Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 2, 441–448

0278-7393/14/$12.00 DOI: 10.1037/a0034977

Testing the Item-Order Account of Design Effects Using the Production Effect

Tanya R. Jonker, Merrick Levene, and Colin M. MacLeod
University of Waterloo

A number of memory phenomena evident in recall in within-subject, mixed-lists designs are reduced or eliminated in between-subject, pure-list designs. The item-order account (McDaniel & Bugg, 2008) proposes that differential retention of order information might underlie this pattern. According to this account, order information may be encoded when a common form of processing is used alone in a list (e.g., reading), but not when an unusual form of processing is used (e.g., generation) or when a common form and an unusual form are mixed within a list. The production effect—better memory for words said aloud than for words read silently—shows this same design-contingent pattern. In 2 experiments, we investigated whether differential order retention might underlie the production effect. Consistent with the item-order account, we found that retention of order information was better in pure silent lists than in either pure aloud lists or mixed lists, as measured using an order reconstruction test. Moreover, in Experiment 2, order was better preserved in free recall of pure silent lists than of either pure aloud or mixed lists. Thus, production joins the set of tasks identified by McDaniel and Bugg (2008), and our findings suggest a role for order processing in explaining the production effect.

Keywords: item-order account, memory, production effect, design effects

People are continually seeking ways to improve their memory. Consider a student highlighting portions of a textbook in an attempt to remember relevant information, a parent who sets an alarm as a cue to pick up the children from school, or an elderly person who leaves a pill bottle on the counter as a reminder to take the medication. These behaviors demonstrate that people recognize the importance as well as the difficulty of remembering in their daily lives. Consequently, research on memory strategies has received considerable attention, and as a result, a variety of encoding strategies for improving memory have been identified. For example, items that are generated from a cue are better remembered than items that are simply read (the generation effect; e.g., Slamecka & Graf, 1978), bizarre images of items are better remembered than common images (the bizarreness effect; e.g., McDaniel & Einstein, 1986), partially masked items are better remembered than items that are unmasked (the perceptual interference effect; e.g., Nairne, 1988), and actions described in a sentence are better remembered when they are enacted than when they are passively read (the enactment effect; e.g., Cohen, 1981).

These encoding techniques produce a distinct advantage for the information that is processed in an unusual manner (e.g., generated, bizarrely imaged, partially masked, or enacted) relative to the information that is encoded in a more common manner (e.g., simply read, normally imaged, viewed unmasked, or not enacted). This reasonably intuitive pattern is evident in free recall when the two types of encoding processes occur in the same study list. Interestingly, however, the pattern is attenuated, eliminated, or occasionally even reversed in free recall when one group of participants processes in an unusual manner while another processes in a more common manner (for examples, see McDaniel & Bugg, 2008). In other words, the phenomenon is most distinct in *mixed-list, within-subject* designs in which each participant experiences both unusually processed and commonly processed items in the same study list; it typically does not appear in *pure-list, between-subjects* designs in which unusually processed and commonly processed items are studied in separate lists. This pattern has been noted for some time (e.g., McDaniel & Einstein, 1986; Serra & Nairne, 1993).

Recently, McDaniel and Bugg (2008) proposed a unifying explanation to address this pattern of results in free recall studies (see also Nairne, Riegler, & Serra, 1991; Serra & Nairne, 1993). Their explanation—the *item-order account*—emphasizes the encoding of two types of information and maintains that list composition (pure vs. mixed) determines the type of information that is encoded during study. The first type of information is item specific, which is the result of rich encoding of each individual item. This item-specific information is thought to be well encoded for unusually processed items because of the nature of the process (e.g., generation) or the properties of the stimulus itself (e.g., bizarre item). Importantly, this type of information is specific to the item itself

This article was published Online First November 11, 2013.

Tanya R. Jonker, Merrick Levene, and Colin M. MacLeod, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

This research was supported by Discovery Grant A7459 from the Natural Sciences and Engineering Research Council (NSERC) of Canada (to Colin M. MacLeod) and by an NSERC Vanier Canada Graduate Scholarship (to Tanya R. Jonker). We thank Bonnie Armstrong, Ailin Huang, Jennah Kerman, Mackenzie Mullen, Mariah Nanton, Bhavan Sandhu, and Trish Varao Sousa for their assistance with data collection. We also thank Trish Varao Sousa and Brianna Groot for assistance with coding data.

