Psychonomic Bulletin & Review
2008, 15 (2), 237-255
doi: 10.3758/PBR.15.2.237

Theoretical and Review Articles

Instability in memory phenomena: A common puzzle and a unifying explanation

Mark A. McDaniel and Julie M. Bugg
Washington University, St. Louis, Missouri

In mixed lists, stable free recall advantages are observed for encoding conditions that are unusual, bizarre, or attract extensive individual item elaboration relative to more common encoding conditions; but this recall advantage is often eliminated or reversed in pure lists. We attempt to explain this ubiquitous memory puzzle with an item-order account that assumes that (1) free recall of unrelated lists depends on order and item information; (2) unusual items attract greater individual item-processing but disrupt order encoding regardless of list composition; and (3) list composition determines differences in order encoding across unusual and common items. We show that the item-order account provides a unifying explanation of five memory phenomena for which the requisite data exist. The account also successfully anticipates pure-list reversals, in which the standard *mixed*-list recall pattern is obtained in pure, structured lists, a finding that competing accounts cannot handle. Extending the item-order account to other "established" recall phenomena may prove fruitful.

Perhaps because general laws of memory are rare at best (Roediger, 2008), the memory field has oriented toward identifying and understanding various phenomena. Phenomena that are well known to students and researchers of memory include the *generation effect* (items that are generated are recalled better than items that are read—Slamecka & Graf, 1978; see also Jacoby, 1978), the *word frequency effect* (high-frequency items are better recalled than are low-frequency ones; Deese, 1960; Gregg, 1976; Hall, 1954), and the *bizarreness effect* (items rendered in a bizarre image are better recalled than those rendered in a common image; for reviews, see Einstein & McDaniel, 1987; Worthen, 2006). In this article we establish that even these and many other "stable" recall phenomena are not as straightforward as many believe. We identify a single experimental factor, one not frequently appreciated, that robustly modulates, and in many cases reverses, the pattern commonly viewed as inviolable. We then propose a framework to unify and explain the effects of this factor on a wide range of recall phenomena.

A novel aspect of our framework is that it highlights the importance of a memory component that has received relatively little attention in the free-recall literature, although it is ubiquitous in daily memory functioning: serial order information. Detailed predictions are generated that focus on the encoding and use of order information and its relation to the free-recall patterns of interest. For five diverse free-recall phenomena that have received extensive interest in the literature, evidence is available to evaluate some or all of these predictions. We consider such evidence from both the perspective of the framework developed herein and from the perspective of an alternative and popular construct, that of distinctiveness (see Hunt & Worthen, 2006, for a recent volume on distinctiveness), that has been invoked to explain many of these findings.

Effects in Free Recall Depend on List Composition

An underappreciated yet widespread finding is that the effects of a given encoding condition on free recall frequently depend on list composition; that is, many effects believed to be stable in actuality hold true only when the manipulation is instantiated with a particular list composition. When the encoding condition is varied within lists, such that both levels of encoding are intermixed in one list, the "standard" effect (for free recall) is typically observed. For instance, the generation effect, the *enactment effect* (subject enacted actions recalled better than experimenter enacted actions), the *perceptual interference effect* (perceptually masked words recalled better than unmasked words), and the bizarreness effect are reported with mixed lists of items. In contrast, when the encoding condition is varied between lists such that pure lists of items are used, the puzzling result is that the just-mentioned recall effects are eliminated or even reversed. (For word frequency, the well-known superiority in free recall of high-frequency over low-frequency words obtains in pure lists but typically not in mixed lists.)

M. A. McDaniel, mmcdanie@artsci.wustl.edu

 Copyright 2008 Psychonomic Society, Inc.

Moreover, the effects of list composition are not limited to the phenomena specified above. Table 1 summarizes a wide range of encoding conditions for which recall effects are modulated by list composition. As can be seen from the table, this pattern appears to be ubiquitous, which makes it all the more surprising that few researchers have appreciated and systematically investigated the underlying causes of these effects. There are five phenomena, however, for which sufficient evidence is available to begin to construct an understanding of the puzzling instability in the patterns observed as a consequence of list composition: the generation, word frequency, enactment, perceptual interference, and bizarreness effects. Using these literatures, we test a unifying explanation that we propose for pure- and mixed-list recall patterns: the item-order account. As will be demonstrated, this account provides a compelling synthesis of these separate literatures based on item-elaboration and order-memory dynamics, a synthesis that we believe could also fruitfully account for the numerous effects documented in Table 1.

The Item-Order Framework

We now describe a theoretical framework for free recall that provides an overarching and parsimonious account for why mixed and pure-list manipulations produce disparate patterns for the variety of manipulations reviewed above. Our basic assumption is that information about individual list elements (more generally individual events) and information about the relations among list elements (events) jointly contribute to free recall. This assumption is based on a wealth of evidence indicating that both organization (a form of relational processing) and elaboration of individual list items (e.g., levels of processing) improve free recall (e.g., Craik & Tulving, 1975; Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993; Mandler, 1969).

Importantly, these two kinds of information appear to have distinguishable functions in free recall. One specific formulation is that relational processing provides structure to guide retrieval of target candidates, and item processing provides information that allows discrimination of possible list candidates from items not presented in the experimental context (see Hunt & Einstein, 1981, and Hunt & McDaniel, 1993, for reviews and detailed development of these ideas). This formulation is similar to a generate–recognize approach to free recall, an approach that remains viable with appropriate theoretical modifications (Guynn & McDaniel, 1999; Jacoby & Hollingshead, 1990; see also Kintsch, 1978). Another formulation is that relational information allows an initial characterization of target items that is general and somewhat underspecified, and that item information allows this general characterization to be fleshed out and fully specified (see Gillund & Shiffrin, 1984, for a similar idea in the search of associative memory [SAM] model of recall). For present purposes, we need not favor one of these formulations over the other. The major point is that relational information and item information have different functional roles in recall, with both contributing to free-recall performance.

Order Information

A second critical assumption of our theoretical account is that serial order information is a particular kind of relational information routinely encoded for episodic

Table 1
Conditions for Which Free Recall Effects Vary for Pure- and Mixed-List Designs

Illustrative Study	Encoding Condition	List Composition: Pure	List Composition: Mixed
Gregg, Montgomery, & Castaño (1980)	High frequency	.56	.48
	Low frequency	.45	.52
Hunt & Elliott (1980)	Orthographically common	.44	.36
	Orthographically distinct	.44	.52
McDaniel & Einstein (1986)	Common sentences	.55	.38
	Bizarre sentences	.52	.62
Hirshman & Bjork (1988)	Read	.45	.41
	Generate	.43	.86
Ransdell & Fischler (1989)*	Abstract prose	.55	.42
	Concrete prose	.61	.71
Zucco, Traversa, & Cornoldi (1984)	Simple pictures	.68	.59
	Complex pictures	.67	.72
Schmidt (1994)	Nonhumorous sentences	.28	.23
	Humorous sentences	.28	.30
Engelkamp & Zimmer (1997)	Experimenter-enacted actions	.65	.43
	Subject-enacted actions	.66	.62
Mulligan (1999; pure);	Intact words	.53	.17
Hirshman & Mulligan (1991; mixed)	Perceptual interference	.48	.26
Hadley & MacKay (2006)	Neutral words	.45	.36
	Taboo words	.44	.45
Zaromb (2007)	Precued prose	.36	.37
	10-sec delay cued prose	.39	.50

Note—Where possible, we offer the results of a single experiment pertaining to each encoding condition. Note that additional studies showing a similar pattern exist in some cases. *The advantage for concrete prose in pure lists is not statistically significant.

events. Outside of the laboratory, organization of events in terms of temporal (serial) order appears to be a property of episodic memory. For instance, retrieval of the temporal ordering of autobiographical events that occur 5 to 8 days apart remains at an above-chance level for up to 9 months (Linton, 1975). Furthermore, for autobiographical events that occurred 20 years earlier, and for which some detail could be recalled, correlations between actual and estimated date and time are reasonably high (r = .47–.72; White, 2002). Neuropsychological theories also embrace the idea that the memory system, in particular medial temporal structures, has a general tendency to bind events into temporal sequences (e.g., Eichenbaum, 2004).

Importantly, we suggest that encoding of order persists in the context of the laboratory, even for relatively impoverished events such as words presented in lists. Although the research on memory for order information for laboratory materials is sparse, the extant findings indicate that participants do encode information about the serial presentation order of target items (Hinrichs, 1970; Tulving & Madigan, 1970; Tzeng, 1976; Tzeng, Lee, & Wetzel, 1979; Underwood, 1969). Further, order memory is observed across a range of instructional conditions. When participants are not forewarned about an order memory test per se, serial position judgments are as accurate as when participants are forewarned (Toglia & Kimble, 1976). When the word lists are as long as 50 items, serial position judgments evidence retention of order information (Toglia & Kimble, 1976; see also Tzeng, 1976). Findings such as these support the claim that "a temporal code about each item is part and parcel of what the subject stores about the material to be remembered" (Tzeng et al., 1979, p. 53).

Our third essential assumption is that participants can rely on this order information to guide free recall. The idea here is that in the absence of salient relational information (such as taxonomic relations among list items), order information provides organizational structure to support search of target items. Put another way, order information provides initial cues to help access target items (cf. Tulving & Pearlstone, 1966). Because the free-recall literature has focused almost exclusively on accuracy of recall for list items per se, the role of order information in free recall may have been underappreciated. However, when recall order has been examined, studies have shown that the output of items in recall corresponds to the original presentation order (Kintsch, 1970b; Mandler, 1969; see also Postman, 1972), suggesting that serial order is guiding recall. Often these patterns have been reported in multiple study–recall trial paradigms with constant presentation order (see Kintsch, 1970b; Mandler, 1969; Mandler & Dean, 1969). For single-trial free-recall paradigms considered herein, the evidence that order guides recall is more mixed (further discussion is provided in the Limits and Boundary Conditions section below; see also Mulligan & Lozito, 2007). Nevertheless, even in single-trial free-recall paradigms, some correspondence between presentation order and output order is observed with lists as long as 32 items (Burns, 1996). Additionally, as noted above, this reliance on order in structuring recall is most evident for lists of unrelated items (Burns, 1996; Postman, 1972).

