performance is determined by functional retrieval set is also a key aspect of some general theories of memory, such as the feature model of memory (Nairne, 1990, 2002). Also, the importance of the retrieval set is highlighted in some definitions of distinctiveness. According to Nairne (2006), distinctiveness is defined as

the extent to which a particular cue (or set of cues) specifies a particular stored event to the exclusion of others. Framed in this way, distinctiveness is not a fixed property of a cue, or a target trace, or even of an interaction between a given cue and a given target. It is a property of a cue in context: given a fixed set of alternatives, a measure of distinctiveness can be assigned to a particular cue with respect to a particular alternative. Change the context—for example, by changing how the cue is perceived or the range of possible responses—and the measure of distinctiveness changes as well. (p. 27)

The present finding—showing that the bizarreness advantage only occurs when the retrieval cue specifies that common and bizarre items be recalled together—is accommodated well by this view of distinctiveness.

An alternative theoretical approach might be offered for explaining why retrieving common and bizarre items together produces the bizarreness effect: One idea is that bizarre items are remembered better than common items in mixed-list designs in part because the recall of common items is compromised by disrupted encoding of order information when bizarre items are also present in the list (see McDaniel & Bugg, 2008, for this idea applied to the bizarreness effect; Nairne, Riegler, & Serra, 1991, and Serra & Nairne, 1993, for this original order disruption theory applied to the generation effect; and Mulligan, 2000, for this idea applied to the perceptual interference effect). Although this posited disruption in the processing of order information that results from including bizarre items in a study list has been assumed to occur at encoding (McDaniel & Bugg, 2008), it is possible that people do encode item order information, but that bizarreness information discourages the use of this order information at retrieval (McDaniel, DeLosh, & Merritt, 2000; see McDaniel et al., 2011, for similar findings and a discussion regarding orthographic distinctiveness effects).

In the present study, the idea is that in the together-recall condition, the inclusion of bizarre items during recall might have disrupted the use of order information that would otherwise have been exploited for recalling the common items (thereby producing a bizarreness effect in the together-recall but not in the separate-recall condition). If so, item order memory for common items should be better in the separate-recall condition than in the together-recall condition. To analyze whether the order of common items was affected by whether they were retrieved with or without bizarre items, we used an input–output analysis (Asch & Ebenholtz, 1962). According to this method, order memory is calculated by counting the number of pairs of adjacent items (in this case, pairs of sentences) recalled in the correct serial order in which they were studied out of the number of adjacent pairs of items recalled in any order. The results showed that input–output correspondence scores were at chance in both the together-recall and the separate-recall conditions. Accordingly, it remains unclear whether modulation of the use of order memory in recall (for common items) was related to the emergence of the bizarreness effect in the together-recall condition. However, as recall for the common items was not strongly attenuated by recalling items together, relative to separately, the results do not readily submit to an interpretation based on the order hypothesis.

Regardless of the particular theoretical approach adopted, more generally, the present results emphasize the importance of examining retrieval processes for understanding the bizarreness effect. This view is consistent, not only with the various theories of distinctiveness noted above, but also with general theories of memory. The notion that “the key to understanding memory is understanding retrieval” (Roediger, 2002, p. 25) has been emphasized by several researchers. The present findings

add another wrinkle to the general idea that memory performance depends on the nature of the retrieval cue. Here, encoding conditions were held constant, as were retrieval conditions (all participants were given a free recall test), but participants were simply told to retrieve from different sets of items (either all or only specific studied items). Thus, varying the retrieval set appears to significantly influence memory performance for identically encoded items. This finding is not unlike findings from implicit memory research in which test cues are held constant (see Roediger & Geraci, 2005, for a review), but performance varies dramatically depending on whether participants are given explicit or implicit memory instructions (which change the retrieval set from the study items [explicit instructions] to all possible words [implicit instructions]). Similarly, the present study showed that memory only benefits from bizarreness when participants retrieve items from both common and bizarre retrieval sets, rather than just the common or bizarre set.

In sum, the present experiment establishes that differential encoding processes are not needed in explanations of the bizarreness effect. Rather, retrieval processes operating when common and bizarre items are combined in the retrieval set play a unique and important role in mediating the bizarreness effect. Given that in mixed-list designs the retrieval set combines common and bizarre items, it appears that retrieval dynamics are sufficient to produce the oft-reported bizarreness advantage in these designs.

Author Note This work was supported by the National Institute on Aging (RO1 AG039502 to L.G.).

Correspondence should be sent to Lisa Geraci, Department of Psychology, Texas A&M University, College Station, TX 77843. Email: lgeraci@tamu.edu.

References

- Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, *106*, 135–163.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576. doi:10.1037/0033-295X.114.3.539
- Bruce, D., & Gaines, M. T. (1976). Tests of an organizational hypothesis of isolation effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *15*, 59–72.
- Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 649–657.
- Einstein, G. O., & McDaniel, M. A. (1987). Distinctiveness and the mnemonic benefits of bizarre imagery. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 78–102). New York, NY: Springer.
- Geraci, L., & Manzano, I. (2010). Distinctive items are salient at encoding: Delayed judgments of learning predict the isolation effect. *Quarterly Journal of Experimental Psychology*, *63*, 50–64.