Correspondence concerning this article should be addressed to Tanya R. Jonker, Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada. E-mail: trjonker@uwaterloo.ca

and is therefore well encoded for unusually processed items irrespective of whether the list design is mixed or pure.

The second type of information is the relations among list items. In particular, a participant might encode and retain the studied order of the items. Critically, the item-order account holds that relational information is not well encoded for unusually processed items, whereas it is well encoded for commonly processed items but only when they are studied in a pure list. According to the account, participants do not encode relational information well in a mixed list because the unusual encoding captures attention. The consequence of this is that item-specific elaborative information is encoded for unusually processed items both in mixed lists and in pure lists, whereas interitem relational information is encoded for commonly processed items in pure lists only. The item-order account posits that it is this difference that produces differing recall patterns for the two types of encoding tasks in mixed versus pure lists.

To elaborate on this concept, consider first pure lists. In the case of unusually processed items in pure lists, the unique, item-specific information assists identification and recall of these items. In the case of commonly processed items, although these items are not elaborately encoded, memory for relational structure (e.g., serial order) guides recall. Thus, both unusually processed and commonly processed items benefit from a unique source of information in pure lists, which works to equate performance under the two encoding conditions. This is not the case for mixed lists. In mixed lists, the relational structure is not well encoded because it is disrupted by the item-specific elaboration that occurs during unusual processing. Consequently, participants can still encode item-specific information and avail themselves of it when recalling the unusually processed items, but their recall of commonly processed items suffers because there is no interitem information to guide recall, leading to a mixed-list effect.

This account is supported by findings from studies exploring memory for studied order (Engelkamp & Dehn, 2000; Mulligan, 1999, 2002; Nairne et al., 1991; Serra & Nairne, 1993), which is a form of relational memory. In these studies, participants are presented with short lists of items, one item at a time. Then, during the test phase for each short list, the set of items appears simultaneously on the screen in a scrambled order. Participants are to reorder the items to match the serial order of the study list. These studies have shown that participants are more accurate at reordering items from pure lists of commonly processed items compared to pure lists of unusually processed items (e.g., Nairne et al., 1991), which supports the claim that participants rely on relational information when recalling from pure lists that were commonly processed. In mixed lists, typically there is no difference in order reconstruction performance for unusually processed and commonly processed items, which supports the claim that relational encoding is disrupted by the unusual processing in mixed lists.

To date, the item-order account has been applied to the generation effect, the enactment effect, the perceptual interference effect, the bizarreness effect, the word-frequency effect (McDaniel & Bugg, 2008), and, more recently, the orthographic distinctiveness effect (McDaniel, Cahill, Bugg, & Meadow, 2011) and the testing effect (Karpicke & Zaromb, 2010; Peterson & Mulligan, 2013). Although these encoding techniques involve very different procedures, they produce similar results on order reconstruction tasks for pure lists compared to mixed lists, suggesting a similar underlying mechanism. In the present article, we examine another memory phenomenon that may well belong to this class of encoding techniques: the production effect.

The Production Effect

Speaking some words aloud results in considerably better memory for those items than for items that are read silently. This advantage has been termed the *production effect* (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), and it has been found to occur both on recognition tests and on recall tests. Until recently, this phenomenon has been reported only sporadically (Conway & Gathercole, 1987; Gathercole & Conway, 1988; Hopkins & Edwards, 1972; MacDonald & MacLeod, 1998), but recent work has delineated and extended the phenomenon (e.g., Castel, Rhodes, & Friedman, 2013; Lin & MacLeod, 2012; MacLeod, 2011; MacLeod et al., 2010; Ozubko, Gopie, & MacLeod, 2012; Ozubko, Hourihan, & MacLeod, 2012; Ozubko, Major, & MacLeod, in press; Quinlan & Taylor, in press).