In considering order information, one can distinguish between encoding and use of *absolute order* versus *relative order*. Absolute order is reflected in memory for the precise location of the item in the list, such as "RAKE was the third item in the list." Relative order does not retain the precise location in the list; instead, the order of the item is remembered relative to its appearance with other list items (RAKE followed TABLE). In the immediate memory span literature (or short-term memory paradigms), healthy debate exists regarding whether order is encoded absolutely or relatively. Some prominent models assume absolute order encoding (see Estes, 1972; Lee & Estes, 1981; Nairne, 1992; see also Fuchs, 1969), whereas other approaches emphasize encoding of relative order (Lewandowsky & Murdock, 1989; Tzeng et al., 1979). For paradigms requiring free recall from long-term memory, less is known. Our account remains open on this issue. For the findings that we review below to evaluate the present theoretical approach, we acknowledge at the outset that absolute or relative order information or both could be playing a role in free recall.

Differential Encoding of Item and Order Information As a Function of Stimulus Type

To recapitulate, our view is that free recall is a function of the individual-item information and the relational information encoded during study. For unstructured lists of items, serial order is the primary form of relational information encoded. Our final assumption is that the nature of the list items differentially influences the encoding of item and order information. For usual, common, or run-of-the-mill stimuli, we suggest that serial order information is noticed and encoded by the learner. Examples of such stimuli include high-frequency words, words presented in intact form (rather than fragmented), and sentences that describe typical or straightforward events (e.g., *The dog chased the bicycle down the street*).

In contrast, we suggest that unusual, less common, or atypical stimuli lure attention in the service of their interpretation. This processing augments the elaboration of the characteristics pertaining to each stimulus (relative to common stimuli), resulting in rich encoding of the individual-item features of the stimuli. However, the focus on the individual features comes at the expense of encoding information regarding the serial order of the stimuli. The idea is that unusual or uncommon items attract focus to the individual item, thereby distracting learners from encoding the order in which the items are presented. (For purposes of exposition, we will label these kinds of stimuli *unusual.*)

Consider now the situation in which common and unusual items are intermixed in a single list. In this case, the level of order encoding ordinarily associated with a particular type of item will be modulated by the presence of the alternative item type. This is because serial order information for an item in a particular position is necessarily influenced by the degree to which serial order information for neighboring items is intact. Accordingly, in a mixed list, the order encoding associated with common items will be somewhat disrupted relative to pure lists, whereas the order encoding for unusual items will be somewhat improved (relative to pure lists).

The foregoing observations are a generalization of seminal work by Nairne and colleagues regarding the generation effect (Nairne, Riegler, & Serra, 1991; Serra & Nairne, 1993). Along the lines outlined above, Nairne et al. (1991; Serra & Nairne, 1993) proposed that generating target items from word fragments produces trade-offs relative to reading target items in the encoding of individual item and order information. In support of this claim, Nairne et al. (1991, see also Burns, Curti, & Lavin, 1993; Mulligan, 2002) found that lists of generated items produced recognition performance superior (reflecting good item encoding) to that of comparable lists of read items, whereas the lists of read items produced more accurate memory for serial order than did the lists of generated items. In mixed lists, memory for serial order was approximately the same for generated and read items, and the levels were intermediate to those found in pure lists.

Following from this work, a parallel pattern has been reported when words are presented in the presence of perceptual interference. Lists of words presented with perceptual interference showed recognition superior to that for the same words presented normally (without perceptual interference); in contrast, order memory was better for the normal presentation relative to the presentation with perceptual interference (Mulligan, 2000a). Note that both the generation-effect and perceptual interference manipulations dovetail with the general ideas presented above. Less typical presentation formats or stimuli attract or require attention for individual item processing and reduce encoding of order information.

Theoretical Predictions

We have now laid the groundwork for a unifying theoretical account of the effects of list composition over a wide range of encoding conditions. Our set of testable predictions follows, focusing first on order encoding and item encoding, and then proceeding to free recall. For ease of accessibility, the predictions are also listed in Table 2. (1) For pure lists of unusual items (or presentations that attract attention to details of each item), memory for serial order will be attenuated relative to pure lists of common items. (2) Unusual items, however, will enjoy superior item elaboration relative to common items. (3) For mixed lists, the levels of serial order information encoded for common and unusual items will approach each other, becoming approximately equivalent. (4) The levels of order information in mixed lists will be intermediate to the order encoding evidenced for common and unusual items in pure lists. (5) Superior item elaboration will remain for unusual items.

These dynamics provide the underpinnings of a principled and parsimonious account of why list composition alters the pattern of free-recall effects produced by the kinds of manipulations reviewed at the outset. In pure lists, our framework anticipates that the superior order encoding enjoyed by common items will provide an advantage in recall relative to unusual items. This is because order information is used to help guide recall, thereby conferring an advantage to pure lists of common items in accessing or generating candidates for free recall. These ideas lead to the following predictions. (6) In pure lists, for common items there should be significant correspondence between the presentation order of the items and the order in which they are recalled. (7) Input–output correspondence for unusual items will be minimal, and less pronounced than observed for common items. Note that in some cases alternative strategies for recalling unusual items (e.g., Knoedler, Hellwig, & Neath, 1999) may offset the order encoding advantage for common items, allowing benefits of superior item elaboration for unusual items to produce approximately equivalent free recall to common items (e.g., McDaniel, DeLosh, & Merritt, 2000).

The primary notions captured above are that common and unusual items are on equal footing with regard to order information in mixed but not in pure lists, and that, therefore, the richer individual-item elaboration for unusual items in mixed lists can be manifested as an advantage in free recall. This pattern is anticipated by the present framework as a consequence of the following changes in recall levels for particular item types. (8) For common items, recall will diminish in mixed lists relative to pure lists (because of diminished order memory). (9) For unusual items, recall will improve in mixed lists relative to pure lists (because of enhanced order memory in mixed lists for unusual items).

For five memory phenomena in which free-recall patterns diverge across pure and mixed list manipulations, the literature has the requisite order memory and recall data to test most or all of the predictions developed above. For each phenomenon, the following sections review these data and evaluate the fruitfulness of the present theoretical account (for ease of exposition, we label it the *item-order account*). Additionally, extant theoretical explanations of each phenomenon are shown to be wanting in light of the comprehensive pattern of evidence assembled.

Table 2
Nine Predictions of the Item-Order Encoding Account

(1) Serial order: pure unusual < pure common
(2) Item elaboration: pure unusual > pure common
(3) Serial order: mixed unusual = mixed common
(4) Order information: pure unusual < mixed < pure common
(5) Item elaboration: mixed unusual > mixed common
(6) Significant input–output correspondence for pure common during recall
(7) Minimal input–output correspondence for pure unusual during recall (< pure common)
(8) Recall: mixed common < pure common (because of diminished order memory)
(9) Recall: mixed unusual > pure unusual (because of enhanced order memory)

The Generation Effect

The seminal ideas for the present framework originated with attempts to explain the effects of generation on memory; accordingly, we begin by examining the generation effect, which refers to the mnemonic advantage for items produced by participants, as in a word stem completion task, compared with items provided by the experimenter, as in a word reading task. Early controversy stemmed from the perplexing finding of a reduction or elimination of the free-recall advantage for generated items in the context of pure lists (Hirshman & Bjork, 1988; McDaniel, Waddill, & Einstein, 1988; Slamecka & Katsaiti, 1987). A series of studies followed that provides data that permit us to examine facets of the item-order account (see Table 3).

In an initial study, Nairne et al. (1991) tested the hypothesis that generation differentially influences item and order information in pure lists. Nairne et al. employed a recognition test to assess item elaboration and an order reconstruction test to assess order encoding. Participants were administered pure generate and pure read word lists. Following each list, participants were required to reconstruct the lists by placing the items from the list in their appropriate presentation positions, or to simply prepare for the next list. Consistent with Prediction 1 of the present framework (see Table 2), order memory, as assessed by performance on the reconstruction test, was superior for pure lists of read items (i.e., common intact words) as compared with pure lists of generated items (i.e., unusual fragmented words). In contrast, item elaboration, as measured by performance on a final recognition test for the lists that participants were not required to reconstruct, was superior for pure lists of generated items as compared with pure lists of read items. Critically, this finding is consistent with Prediction 2 of the order encoding account. Taken together, the reconstruction and recognition results confirm the dissociative effects of generation on item and order memory in pure lists; that is, the findings are consistent with the idea that item memory was enhanced whereas serial order memory was attenuated for generated items in pure lists.

A key set of predictions that stems from the item-order framework concerns the pattern of findings obtained in a mixed-list design. Serra and Nairne (1993, Experiment 1) tested the item-order account by presenting participants with pure read, pure generate, and mixed lists, followed by an order reconstruction task. Reconstruction was significantly better for pure read than for pure generated lists, replicating Nairne et al. (1991). For the mixed lists, order reconstruction was equivalent for read and generated items, and the mixed-list reconstruction scores fell between scores for pure read (highest order reconstruction) and pure generate (lowest order reconstruction) lists. This pattern is perfectly in line with Predictions 3 and 4 of the item-order account and the notion that generation reduces order memory for the read items in a mixed list, and similarly read items bolster order memory for generated items in a mixed list. These dynamics roughly equate order memory for read and generated items and produce an overall drop in order memory for mixed lists to an intermediate level between the pure lists. This order memory pattern appears to be robust, since it is also obtained when a free-recall test is administered prior to the order reconstruction task (Serra & Nairne, 1993, Experiment 2).

Of additional relevance to the item-order account are the free-recall differences found between read and generated items in pure and mixed lists, differences that were at the heart of the controversy discussed above. Just as was found in earlier studies (e.g., Slamecka & Katsaiti, 1987), Nairne and colleagues showed that recall was superior for pure read lists over pure generate lists (Nairne et al., 1991; Serra & Nairne, 1993). To examine whether order memory dynamics could account for the recall pattern, Nairne et al. assessed a second dependent measure, input–output correspondence, which quantifies the degree to which the relative order of items output during recall parallels their order during list presentation (Asch & Ebenholtz, 1962). Input–output correspondence is derived by examining the number of consecutively recalled (i.e., output) pairs of items, relative to the total number of items recalled, that preserve the relative order of input. Confirming Predictions 6 and 7, the input–output correspondence scores revealed that the negative-generation effect was accompanied by the finding of significantly less reliance on serial order information as an output strategy for pure lists of generated items, compared with pure lists of read items.

Table 3
Literature Evaluating Order-Memory Patterns and Recall As a Function of Generation

Study	List Composition	Reconstruction		I–O Correspondence		Free Recall	
		R	G	R	G	R	G
Nairne, Riegler, & Serra (1991)							
Experiment 2	Pure lists	.60	.48	.68	.62	.61	.54
	Mixed lists	–	–	–	–	–	–
Serra & Nairne (1993)							
Experiment 1	Pure lists	.68	.58	–	–	–	–
	Mixed lists	.63	.63	–	–	–	–
Experiment 2	Pure lists	.52	.45	–	–	.62	.55
	Mixed lists	.50	.50	–	–	.52	.58
Mulligan (2002)							
Experiment 2	Pure lists	.10	.04	–	–	.33	.23
	Mixed lists	–	–	–	–	–	–

Note—R, read words; G, generate words; I–O, input–output.