- Geraci, L., & Rajaram, S. (2002). The orthographic distinctiveness effect on direct and indirect tests of memory: Delineating the awareness and processing requirements. *Journal of Memory and Language*, *47*, 273–291.
- Geraci, L., & Rajaram, S. (2004). The distinctiveness effect in the absence of conscious recollection: Evidence from conceptual priming. *Journal of Memory and Language*, *51*, 217–230.
- Green, R. T. (1956). Surprise as a factor in the von Restorff effect. *Journal of Experimental Psychology*, *52*, 340–344.
- Hamilton, M., & Geraci, L. (2006). The picture superiority effect in conceptual implicit memory: The importance of cuing a picture's distinctive features. *The American Journal of Psychology*, *119*, 1–20.
- Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*, 105–112. doi:10.3758/BF03214414
- Hunt, R. R., & Elliott, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, *109*, 49–74.
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445.
- Hunt, R. R., & Toth, J. P. (1990). Perceptual identification, fragment completion, and free recall: Concepts and data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 282–290.
- Jenkins, W. O., & Postman, L. (1948). Isolation and “spread of effect” in serial learning. *The American Journal of Psychology*, *61*, 214–221.
- Knoedler, A. J., Hellwig, K. A., & Neath, I. (1999). The shift from recency to primacy with increasing delay. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 474–487.
- Malhotra, A., & Dixit, R. C. (1982). Recall and recognition measures of the Von Restorff effect and the serial position of the item. *Archivio di Psicologia, Neurologia e Psichiatria*, *43*, 378–389.
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679–690. doi:10.3758/s13428-010-0049-5
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*, 237–255. doi:10.3758/PBR.15.2.237
- McDaniel, M. A., Cahill, M., Bugg, J. M., & Meadow, N. G. (2011). Dissociative effects of orthographic distinctiveness in pure and mixed lists: An item-order account. *Memory & Cognition*, *39*, 1162–1173. doi:10.3758/s13421-011-0097-9
- McDaniel, M. A., DeLosh, E. L., & Merritt, P. S. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1045–1056. doi:10.1037/0278-7393.26.4.1045
- McDaniel, M. A., Dornburg, C. C., & Guynn, M. J. (2005). Disentangling encoding versus retrieval explanations of the bizarreness effect: Implications for distinctiveness. *Memory & Cognition*, *33*, 270–279. doi:10.3758/BF03195316
- McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 54–65. doi:10.1037/0278-7393.12.1.54
- McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 422–435.
- McDaniel, M. A., & Geraci, L. (2006). Distinctiveness and the mnemonic benefits of bizarre imagery. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 65–88). New York, NY: Oxford University Press.
- McLaughlin, J. P. (1968). Recall and recognition measures of the von Restorff effect in serial learning. *Journal of Experimental Psychology*, *78*, 99–102.
- Mulligan, N. W. (2000). Perceptual interference and memory for order. *Journal of Memory and Language*, *43*, 680–697.
- Mulligan, N. W., & Peterson, D. (2008). Assessing a retrieval account of the generation and perceptual-interference effects. *Memory & Cognition*, *36*, 1371–1382. doi:10.3758/MC.36.8.1371
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269. doi:10.3758/BF03213879
- Nairne, J. S. (2002). The myth of the encoding–retrieval match. *Memory*, *10*, 389–395. doi:10.1080/09658210244000216
- Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46). Oxford, UK: Oxford University Press.
- Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 702–709. doi:10.1037/0278-7393.17.4.702
- Neath, I., & Brown, G. D. (2007). Making distinctiveness models of memory distinct. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 125–140). New York, NY: Psychology Press.
- Peterson, D., & Mulligan, N. W. (2010). Enactment and retrieval. *Memory & Cognition*, *38*, 233–243.
- Peynircioglu, Z. F., & Mungan, E. (1993). Familiarity, relative distinctiveness, and the generation effect. *Memory & Cognition*, *21*, 367–374.
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, MA: Blackwell.
- Riefer, D. M., & LaMay, M. L. (1998). Memory for common and bizarre stimuli: A storage–retrieval analysis. *Psychonomic Bulletin & Review*, *5*, 312–317.
- Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*, 601–611.
- Roediger, H. L., III. (2002). An interview with Henry Roediger. *Educational Psychology Review*, *14*, 1–30.
- Roediger, H. L., III, & Geraci, L. (2005). Implicit memory tasks in cognitive research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 129–151). Washington DC: American Psychological Association Press.
- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63–77. doi:10.1037/h0031185
- Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, *21*, 34–40.
- Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, *26*, 589–607.
- Smith, R. E., & Hunt, R. R. (2000). The effects of distinctiveness require reinstatement of organization: The importance of intentional memory instructions. *Journal of Memory and Language*, *43*, 431–446.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Oxford University Press, Clarendon Press.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld: Analyse von Vorgängen im Spurenfeld. *Psychologische Forschung*, *18*, 299–342. doi:10.1007/BF02409636
- Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition*, *26*, 108–120.

- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779–804. doi:10.3758/BF03194105
- Watkins, M. J., LeCompte, D. C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 239–245.
- Wollen, K. A., & Cox, S. D. (1981). The bizarreness effect in a multitrial intentional learning task. *Bulletin of the Psychonomic Society*, *18*, 296–298.
- Wollen, K. A., & Margres, M. G. (1987). Bizarreness and the imagery multiprocess model. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 103–128). New York, NY: Springer.
- Worthen, J. B. (2006). Resolution of discrepant memory strengths: An explanation of the effects of bizarreness on memory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 133–156). New York, NY: Oxford University Press.