Much like the phenomena reported by McDaniel and Bugg (2008), production occurs in recall when aloud and silent items are studied in a mixed list (e.g., Forrin, Jonker, & MacLeod, in press; MacLeod, 2011). However, to our knowledge, there are no reported investigations of a between-subjects production effect on a free recall test (null or otherwise). Instead, research on the production effect between subjects has typically employed a recognition test. In this recognition research, a between-subjects production effect has been reported periodically (e.g., Gathercole & Conway, 1988), and a recent meta-analysis of a number of recognition and list-discrimination experiments by Fawcett (2013) demonstrated that there is a relatively small effect between subjects. Based on these studies of recognition, it might seem that the production effect is not compatible with the encoding techniques explained by the item-order account. However, although the item-order account predicts tradeoff between item-specific information and relational information in recall of pure lists, which might yield a null result, it makes a different set of predictions for recognition testing. Specifically, when testing recognition of items from pure lists, the item-order account predicts no or very little benefit from relational information due to the testing format, which results in impaired recognition of the commonly processed items compared to the unusually processed items. Thus, it seems that the between-subjects recognition patterns in the production effect accord with the predictions of the item-order account. Therefore, we reasoned that the production effect might also behave similarly to McDaniel and Bugg's (2008) class of encoding techniques on order reconstruction tests. If this were the case, then in pure lists, participants should better remember the order of silent items (i.e., commonly processed) than of aloud items (i.e., unusually processed), and this difference should not occur in mixed lists. The purpose of the present study is to explore the effects of production on memory for order information, with the ultimate goal of relating the production effect to these other memory phenomena.

To date, production has been thought of as a *distinctiveness* effect (Forrin et al., 2013; Hourihan & MacLeod, 2008; Lin & MacLeod, 2012; MacLeod, 2011; MacLeod et al., 2010; Ozubko & MacLeod, 2010; Ozubko et al., in press; cf. Bodner, Taikh, & Fawcett, 2013). A distinctiveness account emphasizes intrinsic differences between stimuli or processes in a mixed list. This difference can then cause the processing of some items to stand out relative to others, which will lead to better recall of the distinct items (see Hunt, 2006, 2013, for more on distinctiveness as a theoretical concept). In the case of production, the process of reading aloud is distinct from the process of reading silently. The aloud items might stand out due to the additional speech code that was produced during study, leading to better memory for these items on the final test.

The distinctiveness account has been the preferred explanation of the production effect because it explains why production results in a mixed-list effect typically in the absence of a pure-list effect, as well as capturing other aspects of the findings to date (e.g., Ozubko & MacLeod, 2010; Ozubko et al., in press). Specifically, in the case of a mixed list, the aloud items are processed in a distinctive manner against the backdrop of the silent items. In the case of a pure aloud list, however, the aloud items do not stand out as distinctive because they are not directly contrasted with silent items. However, Fawcett's (2013) meta-analysis showing a small between-subjects production effect challenges the distinctiveness account as the sole explanation of the production effect.

In previous work concerning the generation effect and the bizarreness effect, among others, some have argued for a distinctiveness explanation (e.g., McDaniel & Einstein, 1986). Yet a distinctiveness explanation makes no predictions regarding memory for order information and therefore cannot explain the order reconstruction results for pure lists found in previous work (e.g., Nairne et al., 1991; Serra & Nairne, 1993). In contrast, the item-order account makes specific predictions regarding memory for order information. Therefore, if the production effect produces the same pattern on order reconstruction tests as observed with McDaniel and Bugg's (2008) encoding techniques, this would suggest that the item-order account might provide a better explanation of the production effect than would a distinctiveness account alone. If the predictions of the item-order account are upheld in the present experiments, this would also ally the production effect with the memory phenomena listed by McDaniel and Bugg, which represent a growing class of encoding tasks that influence within-subject mixed-list studies while attenuating, eliminating, or reversing the pattern in between-subjects pure-list studies.

Experiment 1

To explore how list composition influences encoding, we had participants study three types of word lists in a within-subject design. In some lists, all of the items were read aloud; in other lists, all of the items were read silently; and in other lists, half of the items were read aloud, and half were read silently. After studying each list, participants performed an order reconstruction test: They were shown all of the items from that list in a scrambled order and were to reconstruct the order in which the words had appeared during study of that list.

If the production effect fits with McDaniel and Bugg's (2008) class of encoding techniques, then two results should be observed. First, order reconstruction performance should be better for lists where all of the words were read silently compared to lists where all of the words were read aloud. Second, order reconstruction performance should not differ for silent and aloud items in mixed lists.

Method

Participants. Thirty-seven students (nine male, 28 female) from the University of Waterloo (Waterloo, Ontario, Canada) with an average age of 19.7 years (one participant did not report her age) participated in exchange for partial course credit. All reported fluency in written and spoken English, normal or corrected-to-normal vision, and normal color vision.

Materials. A set of 276 common nouns was selected from the MRC psycholinguistic database (Coltheart, 1981). All had word-frequency scores lower than 500. Twenty-four lists of eight words each were randomly drawn from the pool of 276 nouns; words were not repeated across lists. Eight lists were randomly assigned to each of three list types: pure aloud, pure silent, or mixed. The words were presented on a monitor using E-Prime 2.0 software. Words that were to be read aloud were presented in blue; words that were to be read silently were presented in red.