Additionally, we argue, as did Nairne et al., that item processing is enhanced by generation (evidenced by the generation effect in recognition), but that the benefit of such elaboration is mitigated by the relatively impoverished order information presumably used to guide free recall.

The item-order account further predicts that a positive generation effect for free recall should be obtained when order memory is approximately equivalent for each item type, as in the mixed lists in the study of Serra and Nairne (1993). Consistent with this prediction (Prediction 5), in the mixed lists free recall was higher for generated than for read items. Two additional predictions of the item-order theory were also confirmed (Predictions 8 and 9): A greater proportion of read items was recalled for pure than for mixed lists, whereas recall was lower for generated items in the pure- than in the mixed-list condition.

A principled extension of the above ideas is that a generation effect in pure lists should emerge for free recall, if an alternative structure for guiding recall is provided to participants. This prediction has been confirmed in a number of experiments. McDaniel et al. (1988, Experiment 1; see also Experiment 2) found significantly higher recall for generated than for read items in taxonomically structured pure lists for which the categorical nature of the list was emphasized during encoding (see Table 8). Similarly, the use of generation contexts that stimulate participants to rely on other categorically related targets in the list to complete the generation task induces greater clustering during recall, thereby supporting a pure-list generation advantage (deWinstanley, Bjork, & Bjork, 1996). This pure-list pattern has been replicated using long, categorized lists of 32–72 items (Burns, 1990; Burns, 1996, Experiment 3A).

Nairne et al. (1991, Experiment 3) provided more detailed recall evidence supporting this prediction and further assessed the degree to which participants relied on order memory during recall. Both reconstruction scores and input–output correspondence suggested a negative effect of generation on order memory, but a partial reversal of the typical pure-list recall pattern was observed. With categories available to guide recall, a reliable generation effect surfaced for the second half of the pure lists. Nairne et al. reasoned that reliance on categorical information to guide recall may be especially pronounced in the second half of the lists, when reliance on order information declines markedly, as indicated by order reconstruction scores. This reasoning may help explain an exception to the patterns just described. Burns (1996, Experiment 2A; see also Burns, 1990, Experiment 1C), using pure high-structured categorized lists, reported that read items remained better recalled than generated items. In this experiment, however, read items not only enjoyed better order encoding than did generated items, but the category clustering scores were significantly higher for read than for generated items, indicating that, in this case, read items also benefited from greater use of the category information to guide recall.

To summarize: The data in the generation literature provide convincing evidence in support of the nine predictions of the item-order account. One might object that the findings reviewed above are limited to conditions in which participants are expecting an order test (see also Mulligan, 2002), and that, otherwise, participants do not spontaneously encode order information to such a degree. Disfavoring this possibility, order reconstruction scores show the same pattern of pure read greater than pure generate, and read equal to generate in mixed lists, even when an incidental reconstruction task is given (Serra & Nairne, 1993). Another point that at first blush seems to disfavor the item-order account is the results (reviewed above) showing the generation effect in a pure-list design. However, the item-order theory anticipates these results, because in these studies recall could be guided by information other than serial order (i.e., taxonomic information).

We now turn to the word frequency effect in an attempt to extend the item-order theory to a general account of the divergent patterns across pure and mixed lists of a variety of manipulations that have captured the interest of the field. If the item-order theory is fruitful, we would expect to see the entire constellation of patterns for word frequency that has been reported in the generation effect literature.

The Word Frequency Effect

Perhaps one of the best known and most accepted findings in the memory literature is that the frequency of words in natural language occurrence affects the memorability of words. Specifically, the pattern most often and persistently summarized is that high-frequency words are recalled better than are low-frequency words, whereas low-frequency words are better recognized than high-frequency words (e.g., Kintsch, 1970a; Neath & Surprenant, 2002). Closer examination reveals, however, that the high-frequency words are recalled better than are low-frequency words when the experimental design uses pure lists of items, whereas low-frequency words are often recalled better (and no worse) than are high-frequency words with mixed lists of items (see Merritt, DeLosh, & McDaniel, 2006, and Watkins, LeCompte, & Kim, 2000, for detailed references; see Benjamin, 2003, Experiment 3, for an exception). Briefly, the item-order account suggests that the superior free recall of high-frequency (common) relative to low-frequency (unusual) words in pure lists rests on better order encoding for pure high-frequency than for pure low-frequency lists. For mixed lists, order encoding is approximately equivalent for high- and low-frequency words, thereby allowing the enhanced item encoding of low-frequency items to manifest in better free recall (than for high-frequency items).

The left-hand columns of Table 4 provide the available order memory findings. As can be seen, across four experiments, in every case the predicted patterns outlined earlier were obtained. For pure lists, order memory was better for high-frequency words than for low-frequency words (Prediction 1); for mixed lists, order memory was nearly identical for high- and low-frequency words (Prediction 3); and the levels of order memory in mixed lists were intermediate to that observed across high- and low-frequency items in pure lists (Prediction 4). Two aspects of this body of evidence are noteworthy. First, in

Table 4
Literature Evaluating Order-Memory Patterns and Recall As a Function of Word Frequency

Study	List Composition	Reconstruction HF	Reconstruction LF	I–O Correspondence HF	I–O Correspondence LF	Free Recall HF	Free Recall LF
Merritt et al. (2006)							
Experiment 1	Pure lists	.29	.19	.67	.59	.47	.31
	Mixed lists	.24	.25	(.59)	(.59)	.35	.39
DeLosh & McDaniel (1996)							
Experiment 1A	Pure lists	.72	.54	.81	.66	.69	.53
	Mixed lists	.60	.62	(.73)	(.73)	.56	.65
Experiment 2A	Pure lists	.59	.47	.69	.56	.55	.43
	Mixed lists	.50	.53	(.65)	(.65)	.51	.59
Experiment 3	Pure lists	.63	.49	.73	.63	.62	.48
	Mixed lists	.53	.55	(.63)	(.63)	.42	.51

Note—HF, high-frequency words; LF, low-frequency words; I–O, input–output. The I–O correspondence values given in parentheses pertain to mixed lists as a whole. The mixed list in Experiment 3 was composed of 2 HF words and 6 LF words.

most experiments, the list lengths were relatively short (8 items; DeLosh & McDaniel, 1996), raising the issue of whether the order encoding dynamics extend to more typical list lengths used in free-recall experiments. Merritt et al. (2006, Experiment 1) addressed this issue by using a more typical list length (16 items), and reported converging results with those reported for shorter lists (see Table 4). Second, the order encoding patterns emerged under conditions in which the instructions did not encourage order encoding, nor were participants expecting an order memory test (Merritt et al., 2006, Experiment 1). These results indicate that order encoding was spontaneous and not limited to situations in which participants were possibly preparing for an order reconstruction test.

Turning to the expectations regarding differential item elaboration for the high- (common) and low-frequency (unusual) words, item elaboration is typically indexed by recognition performance (see Einstein & Hunt, 1980; Hunt & Einstein, 1981). Consistent with Predictions 2 and 5 of the item-order theory, the literature has uniformly reported that in pure-list designs and in mixed-list designs (e.g. Gregg, 1976), low-frequency words are better recognized than are high-frequency ones (e.g., Kinsbourne & George, 1974; Rao & Proctor, 1984). Interestingly, in a novel process dissociation procedure, Nairne and Kelley (2004) estimated the independent effects of word frequency on the use of item and order information in serial recall tasks. Consistent with the item-order account and the above results, low-frequency lists significantly increased item information relative to high-frequency lists.

The middle columns of Table 4 display the correspondences between the input order of the list items and the order output in recall (Asch–Ebenholtz, 1962, scores). All of the scores were above chance levels (.50), indicating that people were using order information to help guide recall. In line with Predictions 6 and 7, the input–output correspondences for pure lists of high-frequency words were significantly above those found for pure lists of low-frequency words. Also worth noting is that the input–output correspondences for the mixed lists were lower than they were for pure high-frequency lists. These patterns support the idea that the enhanced order encoding promoted by pure high-frequency lists was exploited in the service of free recall.

The right-hand columns of Table 4 show the free-recall patterns. The basic finding of high-frequency advantages in pure lists and low-frequency advantages in mixed lists is clearly evident. Importantly, in every case, these patterns were produced because recall of high-frequency (common) items declined in mixed lists relative to pure lists, whereas recall of low-frequency items (unusual) increased in mixed lists relative to pure lists. Note that the changes in recall across list composition for each type of item are precisely as anticipated by the item-order framework (Predictions 8 and 9).

Consideration of Alternative Accounts

There exist a number of alternative theoretical accounts for the superior recall of high-frequency relative to low-frequency words in pure-list designs. However, because none of these can readily account for the reversal of the effect in mixed-list designs (see DeLosh & McDaniel, 1996, and Merritt et al., 2006, for detailed discussions), contemporary alternative accounts do not provide a complete explanation of the patterns of effects reviewed above. Still, aspects of these accounts may provide a more accurate or precise description of the mnemonic advantages of high-frequency items than does the present focus on order encoding. Accordingly, we review these accounts, and introduce additional findings to distinguish the item-order view from the alternative perspectives.

One classic explanation of the recall advantage of high-frequency words rests on the assumption that high-frequency words have stronger or more numerous preexperimental associations than do low-frequency words (Deese, 1960; Gregg, 1976; Nelson & Xu, 1995). Supporting this assumption, average interitem associations are reported to be greater for high-frequency than for low-frequency words. Further, preexperimental associations among target words can substantially influence episodic recall (Nelson & Zhang, 2000), probably because a recalled target can cue retrieval of an associated item. By this explanation, therefore, superior relational information for high-frequency items, in the form of preexisting associations, supports the recall advantage for these items in pure lists. In contrast, the item-order view suggests that the superior relational information for high-frequency items (in pure lists) is based on serial order encoding established during the study episode.

Note that the advantage in preexisting associations for high-frequency words relative to low-frequency words would not be changed by mixing items within a list, and that high-frequency words should still be better recalled in mixed lists. Still, additional dynamics may be present in mixed lists (e.g., distinctive or additional processing of the low-frequency items; see Watkins et al., 2000) that override the association advantage of high-frequency items. Thus, consideration is warranted as to whether a higher degree of preexisting associations in pure lists of high-frequency items than in those of low-frequency items underlies the high-frequency recall advantage.

Much of the older experimental literature is ambiguous on this point, because—although haphazard selection of high- and low-frequency word lists may have resulted in more preexisting interitem associations for high-frequency lists—it is not certain that these associations, if present, were responsible for high-frequency advantages in recall. Order information may still have been at least partially involved in mediating the advantage for high-frequency items. Consistent with this possibility, when semantic associations are purposefully introduced into the word lists (semantically related words from ad hoc categories), for high-frequency lists recall appears to be strongly guided by order information, as evidenced by quite high input–output correspondence scores (.71) (Merritt et al., 2006, Experiment 2A). Moreover, based on a close examination of word-association norms and experimental results with low- and high-frequency cues, Nelson and McEvoy (2000) questioned whether lists composed of high-frequency words necessarily enjoy stronger preexperimental interitem associations than do low-frequency word lists.

One direct way to test the idea that preexisting associations mediate the high-frequency advantage is to construct lists in which the items have no direct or minimal preexperimental connections. If preexisting associations are responsible for the high-frequency advantage, the frequency effect should be eliminated with these lists. Nelson and Xu (1995) reported just such a finding; however, because they used a mixed-list design, the interpretation is uncertain. The critical study was conducted by Merritt et al. (2006). In their study, pure lists of high-frequency words with minimal interitem associativity were constructed, and the minimal associativity was not statistically different from that of the pure low-frequency lists. Across two experiments (when unstructured lists were examined), a robust advantage for high-frequency lists was obtained (see Table 4 for Experiment 1 recall levels; for Experiment 2, recall averaged 70% for high-frequency lists and 53% for low-frequency lists). Clearly, stronger preexperimental associations for high-frequency items cannot be responsible for the recall advantage of the high-frequency items.

Another account of the word frequency effect is that high-frequency items support the formation of more interitem associations during study of the list than do low-frequency items. This idea differs conceptually from the order account, in that the enhanced associations for high-frequency items are not limited to (and do not necessarily include) relations based on serial order, but instead are potentially formed among all of the items in the list. Formalized within the SAM model (Gillund & Shiffrin, 1984, pp. 32–36), the idea is that a retrieved high-frequency word is a stronger cue for stimulating retrieval of another item from the list than is a retrieved low-frequency word (but see Nelson & McEvoy, 2000, for evidence to the contrary). SAM generates two key predictions with regard to recall of mixed lists. First, the recall of high- and low-frequency words should not differ, because the cue strength of any particular high-frequency item used to probe recall will extend to both high- and low-frequency words; similarly, the cue strength of any particular low-frequency item will apply to both high- and low-frequency words. This prediction does anticipate the elimination of the high-frequency advantage in mixed lists, but it is not in line with the finding that low-frequency items are *better* recalled than are high-frequency items in mixed lists (see Table 4).

A second intriguing prediction is that overall recall of a list should improve as the proportion of high-frequency items in the list increases, because with more high-frequency items the cue strength of recovered items will on average be higher, thereby prompting additional recall. To test this prediction, DeLosh and McDaniel (1996) used lists in which 100% (pure list), 75%, 25%, or 0% (pure low-frequency list) of the items were high-frequency words. Total recall was significantly better for the pure high-frequency list than for the other lists (.62); however, countering the prediction, no significant differences in recall were found among the other lists (.50, .46, and .48, for 75%, 25%, and 0% lists, respectively). This finding, in conjunction with superior recall of low-frequency items in mixed lists, counters the interpretation that the word frequency effect (for recall) is based solely on enhanced cuing effectiveness for high-frequency words (as a consequence of stronger item-to-item associations formed during encoding) relative to low-frequency words.

Final Evidence Compelling the Item-Order Theory of the Word Frequency Effect

To provide the strongest possible challenge to the item-order theory, we suggest a straightforward modification to the just-reviewed account (stronger formation of interitem associations for high-frequency words). In mixed lists, low-frequency items may attract additional elaboration or attention during encoding, because they are distinctive relative to the high-frequency items in the list (cf. McDaniel & Geraci, 2006; Schmidt, 1991). More specifically, Watkins et al. (2000) proposed that, when studying mixed lists, participants recognize the difficulty of learning the low-frequency items and, accordingly, expend additional attention on those items. The additional processing afforded low-frequency items in mixed lists would then boost their recall beyond that for high-frequency items in recall (which would otherwise be recalled equivalently, according to the SAM model). By contrast, in pure low-frequency lists, without the contrast with high-frequency items, low-frequency items do not attract additional attention. Accordingly, the high-frequency advantage in interitem associations would lead to superior recall for high-frequency words. With this not unreasonable modification, the mixed list reversal of the recall advantage for high-frequency items could be explained, providing that the associated order encoding effects were viewed as correlated and not causal.

Critically, the item-order framework differs from the above account by asserting that low-frequency words always enjoy more item-specific processing than do high-frequency words. Thus, the framework anticipates in a principled fashion that in pure-list designs, low-frequency words would be better recalled than would high-frequency words, if recall were guided by relational information other than order information (and if, of course, relational information were equally available for low- and high-frequency words). For instance, if the words in the list had some categorical structure, the item-order framework predicts that low-frequency words would be better recalled than would high-frequency words in pure lists, because the theoretical item-specific advantage for low-frequency words would not be offset by deficits in relational (order) information. This prediction is provocative and provides a competitive test of the modified view, as well as of all other views, just sketched. On the modified view (or any other alternative view), in pure lists, high-frequency items can never be recalled at lower levels than can low-frequency items in pure lists; indeed, the high-frequency items would still be expected to be better recalled than low-frequency items, because the categorical structure would only augment the interitem encoding advantages assumed to underlie the high-frequency advantage in pure lists.

Merritt et al. (2006, Experiment 2B) recently conducted the crucial experiment. High- and low-frequency word lists were composed of items from ad hoc categories (e.g., *found in water*; *paper products*); prior to study and to recall, participants were provided with the category label for each list. In this pure-list situation, low-frequency words (M = .65) were significantly better recalled than were high-frequency words (M = .60). Only the item-order framework anticipated this novel finding (see also Merritt et al., 2006, Experiment 2A).

In summary, the item-order account of the word frequency effects in free recall explains and predicts a range of patterns that is difficult if not impossible for alternative long-standing explanations to handle. Approaches presuming that more common (high-frequency) words relative to less common (low-frequency) words have inherent advantages in retrievability (Underwood & Schulz, 1960, p. 86) or in interitem associativity (Deese, 1960; Gregg, 1976; Nelson & Xu, 1995) are disfavored by patterns of superior (or equivalent) recall of low-frequency words as compared with high-frequency words in mixed lists (for an exception, see Watkins et al., 2000, Experiments 3 and 4). To retain these sorts of explanations, one might assume that mixed lists stimulate preferential processing of low-frequency items over high-frequency items, thereby overriding the above factors and conferring the low-frequency advantage in mixed lists (e.g., May, Cuddy, & Norton, 1979; Watkins et al., 2000). However, these modifications neither anticipate nor explain the just-noted superior recall of low-frequency words in *pure* lists that are composed to provide some organizational structure (Merritt et al., 2006). Considering these findings in conjunction with the patterns of order encoding and use of order information across pure and mixed lists, the item-order theory arguably provides the most comprehensive and coherent account of word frequency effects on free recall.

The Enactment Effect

The enactment effect refers to the finding that subject-enacted (or performed) actions (e.g., "flip the pancake") are better remembered than experimenter-enacted actions. Like the generation and word-frequency effects, however, the recall patterns are not as straightforward as this simple summary suggests. Rather, the enactment effect reflects a third phenomenon, in which the recall advantage observed for the encoding condition that attracts extensive individual item elaboration (subject-enacted) relative to a more common encoding condition (experimenter-enacted) varies as a function of list composition. Engelkamp and Zimmer (1997) noted discordant patterns when free recall was compared for pure lists of subject-enacted and experimenter-enacted actions versus mixed lists of subject- and experimenter-enacted actions. Specifically, they found that subject-enacted actions were better recalled than experimenter-enacted actions, when mixed lists were used. However, this enactment effect disappeared when pure lists were used (see also Cohen, 1981, 1983).

In a vein similar to the present item-order theory, Engelkamp and Zimmer (1997) suggested an explanation focusing on differential item-specific and relational encoding. They argued that an advantage in item-specific encoding exists for subject-enacted actions, whereas an advantage in relational encoding exists for experimenter-enacted actions. Although the advantages cancel each other out in a pure list, approximately equating recall, item-specific encoding tends to detract from relational encoding in a mixed list permitting the subject-enacted actions to "win out." A critical difference, however, between Engelkamp and Zimmer's view and an extension of the item-order account to the enactment-effect phenomenon is the particular type of relational information believed to be impaired in the context of mixed lists. Whereas the item-order account uniquely emphasizes serial-order information, Engelkamp and Zimmer assumed that a disruption in episodic-relational information, novel associations not necessarily based on order formed during the presentation of the items, mediated the mixed-list effect. Several recent studies examining recall, recognition, and order memory allow us to evaluate the fruitfulness of our focus on order encoding, relative to Engelkamp and Zimmer's reliance on more general relational information to explain the pattern of enactment findings.

The critical data for examining the predictions of the item-order account are found in a set of experiments conducted by Engelkamp and Dehn (2000) and Golly-Häring and Engelkamp (2003). Engelkamp and Dehn first confirmed that the enactment effect in free recall was obtained in mixed but not in pure lists. Consistent with the assumption of both views regarding enhanced item-specific encoding for subject-enacted actions, recognition performance was enhanced for those items relative to experimenter-enacted actions, regardless of list composition. Note that this confirms Prediction 2 (see Table 2) of the item-order account. The critical data regarding order memory are displayed in Table 5. As anticipated by the item-order account (Prediction 1), in pure lists order reconstruction was superior for experimenter-enacted as opposed to subject-enacted actions (see Engelkamp, Jahn, & Seiler, 2003, for a similar finding across a range of order memory measures). In contrast, in mixed lists, order reconstruction was equivalent for subject-enacted and experimenter-enacted actions (Prediction 3). Additionally, consistent with Prediction 4, in mixed lists order reconstruction scores were intermediate to those obtained in pure experimenter-enacted and subject-enacted lists, with one slight exception (see Engelkamp & Dehn, 2000, Experiment 2).

The order reconstruction patterns are not necessarily predicted by Engelkamp and Zimmer's (1997) episodic-relational account. However, these patterns do not disconfirm the notion that episodic-relational information (i.e., information that is not order based) is used to guide retrieval. Unfortunately, Engelkamp and Dehn (2000) did not compute the input–output correspondence scores needed to inform this issue. Instead, they calculated the correlations between order reconstruction and free-recall performance, providing a less direct assessment of the degree to which order information is used to guide recall. Nevertheless, the correlations were consistent with Predictions 6 and 7 of the item-order account. For pure lists of experimenter-performed actions, the correlation was .83, whereas for subject-performed actions the correlation was lower ($r = .68$), suggesting that participants relied on order to guide retrieval.