Procedure. Participants completed 24 study–test blocks. Each block began with a study list during which words appeared one at a time for 2 s each at the center of the screen. Participants were instructed to say blue words aloud and to read red words silently. A research assistant was present throughout the experiment to ensure that participants followed instructions. After each study phase, participants engaged in a 30-s distractor task, during which they saw a series of single digits on the screen. For each digit, they were to indicate whether it was odd or even. The trials were self-paced, and participants were shown single digits until 30 s had passed.

During the test phase, participants were presented with all eight study words in a scrambled order in a vertical list. Test words were presented in black font on a white background. Participants were provided with a sheet of paper with the numbers 1–8 and were instructed to write the words in the order that they were studied. Upon completion of the test phase, a new study–test block began. Prior to beginning the 24 blocks, participants were given a practice phase to familiarize them with the experiment. In the practice phase, participants were exposed to all three tasks (study, distractor, test).

Results and Discussion

We used a strict scoring criterion: Items were considered correct only if they were recalled in their exact serial position (i.e., if *cushion* was studied as Item 4, it would be scored as correct on the final test only if it were placed on the fourth line of the test sheet; this method was used by Nairne et al., 1991). The proportion of items correctly ordered was computed for each list. For the mixed lists, the proportions of correctly ordered aloud items and silent items were computed separately; thus, for each mixed list, there was a proportion of aloud items correctly ordered (out of a total of four) and a proportion of silent items correctly ordered (also out of a total of four).

A 2 × 2 repeated measures analysis of variance assessed the effects of List Type (mixed, pure) and Item Type (aloud, silent) on the proportion assigned to the correct serial positions on the order reconstruction test. The main effects of List Type and Item Type both were marginally significant, $F(1, 36) = 2.86$, $MSE = .01$, $p = .10$, and $F(1, 36) = 3.08$, $MSE = .01$, $p = .09$, respectively. Of main interest, however, was the List Type × Item Type interaction, which was significant, $F(1, 36) = 5.93$, $MSE = .01$, $p = .02$,

$\eta_p^2 = .14$. To interpret the interaction, we performed two paired-samples *t* tests, comparing aloud to silent for each of the list types. Aloud and silent did not differ from each other in mixed lists, $t(36) = 0.41$, $SE = .02$, $p = .68$, but there was a significant difference between aloud and silent in pure lists, $t(36) = 2.97$, $SE = .02$, $p < .01$, $d = 0.49$. As shown in Figure 1, participants were better able to reconstruct the studied order of silent items than of aloud items when they had appeared in pure lists, but not when they had appeared in mixed lists.

These findings support the prediction made by the item-order account. Thus, it appears that the production effect fits with the class of encoding techniques identified by McDaniel and Bugg (2008).

In our second experiment, we had two goals: First, we sought to replicate the order reconstruction effects; second, we wanted to ensure that the production effect would still occur in free recall under this procedure. In Experiment 1, participants might have shifted their encoding strategies to prepare for the order reconstruction tests. If they were encoding differently, this could have changed their free recall pattern. To determine whether this was the case, we tested both free recall and order reconstruction in Experiment 2. The test type (free recall or order reconstruction) was randomly assigned to each study–test block, and therefore, participants were not able to predict how their memory would be tested.

Experiment 2

We modeled our second experiment after an experiment by Nairne et al. (1991, Experiment 2): A random half of the study trials were followed by the order reconstruction task, and the other half were followed by a free recall task. The benefits of this method were threefold. First, the free recall test allowed us to replicate the standard production effect, where free recall of the aloud items is better than that of the silent items in mixed lists, with no (or little) difference in pure lists. Indeed, it allowed us to present mixed lists and pure lists to the same participants within the same session. Second, test type was randomly assigned, and therefore, participants could not predict how they would be tested while they were studying the items. Third, this method allowed us to examine the free recall responses for order information. According to the item-order account, when recalling from a commonly encoded pure list, participants use the relational structure to guide their free recall. Therefore, the free-recall output order of pure silent lists should more closely match the study order than should the output order of pure aloud lists. If this were the pattern observed, it would provide a converging analysis to the order reconstruction measure and further support for the item-order account.