Regarding the final two predictions of the item-order account specifying changing recall levels across list compositions, recall was higher for experimenter-enacted actions in pure lists than in mixed ones (Prediction 8), and recall was higher for subject-enacted actions in mixed lists than in pure ones (Prediction 9).

Golly-Häring and Engelkamp (2003) provided further evaluation of the item-order account. They reasoned that if the list of target actions contained salient relational information to supplant order information in guiding recall, the item-specific encoding advantage for subject-enacted actions should prevail in a pure list. Consequently, enactment effects should now emerge in free recall, even in pure lists. The entire pattern of results from their experiments using categorized word lists was consistent with this reasoning. First, categorical clustering scores were substantial and equivalent for pure lists of subject-enacted and experimenter-enacted actions, indicating use of a non-order-based relational strategy to guide recall. Second, and in contrast, input–output correspondence scores for recalled items were low and near chance, and order reconstruction scores showed a similarly minimal reliance on order information for both types of pure lists and for mixed lists. Third, the recognition test indicated that item-specific encoding was bolstered for the subject-enacted items.

These patterns capture the dynamics anticipated by the item-order theory and converge on the major prediction, which was confirmed: For categorized lists of action phrases, pure lists of subject-enacted actions were recalled better than were pure lists of experimenter-enacted actions, reversing the typical common item (experimenter-enacted) advantage obtained in pure lists. Dovetailing nicely with this finding, in other contexts where reliance on order information to guide retrieval of action phrases is very low (e.g., long, unstructured lists), the typical pure-list common item advantage has been shown to reverse (Engelkamp & Dehn, 2000). As discussed in the previous section on word frequency, only the item-order account anticipates such pure-list reversals.

In summary, the enactment effect is a third phenomenon for which the item-order theory explains and anticipates the effects of encoding condition, list composition, and, most critically, the interaction of these two variables on recognition, order reconstruction, and free-recall performance.

The Perceptual Interference Effect

In prior sections, we have presented evidence supporting the notion that various unusual or atypical encoding conditions facilitate item-specific processing at the expense of order memory, thereby producing different patterns in mixed- and pure-list designs. We next present evidence regarding the effects of perceptual interference during encoding that supports this same theme. In a typical perceptual interference paradigm, participants are asked to read

Table 5
Literature Evaluating Order-Memory Patterns and Recall As a Function of Enactment

Study	List Composition	Reconstruction EE	Reconstruction SE	I–O Correspondence EE	I–O Correspondence SE	Free Recall EE	Free Recall SE
Engelkamp & Dehn (2000)							
Experiment 1	Pure lists	–	–	–	–	.60	.53
	Mixed lists	–	–	–	–	.52	.57
Experiment 2	Pure lists	.52	.40	.83*	.68*	.52	.47
	Mixed lists	.41	.39	.33*	−.13*	.46	.54

Note—SE, subject enacted (performed); EE, experimenter enacted; I–O, input–output. *Value reflects correlation (r) between reconstruction performance and free recall, not I–O correspondence.

words that are presented for 100 msec and then backward masked for 2.4 sec (perceptual interference condition) or words that are continuously presented for 2.5 sec (intact condition). The perceptual interference effect refers to the finding that recall is surprisingly higher for words presented in the perceptual interference condition (Hirshman & Mulligan, 1991; Mulligan, 1996; Nairne, 1988). Stimulated by the item-order account, Mulligan (1999) subsequently demonstrated that the perceptual interference effect was reversed in pure lists. Because the perceptual interference effect held in mixed but not pure lists, Mulligan (1999, 2000a, 2000b) applied the item-order theory to further investigate this phenomenon. The results of the relevant experiments are summarized in Table 6.

We begin examining the predictions of the item-order account by considering the initial experiments conducted by Mulligan (1999; Experiments 4 and 5). In these experiments, participants were presented with a word list, followed by a brief distractor, then administered either an order-reconstruction or a free-recall test. The order-reconstruction scores from Experiment 4 indicate that order memory was significantly higher for pure lists of intact words than for pure lists of perceptual interference words (Prediction 1, from Table 2). This advantage diminished, though remained significant, in the mixed-list design of Experiment 5, in line with Prediction 3. Consistent with Prediction 4, order information was most preserved for pure lists of intact words and was worst for pure lists of perceptual interference words, with order information for mixed lists at a level intermediate to the pure lists. These patterns suggest that order information is disrupted by perceptual interference.

Equally important to the item-order account is establishing the effects of perceptual interference on item elaboration. Although Mulligan (1999) did not directly evaluate the degree to which item elaboration varied as a function of encoding condition or list composition, other studies have addressed this question. Consistent with Prediction 2, Mulligan (2000a) showed that item elaboration (recognition) was higher for pure lists of words presented briefly for 100 msec before being masked than for intact words. Further, Mulligan (2000b) found a similar advantage for perceptual interference words, using a different measure of item elaboration: item gains (item gains refer to the number of words recalled on a second recall test that were not recalled on an initial recall test). Data also exist to evaluate item elaboration for mixed lists of perceptual interference and intact words. Consistent with Prediction 5, Hirshman and Mulligan (1991) demonstrated that perceptual interference led to superior recognition performance relative to intact presentation in mixed lists. The extant results are therefore consistent with the notion that perceptual interference enhances item elaboration in pure and mixed lists.

The remaining predictions of the item-order account pertain to the recall and input–output correspondence patterns. In pure lists, recall was significantly higher for intact words than for words presented with perceptual interference, and the input–output correspondence scores mirrored this pattern (Mulligan, 1999). Specifically, a greater reliance on order information to guide recall was found in pure lists of intact words relative to pure lists of words presented with perceptual interference. These findings confirm Predictions 6 and 7. For mixed lists, recall was unexpectedly equated for words in the intact and perceptual interference conditions. This finding is contrary to the mixed list recall advantage observed for perceptual interference items in an earlier study (Hirshman & Mulligan, 1991), and is a puzzling result from the perspective of the item-order account. However, the finding is not as surprising as it may initially seem when the order reconstruction scores are considered. In the mixed lists, reconstruction was worse for the perceptual interference items than for the intact items, and this may explain the recall pattern (see Mulligan, 1999).

Comparing the means for each type of encoding condition between the pure- and mixed-list compositions, the expected recall dynamics are partially confirmed: Intact words were recalled at a higher rate in pure than in mixed lists. This is the pattern anticipated by the item-order account (Prediction 8), which attributes this dynamic to changes in order memory; and order reconstruction scores were somewhat higher for intact words in pure (.48) than in mixed (.46) lists. On the other hand, whereas order reconstruction was more accurate for words encoded with perceptual interference in mixed (.43) than in pure (.39) lists—and correspondingly, words encoded with perceptual interference were expected to be recalled at a higher rate in mixed than in pure lists—this prediction (Prediction 9) was not confirmed. Rather, recall was equivalent for words encoded with perceptual interference across list compositions. This may be partially attributable to the relatively low order memory for perceptual interference items in mixed lists, as discussed above.

One final result that is expected by the item-order account is notable. Following the logic of experiments

Table 6
Literature Evaluating Order-Memory Patterns and Recall As a Function of Perceptual Interference

Study	List Composition	Reconstruction		I–O Correspondence		Free Recall	
		I	PI	I	PI	I	PI
Mulligan (1999)							
Experiment 4	Pure lists	.48	.39	.69	.60	.53	.48
Experiment 5	Mixed lists	.46	.43	–	–	.49	.48
Hirshman & Mulligan (1991)	Mixed lists					.17	.26

Note—I, intact presentation; PI, perceptual interference; I–O, input–output.

reviewed in previous sections on the generation, word-frequency, and enactment effects, Mulligan (1999; Experiment 3) showed that the pure-list recall advantage for intact words over perceptual interference words was eliminated by using categorized word lists. Thus, for four seemingly disparate phenomena, the item-order theory not only accounts for a constellation of pure-list effects (order memory, free recall), but also predicts the obtained reversal or elimination of those effects when the pure lists are constructed to provide alternative retrieval strategies (i.e., categorical information) to that of using order information. As proposed in the theory, these alternative strategies offset the disadvantage that is generally observed for unusual or atypical items in recall of pure lists, on account of disrupted order information, thereby allowing the item-specific advantage for these items to be revealed.

The Bizarreness Effect

From the time of the Greeks, mnemonists have suggested that imagining content in an illogical or bizarre fashion will enhance recall of that content (see Cermak, 1975; Lorayne & Lucas, 1974). Over the past 30 years, memory researchers have experimentally tested this claim. Typically, target noun triplets are presented in sentences specifying common scenes, such as "The MAID spilled AMMONIA on the TABLE" or specifying bizarre scenes such as "The MAID licked AMMONIA off the TABLE." In most experiments, subjects are instructed to form an interactive visual image of the meaning of each sentence presented in a list. For free recall, a remarkably consistent pattern has been obtained. In line with the many phenomena described at the outset of this article, when bizarre and common sentences are presented in mixed lists, recall of content in bizarre frames is significantly better than is recall of content in common frames (McDaniel, Anderson, Einstein, & O'Halloran, 1989; McDaniel & Einstein, 1986; McDaniel, Einstein, DeLosh, May, & Brady, 1995; Merry, 1980; Pra Baldi, de Beni, Cornoldi, & Cavedon, 1985; Webber & Marshall, 1978; Wollen & Cox, 1981). By contrast, with pure-list presentations, the bizarreness advantage is either eliminated or reversed (Collyer, Jonides, & Bevan, 1972; Cox & Wollen, 1981; Hauck, Walsh, & Kroll, 1976; McDaniel & Einstein, 1986; Senter & Hoffman, 1976; Wollen, Weber, & Lowry, 1972).

The item-order theory can be straightforwardly applied to this seemingly curious (from the mnemonist's perspective) pattern. Bizarre sentences presumably attract more individual-item elaboration than do common sentences (see Wollen & Margres, 1987; Worthen, Garcia-Rivas, Green, & Vidas, 2000); however, this detracts from order encoding in pure lists relative to the order encoding afforded common sentences. As detailed throughout, this dynamic in pure lists would produce equivalent or slightly better recall for common than for bizarre sentences. In mixed lists, when order memory is more or less equal for bizarre and common sentences, the more extensive individual-item encoding of bizarre items (relative to common items) would be manifested in superior free recall for bizarre items.

Two published studies have tested most of the detailed predictions (as listed in Table 2) associated with an item-order interpretation of the pattern of bizarreness effects in pure and mixed lists. Table 7 summarizes the results from the pertinent experiments. Consider first the order reconstruction scores. Consistent with Prediction 1, in pure lists, memory for serial order of common sentences was, without exception, significantly better than was memory for order of bizarre sentences. Similarly consistent with Prediction 3, in mixed lists, memory for serial order was approximately equivalent for common and bizarre sentences (with the order reconstruction values not significantly different in any experiment). Finally, with regard to Prediction 4, the levels of order memory in the mixed lists were (with one exception—see McDaniel et al., 2000, Experiment 1) intermediate to that in pure lists.

With regard to the assumption that bizarre sentences attract more individual-item processing than do common sentences, it is worth noting that, in cases where sentence presentation is subject-paced, more attention is devoted to bizarre than to common sentences, regardless of list composition. For instance, in McDaniel and Einstein (1986, Experiment 1), in pure lists processing time averaged 11.7 sec for bizarre sentences and 8.8 sec for common sentences. In mixed lists, processing time was 13.8 sec on average for bizarre sentences and 11.1 sec for common sentences (Experiment 3). To the extent that processing time reflects

Table 7
Literature Evaluating Order-Memory Patterns and Recall As a Function of Bizarreness

Study	List Composition	Reconstruction Com	Reconstruction Biz	I–O Correspondence Com	I–O Correspondence Biz	Free Recall Com	Free Recall Biz
McDaniel et al. (1995)							
Experiment 4	Pure lists	.30	.17	–	–	–	–
	Mixed lists	.28	.29	–	–	–	–
Experiment 5	Pure lists	.44	.33	.58	.51	.60	.50
	Mixed lists	.34	.39	(.53)	(.53)	.45	.56
McDaniel et al. (2000)							
Experiment 1	Pure lists	.62	.55	.66	.58	.67	.62
	Mixed lists	.52	.54	(.58)	(.58)	.59	.69
Experiment 2	Pure lists	.66	.53	.72	.61	.67	.59

Note—Com, common sentences; Biz, bizarre sentences; I–O, input–output. The I–O correspondence values given in parentheses pertain to mixed lists as a whole.

item elaboration, this finding is in line with Prediction 2 that uncommon (bizarre) items enjoy superior item elaboration relative to common items, even in pure lists.

To get a sense of the degree to which the encoded order information played a role in recalling the sentences, we turn to the input–output correspondence scores (McDaniel et al., 1995, provide details of how sentence order was constructed from recall of target nouns). The patterns were again completely consistent with a theoretical analysis based on the item-order theory. In all experiments, the recall order of common sentences from pure lists significantly corresponded to the input order (Prediction 6). In contrast, recall order of bizarre sentences from pure lists was not significantly associated with input order, with input–output correspondence scores significantly reduced from those with common sentences (Prediction 7).

Finally, the two experiments that collected free recall for both pure and mixed lists verified that for common sentences recall was attenuated in mixed relative to pure lists (Prediction 8). For bizarre sentences, by contrast, recall was augmented in mixed relative to pure lists (Prediction 9).

Thus, the item-order theory correctly anticipates patterns of order memory and associations between order memory and free recall that are in line with the typically observed bizarreness effect (in free recall) in mixed lists, and the elimination or reversal of the bizarreness effect in pure lists. In contrast, a venerable account that many theorists have found parsimonious and intuitively appealing (see Worthen, 2006, for an excellent review) does not readily explain this pattern. By this account, the positive effects of bizarreness stem from additional cognitive resources (e.g., attention) allocated to the bizarre items rather than to the common ones. However, as noted above, bizarre items in pure lists can receive more processing than common items can, yet they may be recalled less well than the common items (McDaniel & Einstein, 1986, Experiment 1).

An Alternative Account: The Distinctiveness Explanation

A widely embraced construct for explaining many episodic memory effects is the idea that distinctiveness enhances memory (e.g., see the Hunt & Worthen, 2006, book). This idea can be developed to account for the general change in patterns of free recall across list composition that have been reviewed herein. When items with unusual or uncommon features—for example, fragmented items (generation task), low-frequency words, enacted actions, items with perceptual interference, and bizarre items—are presented in mixed lists, the unusual or distinctive nature of these items becomes especially prominent. That is, the presence of common items in the mixed list highlights and attracts attention to the relatively unusual nature of the uncommon items. This favored processing of the unusual items then supports better recall in mixed lists. By contrast, when the unusual items are presented in pure lists, the distinctiveness of those items becomes muted. That is, the items no longer "stand out" when in the presence of other unusual items, and thus do not attract more extensive processing or encoding. Consequently, unusual items presented in pure lists do not show advantages in free recall relative to common items.

The general theme just outlined has been a ready explanation for a majority of the effects detailed in previous sections. For the bizarreness effects, an account often invoked is that bizarre items become functionally distinctive in mixed but not in pure lists (McDaniel & Einstein, 1986), with concomitant increases in elaborative encoding for the bizarre items (see Merry, 1980; Wollen & Cox, 1981; Wollen & Margres, 1987). Regarding the generation effect, Slamecka and Katsaiti (1987) specifically suggested that in mixed lists the items to be generated draw attention away from the read items, thereby stimulating more elaborate encoding of generated items and more impoverished encoding of read items. A similar proposal has been offered for the recall advantage of low-frequency words in mixed lists (Watkins et al., 2000; see the Word Frequency Effect section for details). These putative encoding dynamics can straightforwardly accommodate the increase in recall of generated (low-frequency, bizarre) items in mixed relative to pure lists and the increase in recall of read (high-frequency, common) items in pure lists relative to mixed lists.

However, the distinctiveness account, and its associated assumptions regarding more pronounced elaboration of unusual items in mixed but not pure lists, encounters difficulties with the entire pattern of findings for the phenomena considered herein. First, recognition testing reveals superior performance in pure-list designs for generated, low-frequency, enacted, and perceptual interference items relative to read, high-frequency, experimenter-performed, and intact items, respectively (e.g., Engelkamp & Dehn, 2000; Kinsbourne & George, 1974; Mulligan, 2000a; Nairne et al., 1991). These findings suggest that, consistent with the item-order theory's assumption, unusual items stimulate richer encoding of individual-item features than do common items, regardless of list composition.

Second, and in a related vein, the distinctiveness account offers no explanation why unusual items are better recalled than are common items in pure lists with relational structure. As detailed in preceding sections and summarized in Table 8, when pure lists are constructed with categorical structure, generated items are now better recalled than are read items (Burns, 1990; McDaniel et al., 1988; Nairne et al., 1991), low-frequency items are now better recalled than are high-frequency items (Merritt et al., 2006), and enacted items are now better recalled than are experimenter-performed items (Golly-Häring & Engelkamp, 2003). These findings disfavor the distinctiveness approach's assumption that in pure lists, unusual items—by virtue of being presented in a context with other unusual items—do not attract extensive item-specific processing. Note that these patterns were predicted by the item-order account; indeed, the novel findings using pure categorized lists emerged from testing principled extensions from the item-order theory.

It is important to note that the distinctiveness of an item might also be defined relative to background knowl-

Table 8
Literature Showing Recall Advantage of Unusual Items in Pure Structured Lists

Study	List Length	Encoding Condition	Pure Structured List
McDaniel, Waddill, & Einstein (1988)	40	Read	.41
		Generate	.55
Burns (1990)	72	Read	.16
		Generate	.24
Burns (1996)	32	Read	.20
		Generate	.30
Golly-Häring & Engelkamp (2003)			
Experiment 1	24	Experimenter-enacted actions	.49
		Subject-enacted actions	.61
Experiment 3	8	Experimenter-enacted actions	.82
		Subject-enacted actions	.86
Merritt, DeLosh, & McDaniel (2006)	8	High frequency	.60
		Low frequency	.65

edge and experience rather than to local list context (see Schmidt, 1991). By this formulation, unusual items would be distinctive even in pure lists, thereby enjoying relatively rich item-specific encoding in pure lists. This idea alone cannot explain the patterns considered in this article, because unusual (distinctive) items should produce better free recall, regardless of list composition (it is for this very reason that some theorists assumed that local list context created functional distinctiveness; e.g., McDaniel & Einstein, 1986; Schmidt, 1991). However, this approach to distinctiveness is fruitful when considered in conjunction with order encoding dynamics, as detailed throughout.

Limits and Boundary Conditions

Serial Order Information and Free Recall

A strong version of the item-order framework would assume that free recall of unstructured lists is always guided, at least in part, by serial order information. This assumption may not be entirely correct (e.g., Engelkamp et al., 2003; Mulligan & Lozito, 2007). Most generally, learners appear to exploit a variety of strategies in guiding free recall (cf. Gronlund & Shiffrin, 1986). More specifically, some factors appear to discourage encoding of or subsequent reliance on serial order information. One factor that has been identified in research testing the item-order framework has been list length. As lists increase in length, encoding of order information may be less robust and accordingly not as useful for supporting free recall (Mulligan & Lozito, 2007). Alternative retrieval strategies are believed to guide recall under such conditions. Engelkamp and Dehn (2000) provided evidence supporting these assertions by showing first that in long lists (24 items), order reconstruction scores were equally low ($<$.15) for subject- and experimenter-performed actions. (See Mulligan & Lozito, 2007, for further demonstrations that in lists of 24 unrelated words or unrelated pictures, order information appears to contribute little to free recall.) Second, in conjunction with the low order reconstruction scores, pure lists of subject-enacted actions were recalled better (M = .39) than were pure lists of experimenter-enacted actions (M = .33), and the same held true in mixed lists where the order reconstruction score was .10. This reverses the finding demonstrated with pure short lists of subject- and experimenter-performed actions. Note that the item-order account acknowledges that when recall strategies are not based on order (or in the event that order encoding can be equated across pure lists of items), the items attracting relatively more item-specific processing will be better recalled in pure lists.[1] Thus, the Engelkamp and Dehn finding is not incompatible with the item-order encoding account in terms of the expected relation between order encoding (uniformly poor and equivalent in this instance) and effects of item type in free recall.

Similarly, the item-order framework has been examined for the generation effect, using longer lists of 40 pure read and generated items (Burns, 1996). Although Burns (1996) labeled his measure of order memory *order reconstruction*, the scoring was different from order reconstruction tasks discussed in this article thus far. Specifically, Burns's (1996) order reconstruction measure was more similar to input–output correspondence, in that it assessed the number of times items were recalled in adjacent output positions presented in adjacent input positions. In pure lists, order reconstruction scores were again low in both conditions ($<$.10 for generated and read items), indicating that little order information was available to guide retrieval. Nevertheless, order reconstruction scores were significantly higher for read items than they were for generated items, and recall correspondingly showed read items to be superior to generated items. This result instantiates the notion that even for long lists, order can at least partially exert an influence on recall performance. Thus, the existing results with longer lists are not entirely incompatible with the item-order account. At this point, the most definitive evidence for the order encoding patterns considered in this article exists primarily for relatively short lists ranging in length from 8 to 16 items (Merritt et al., 2006). However, these list lengths are representative of list lengths often used in research documenting the phenomena to which we have applied the item-order theory.