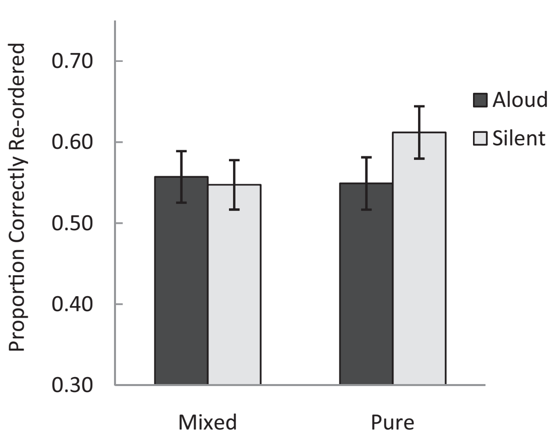

Figure 1. Experiment 1: Proportion of items correctly reordered on the order reconstruction test as a function of list type and encoding condition. Error bars represent one standard error.

Method

Participants. Thirty-eight students (19 male, 29 female) from the University of Waterloo, with a mean age of 19.8 years, participated in exchange for partial course credit. All reported fluency in English, normal or corrected-to-normal vision, and normal color vision.

Materials. The materials were the same as those used in Experiment 1.

Procedure. The procedure differed from Experiment 1 in that, rather than receiving an order reconstruction test in each of the 24 study–test blocks, participants completed a free recall test in half of the blocks and the order reconstruction test in the other half of the blocks. Test type was randomly assigned to blocks. The order reconstruction test was conducted as in Experiment 1. For the free recall test, participants were provided with a test sheet with eight lines and were instructed to recall as many items as possible. It was emphasized that they should write these items as they came to mind, rather than attempting to recreate the order in which they had studied the items.

Prior to completing the 24 study–test blocks, participants first completed a practice phase. During this practice phase, they studied eight items in a mixed list, completed the 30-s distractor task, and completed both tests, with free recall preceding order reconstruction. Both tests were included in the practice phase to ensure understanding of the instructions, but only one of the two tests occurred in each of the 24 experimental blocks.

Results and Discussion

Order reconstruction. We analyzed the data separately for each test type, focusing first on the order reconstruction tests. We found no effect of List Type, $F(1, 47) = 0.11$, $MSE = .03$, $p = .75$, or of Item Type, $F(1, 47) = 1.93$, $MSE = .03$, $p = .17$, but we found a significant List Type × Item Type interaction, $F(1, 47) = 11.95$, $MSE = .01$, $p < .01$, $\eta_p^2 = .20$. We again performed two paired-samples *t* tests, comparing aloud to silent for each of the list types. There was a tendency toward better order reconstruction for aloud than for silent items in mixed lists, but the difference was not significant, $t(47) = 1.80$, $SE = .02$, $p = .08$. Importantly, however—and replicating the result of Experiment 1—we observed a significant advantage in order reconstruction accuracy for silent items from pure lists over aloud items from pure lists, $t(47) = 2.23$, $SE = .03$, $p = .03$, $d = 0.32$. These data are shown in Figure 2A.

Free recall. In our analysis of the free recall responses, there was no effect of List Type, $F(1, 47) = 0.36$, $MSE = .01$, $p = .55$, and a significant effect of Item Type, $F(1, 47) = 40.01$, $MSE = .02$, $p < .001$, $\eta_p^2 = .46$. Of main interest, however, was the List

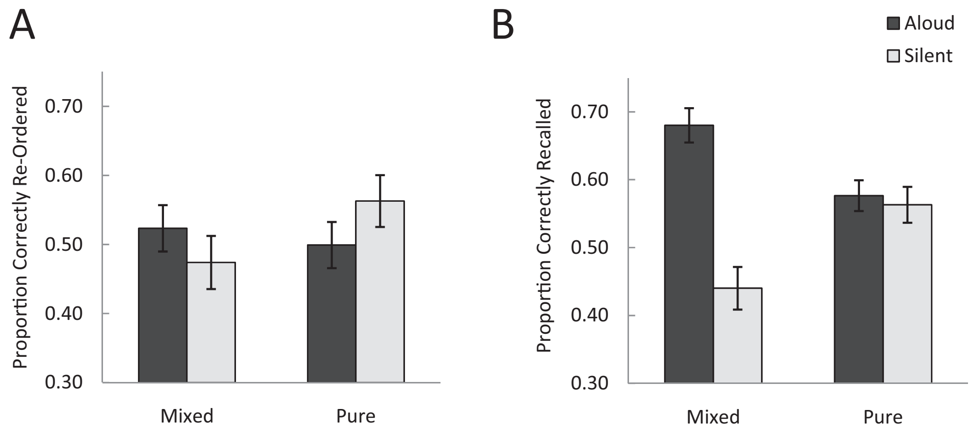

Figure 2. Experiment 2: Proportion of items that were reordered correctly on the order reconstruction test (Panel A) and proportion of items correctly recalled on the free recall test (Panel B) as a function of list type and encoding condition. Error bars represent one standard error.