Another factor that may deflect strategies from use of order information is the nature of the individual stimuli.

For instance, regarding the bizarreness effects, it appears that bizarre sentences may produce a dominant reliance on item distinctiveness in recall in mixed lists (see Knoedler et al., 1999, for a general view of the privileged use of distinctiveness in recall, and McDaniel, Dornburg, & Guynn, 2005, for a specific view with regard to bizarreness). In line with this claim, input–output correspondence for mixed lists of bizarre and common items has been shown to not exceed chance (McDaniel et al., 2000, Experiment 1; McDaniel et al., 1995, Experiment 5), disfavoring the view that order information was used to guide retrieval of these mixed lists. Still, consistent with the idea that order information plays a role in the pure-list free-recall patterns for bizarre and common items, input–output correspondence for pure lists of common sentences is consistently greater than chance (see Table 7).

Absolute Versus Relative Order

One of the main tenets of the item-order account is that unusual or atypical encoding conditions facilitate item-specific processing while reducing the encoding of order information. Our theory does not specify, however, whether it is absolute, relative, or both types of order information that are disrupted by unusual encoding conditions, and that are generally used to guide recall. Some evidence suggests that absolute but not relative order memory is disrupted by the unusual encoding conditions considered herein. For instance, generation has been shown to disrupt order reconstruction, an absolute measure, but not relative recency, a relative order measure (Greene, Thapar, & Westerman, 1998). Similarly, perceptual interference was found to negatively affect order reconstruction, but relative recency was equivalent for words from intact and perceptual interference (pure) lists (Mulligan, 2000a; Experiment 3). Finally, pure low-frequency lists produced significantly lower order reconstruction than did high-frequency lists, but relative recency did not differ between high- and low-frequency items (Mulligan, 2001).

However, pending further work, we remain open on this issue, as several studies have demonstrated that both order reconstruction (i.e., absolute order memory) and input–output correspondence scores, a measure of the degree to which the *relative* order of items output during recall parallels their order during list presentation, consistently reveal a similar pattern of influence on account of encoding conditions. Focusing on data from pure-list designs, based on input–output correspondence scores, generation (Nairne et al., 1991), low-frequency words (DeLosh & McDaniel, 1996), perceptual interference (Mulligan, 1999), and bizarreness (McDaniel et al., 1995) all appear to disrupt both absolute and relative order memory. The same may hold true for the enactment effect, but input–output correspondence scores have not yet been examined.

Dissociation of Item-Enhancing and Order-Disrupting Effects

The item-order account posits that unusual or atypical encoding conditions lead to a trade-off in processing, such that item-specific processing is enhanced and processing of serial order is attenuated; and we have evaluated data that support the notion that the effects are causally related. However, additional results suggest that in some cases unusual encoding conditions disrupt serial encoding without enhancing item-specific processing. For the most part, these situations involve low-meaningful stimuli. For instance, Mulligan (2002) showed that generation did not enhance item memory, though it did impair order memory in pure lists of nonwords or unfamiliar compounds. Exploring word frequency effects, Mulligan (2001) reported a similar finding: Very low-frequency words (e.g., *loess*) disrupted order memory more than either high- or low-frequency words did, but without a corresponding facilitation of item memory (as assessed on a final recognition test of previously untested lists). Note that these results do not necessarily disfavor the notion that unusual stimuli attract additional attention toward the item per se. It may be that learners' attempts to elaborate nonwords or very low-frequency words are minimally successful because of impoverished semantic information. Still, it may be that a simple processing trade-off between item-specific and order memory may not always be the most accurate conceptualization (see Mulligan, 2000a, for related work with perceptual interference effects).

Item Encoding for Unusual Stimuli

A final point on which the item-order framework is underspecified concerns what we mean by *item elaboration* and the dynamics by which unusual stimuli provoke additional item elaboration.[2] *Item elaboration* generally refers to an enrichment of features encoded for the item. However, some preliminary evidence suggests that for one of the manipulations considered in this article (generation), enhanced item processing may be at the level of item familiarity, rather than richer feature elaboration (Nairne & Kelley, 2004).

A range of operations producing additional elaboration is likely possible. Stimuli that are clearly unusual may lure additional attention because of orienting or startle reactions (as with taboo words or bizarre stimuli; Hirshman, Whelley, & Palij, 1989). In other cases, the unusual presentation format may require additional attention and elaboration to interpret the stimuli (as with generation and perceptual interference) or complete the required task (as in enactment). Even when the presentation format is not unusual, as in the case of low-frequency words, learners may allocate additional processing to ensure interpretation or acquisition of these words (Watkins et al., 2000). Sometimes the stimulus presentation may itself be enriched, thereby affording more item encoding (as in detailed pictures relative to line drawings; Zucco, Traversa, & Cornoldi, 1984).

Conclusion

In summary, this article has assembled a wide range of free-recall findings that converge on the fundamental impact of a ubiquitous factor in memory research (and more generally in experimental psychology): the manipulation of a variable between lists of stimuli (pure lists) versus within lists of stimuli (mixed lists). Further, the effects of pure versus mixed lists are remarkably consistent, when viewed from the following perspective: Items that are unusual (low-frequency items, words presented with perceptual interference), bizarre, or that attract relatively extensive individual item elaboration (e.g., subject-performed tasks; targets that must be generated from fragmented information; sentences that are difficult to comprehend; detailed pictures) are better recalled when mixed within lists of stimuli that are more typical of the dimension of interest. By contrast, the unusual items enjoy no advantage in recall, or are recalled less well than the common items, when pure lists of stimuli are used.

At least two important implications emerge from this observation. First, well-known and extensively cited free-recall effects that have been "established" in the literature—bizarreness effects, word frequency effects, enactment effects, generation effects, and others enumerated in Table 1—are all limited and bounded by the composition of the experimental stimulus lists. Thus, these effects cannot be described as general, immutable memory findings. Second, and perhaps more importantly, the stability of the pattern across an array of stimulus manipulations strongly suggests that there are systematic and principled properties of memory revealed by list composition.

Our proposed item-order framework rests on the assumptions that (1) free recall of lists of unrelated items is mediated in part by retained information about the presentation order of the items; (2) list composition affects the degree to which order information differs across unusual and common items; and (3) unusual or complex items attract more individual-item elaboration than do more common or simple items, regardless of list composition. The theory generates a set of novel predictions regarding the encoding and use of order information as a joint function of stimulus type and list composition, and relates these order-encoding and retrieval dynamics to the free-recall patterns in a principled way. The detailed review of the five phenomena for which the order memory data were available found that the data consistently supported the theory's predictions. Even in cases when longer lists were used, thus discouraging the use of order information (e.g., see Mulligan & Lozito, 2007), the free-recall patterns were in line with what would be expected when order information is minimally used in free recall (Engelkamp & Dehn, 2000).

A long-standing and still popular alternative interpretation of list composition effects of at least some free-recall phenomena is the idea that in mixed lists, unusual items receive more study time or individual item elaboration than the common items do. In pure lists, however, more rehearsal would not accrue to unusual items than to common items, because there are no common items in the pure unusual list from which to "steal" study time. This idea is often couched in terms of rehearsal being displaced from common items to unusual items (see Underwood, 1983), and has been specifically applied to the list composition effects for generation (Slamecka & Katsaiti, 1987) and word frequency (Watkins et al., 2000). Indeed, the finding that unusual items are better recalled in mixed but not in pure lists is seen as diagnostic that subjects did displace rehearsal in the mixed list (Underwood, 1983, p. 158). A similar but more general idea using contemporary framing is that unusual items become distinct in the context of common items in the mixed list, and therefore are better recalled in mixed but not in unmixed list designs (see, e.g., McDaniel & Einstein, 1986, for bizarreness effects, and Schmidt, 1994, for humor effects).

Several critical findings from the literature adjudicate between the item-order view and the just-mentioned alternative view. According to the alternative view, unusual items could never show recall superior to that for common items in pure lists, because pure lists do not displace rehearsal to unusual items, or do not provide a context in which the unusual items become distinctive. By contrast, the item-order framework clearly anticipates that, if relational information that can be used to guide recall is embedded in pure lists (or, as mentioned above, if order information in longer lists is impoverished so that order cannot be relied upon to help guide recall), unusual items will be better recalled than will common items in pure lists. Note that this is a strong prediction, as it represents a pattern that had not been typically reported prior to development of the item-order theory. When relational semantic information (usually taxonomic structure) is introduced into pure lists, generated items are indeed recalled better than are read items (Burns, 1990, 1996; McDaniel et al., 1988; Nairne et al., 1991); low-frequency words are better recalled than are high-frequency words (Merritt et al., 2006); and subject-performed actions are better recalled than are experimenter-performed actions (Golly-Häring & Engelkamp, 2003; see Table 8). These findings strongly favor the item-order view and are difficult to reconcile with the alternative view.

We close by acknowledging that some details of the item-order theory remain to be fully specified, such as the conditions under which order information is incorporated into free recall (cf. McDaniel et al., 2000; Mulligan & Lozito, 2007), the nature of the order information that is encoded (e.g., Mulligan, 2001), and how that order information is disrupted by unusual items (e.g., Mulligan, 2002). Certainly, some of the initial ideas can be further refined. Moreover, the theory may not be limited to recall, but could also apply to recollective processes in recognition. For instance, recollection of emotionally charged versus neutral words is better in mixed but not in pure lists (Dewhurst & Parry, 2000). Taking a broad perspective, however, we believe that the theory provides a compelling explanation for how list composition influences the five well-known phenomena for which the requisite evidence exists and provides an exciting agenda for fruitful exploration of the many other phenomena listed in Table 1 for which list composition produces divergent patterns of effects in free recall.

AUTHOR NOTE

The authors thank William Hockley, Neil Mulligan, and Jim Nairne for helpful comments on an earlier version of the manuscript, and the latter two for sharing unreported means from their work. We also thank Lindsay Brockmeier for assistance with technical aspects of the manuscript. J.M.B. was supported by National Institute on Aging Grant 5T32AG00030 during writing of this manuscript. Correspondence concerning this article should be addressed to M. A. McDaniel, Washington University, Department of Psychology, Campus Box 1125, One Brookings Dr., St. Louis, MO 63130 (e-mail: mmcdanie@artsci.wustl.edu).

REFERENCES

Asch, S. E., & Ebenholtz, S. M. (1962). The process of free recall: Evidence for non-associative factors in acquisition and retention. *Journal of Psychology: Interdisciplinary & Applied*, **54**, 3-31.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, **31**, 297-305.

Burns, D. J. (1990). The generation effect: A test between single and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1060-1067.

Burns, D. J. (1996). The item-order distinction and the generation effect: The importance of order information in long-term memory. *American Journal of Psychology*, **109**, 567-580.

Burns, D. J., Curti, E. T., & Lavin, J. C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition*, **21**, 846-852.

Cermak, L. S. (1975). *Improving your memory.* Oxford: Norton.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, **22**, 267-281.

Cohen, R. L. (1983). The effect of encoding variables on the free recall of words and action events. *Memory & Cognition*, **11**, 575-582.

Collyer, S. C., Jonides, J., & Bevan, W. (1972). Images as memory aids: Is bizarreness helpful? *American Journal of Psychology*, **85**, 31-38.

Cox, S. D., & Wollen, K. A. (1981). Bizarreness and recall. *Bulletin of the Psychonomic Society*, **18**, 244-245.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.

Deese, J. (1960). Frequency of usage and number of words in free-recall: The role of association. *Psychological Reports*, **7**, 337-344.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1136-1146.

Dewhurst, S. A., & Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology*, **12**, 541-555.

deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory*, **4**, 31-48.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, **44**, 109-120.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Addictive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 588-598.

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: Springer.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 671-682.

Engelkamp, J., Jahn, P., & Seiler, K. H. (2003). The item-order hypothesis reconsidered: The role of order information in free recall. *Psychological Research*, **67**, 280-290.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, **96**, 43-60.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161-190). Washington, DC: Winston.

Fuchs, A. H. (1969). Recall for order and content of serial word lists in short-term memory. *Journal of Experimental Psychology*, **82**, 14-21.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Golly-Häring, C., & Engelkamp, J. (2003). Categorical-relational and order-relational information in memory for subject-performed and experimenter-performed actions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 965-975.

Greene, R. L., Thapar, A., & Westerman, D. L. (1998). Effects of generation on memory for order. *Journal of Memory & Language*, **38**, 255-264.

Gregg, V. H. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183-216). Oxford: Wiley.

Gregg, V. H., Montgomery, D. C., & Castaño, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning & Verbal Behavior*, **19**, 240-245.

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 550-561.

Guynn, M. J., & McDaniel, M. A. (1999). Generate—sometimes recognize, sometimes not. *Journal of Memory & Language*, **41**, 398-415.

Hadley, C. B., & MacKay, D. G. (2006). Does emotion help or hinder immediate memory? Arousal versus priority-binding mechanisms. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 79-88.

Hall, J. F. (1954). Learning as a function of word frequency. *American Journal of Psychology*, **67**, 138-140.

Hauck, P., Walsh, C., & Kroll, N. (1976). Visual imagery mnemonics: Common vs. bizarre mental images. *Bulletin of the Psychonomic Society*, **7**, 160-162.

Hinrichs, J. V. (1970). A two-process memory strength theory for judgment of recency. *Psychological Review*, **77**, 223-233.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 484-494.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 507-513.

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory & Language*, **28**, 594-609.

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

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 & Language*, **32**, 421-445.

Hunt, R. R., & Worthen, J. (Eds.) (2006). *Distinctiveness and memory.* New York: Oxford University Press.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**, 649-667.

Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory & Language*, **29**, 433-454.

Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **13**, 63-69.

Kintsch, W. (1970a). *Learning, memory, and conceptual processes*. Oxford: Wiley.

Kintsch, W. (1970b). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 331-373). New York: Academic Press.

Kintsch, W. (1978). More on recognition failure of recallable words: Implications for generation–recognition models. *Psychological Review*, **85**, 470-473.

Knoedler, A. J., Hellwig, K. A., & Neath, I. (1999). The shift from recency to primacy with increasing delay. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 474-487.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 149-169.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, **96**, 25-57.

Linton, M. (1975). Memory for real-world events. In D. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376-404). San Francisco: W. H. Freeman.

Lorayne, H., & Lucas, J. (1974). *The memory book.* New York: Stein & Day.

Mandler, G. (1969). Input variables and output strategies in free recall of categorized lists. *American Journal of Psychology*, **82**, 531-539.

Mandler, G., & Dean, P. J. (1969). Seriation: Development of serial order in free recall. *Journal of Experimental Psychology*, **81**, 207-215.

May, R. B., Cuddy, L. J., & Norton, J. M. (1979). Temporal contrast and the word frequency effect. *Canadian Journal of Psychology*, **33**, 141-147.

McDaniel, M. A., Anderson, D. C., Einstein, G. O., & O'Halloran, C. M. (1989). Modulation of environmental reinstatement effects through encoding strategies. *American Journal of Psychology*, **102**, 523-548.

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, & Cognition*, **26**, 1045-1056.

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.

McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 54-65.

McDaniel, M. A., Einstein, G. O., De Losh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 422-435.

McDaniel, M. A., & Geraci, L. (2006). Encoding and retrieval processes in distinctiveness effects: Toward an integrative framework. In J. Worthen & R. R. Hunt (Eds.), *Distinctiveness and memory* (pp. 65-88). New York: Oxford University Press.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory & Language*, **27**, 521-536.

Merritt, P. S., DeLosh, E. L., & McDaniel, M. A. (2006). Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition*, **34**, 1615-1627.

Merry, R. (1980). Image bizarreness in incidental learning. *Psychological Reports*, **46**, 427-430.

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1067-1087.

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 54-69.

Mulligan, N. W. (2000a). Perceptual interference and memory for order. *Journal of Memory & Language*, **43**, 680-697.

Mulligan, N. W. (2000b). Perceptual interference at encoding enhances item-specific encoding and disrupts relational encoding: Evidence from multiple recall tests. *Memory & Cognition*, **28**, 539-546.

Mulligan, N. W. (2001). Word frequency and memory: Effects on absolute versus relative order memory and on item memory versus order memory. *Memory & Cognition*, **29**, 977-985.

Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition*, **30**, 850-861.

Mulligan, N. W., & Lozito, J. P. (2007). Order information and free recall: Evaluating the item-order hypothesis. *Quarterly Journal of Experimental Psychology*, **60**, 732-751.

Nairne, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 248-255.

Nairne, J. S. (1992). The loss of positional certainty in long-term memory. *Psychological Science*, **3**, 199-202.

Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory & Language*, **50**, 113-133.

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 702-709.

Neath, I., & Surprenant, A. M. (2002). *Human memory: An introduction to research, data, and theory.* Belmont, CA: Thompson Wadsworth.

Nelson, D. L., & McEvoy, C. L. (2000). What is this thing called frequency? *Memory & Cognition*, **28**, 509-522.

Nelson, D. L., & Xu, J. (1995). Effects of implicit memory on explicit recall: Set size and word-frequency effects. *Psychological Research*, **57**, 203-214.

Nelson, D. L., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. *Psychonomic Bulletin & Review*, **7**, 604-617.

Postman, L. (1972). A pragmatic view of organization theory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 3-38). New York: Academic Press.

Pra Baldi, A., de Beni, R., Cornoldi, C., & Cavedon, A. (1985). Some conditions for the occurrence of the bizarreness effect in free recall. *British Journal of Psychology*, **76**, 427-436.

Ransdell, S. E., & Fischler, I. (1989). Effects of concreteness and task context on recall of prose among bilingual and monolingual speakers. *Journal of Memory & Language*, **28**, 278-291.

Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 386-394.

Roediger, H. L., III (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, **50**, 225-254.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, **19**, 523-542.

Schmidt, S. R. (1994). Effects of humor on sentence memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 953-967.

Senter, R. J., & Hoffman, R. R. (1976). Bizarreness as a nonessential variable in mnemonic imagery: A confirmation. *Bulletin of the Psychonomic Society*, **7**, 163-164.

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., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, **26**, 589-607.

Soraci, S. A., Jr., Franks, J. J., Bransford, J. D., Chechile, R. A., Belli, R. F., Carr, M., & Carlin, M. (1994). Incongruous item generation effects: A multiple-cue perspective. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 67-78.

Toglia, M. P., & Kimble, G. A. (1976). Recall and use of serial position information. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 431-445.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, **21**, 437-484.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, **5**, 381-391.

Tzeng, O. J. (1976). A precedence effect in the processing of verbal information. *American Journal of Psychology*, **89**, 577-599.

Tzeng, O. J., Lee, A. T., & Wetzel, C. D. (1979). Temporal coding in verbal information processing. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 52-64.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, **76**, 559-573.

Underwood, B. J. (1983). *Attributes of memory*. Glenview, IL: Scott Foresman.

Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning.* Philadelphia: Lippincott.

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, & Cognition*, **26**, 239-245.

Webber, S. M., & Marshall, P. H. (1978). Bizarreness effects in imagery as a function of processing level and delay. *Journal of Mental Imagery*, **2**, 291-300.

White, R. (2002). Memory for events after twenty years. *Applied Cognitive Psychology*, **16**, 603-612.

Wollen, K. A., & Cox, S. (1981). Sentence cuing and the effectiveness of bizarre imagery. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 386-392.

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: Springer.

Wollen, K. A., Weber, A., & Lowry, D. H. (1972). Bizarreness versus interaction of mental images as determinants of learning. *Cognitive Psychology*, **3**, 518-523.

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: Oxford University Press.

Worthen, J. B., Garcia-Rivas, G., Green, C. R., & Vidas, R. A. (2000). Tests of a cognitive-resource-allocation account of the bizarreness effect. *Journal of General Psychology*, **127**, 117-144.

Zaromb, F. M. (2007). *The effects of effort after meaning on recall in within and between subjects designs.* Unpublished master's thesis, Washington University, St. Louis, MO.

Zucco, G., Traversa, A. B., & Cornoldi, C. (1984). *Il ruolo di dettagli non essenziali e della loro predicibilita contestuale nella rievocazione di nemi di figure* [The role of nonessential details and of their contextual predictivity in the recall of names of pictures]. *Ricerche di Psicologia*, **4**, 44-58.

NOTES

1. A similar argument has been proposed by Soraci et al. (1994) to explain the finding that incongruous generation effects in free recall (target items are generated from cues that are not normatively related to the target) are observed for pure lists (of generated and read items). They suggested that having to generate items unrelated to the cues might produce greater interitem organization. Consistent with this idea, greater clustering in recall was observed for an incongruent generation condition relative to a congruent generation condition. Thus, the incongruous generation effect in pure lists appears to reflect reliance on interitem associations other than order to guide recall.

2. We thank William Hockley for raising this issue.

(Manuscript received June 27, 2007;
revision accepted for publication September 1, 2007.)