Type × Item Type interaction, which was significant, $F(1, 47) = 52.56$, $MSE = .01$, $p < .001$, $\eta_p^2 = .53$. We performed two paired-samples *t* tests, comparing aloud to silent separately for each of the list types. Recall of aloud items was significantly better than that of silent items in mixed lists, $t(44) = 7.97$, $SE = .03$, $p < .001$, $d = 1.15$, but recall did not differ significantly between pure aloud lists and pure silent lists, $t(44) = 0.73$, $SE = .02$, $p = .47$. Thus, as shown in Figure 2B, we reproduced the standard production effect in the free recall data.

Another feature of the free recall data is worth highlighting as it relates to debate about whether the production effect occurs in pure lists. Here, we have the advantage of all three conditions within subject, and we observe a very clear pattern: no effect of production in pure lists (between subjects) but a robust effect in mixed lists (within subject). As Figure 2B clearly portrays, the difference in mixed lists is due to both a cost for the silent items (mixed silent < pure silent), $t(47) = 5.74$, $SE = .02$, $p < .001$, $d = 0.83$, and a benefit for the aloud items (mixed aloud > pure aloud), $t(47) = 4.39$, $SE = .02$, $p < .001$, $d = 0.63$.

Order in free recall. We were also interested in participants' use of order information during free recall and whether the patterns observed would complement those observed on the order reconstruction test. According to the item-order account, participants use relational information (i.e., order) to guide their free recall from pure silent lists. Therefore, the study order should be better retained in the free recall outputs from pure silent lists than from pure aloud lists.

To examine memory for order in the free recall responses, we performed two separate analyses. The first analysis examined whether recall was clustered in a way that paralleled study order. For example, if a participant recalled an item from Serial Position 3 of the study list, the subsequent recall could be from a nearby position, such as Position 2 or 4, which would indicate better retention of the study order, or it could be from a more distant position, such as Position 7 or 8, which would indicate poorer retention of the study order. To assess the average distance between recalled items, we scored the distance between the two members of each pair of sequentially recalled items. For example, if a participant recalled four items in the order shown in Table 1, the distance on the original study list between the two members of each successive pair of items could be calculated. The example set of recalled items would produce distance scores of 1 (flower–cushion), 3 (cushion–kitten), and 1 (kitten–auto), for an average distance of 1.67. A higher score indicates a greater average distance between items recalled.[1]

We could not compute separate distance scores for aloud and silent items in mixed lists precisely because the two item types were intermixed during recall; therefore, we computed one distance score for mixed lists, collapsing across item type. Overall, the average distance between recalled items differed for pure silent ($M = 2.22$), pure aloud ($M = 2.50$), and mixed lists ($M = 2.60$), $F(2, 94) = 5.73$, $MSE = 0.32$, $p < .01$, $\eta_p^2 = .11$. Follow-up analyses revealed that the average distance between recall outputs was smaller for pure silent lists than for either mixed lists or pure aloud lists, $t(47) = 3.35$, $SE = .11$, $p < .01$, $d = 0.48$, and $t(47) = 2.28$, $SE = .12$, $p = .03$, $d = 0.33$, respectively, but that mixed lists and pure aloud lists did not differ from each other, $t(47) = 0.92$, $SE = .11$, $p = .36$. Thus, when recalling from pure silent lists, each subsequent recall was from a nearer serial position from the study list compared to when recalling from pure aloud lists or mixed lists.

The second part of the follow-up analyses of recall order examined the input–output correspondence (adapted from Asch & Ebenholtz, 1962). To analyze free recall data in this way, we determined whether the second item in each successive pair of items occurred later in the study list than the first item in that pair. If so, the pair was identified as being recalled in a forward direction; otherwise, the pair was identified as being recalled in a backward direction. In the example given in Table 1, the first two pairs of items were recalled in a forward direction (i.e., flower–

[1] We did not analyze the last response if participants recalled all eight items because if participants had recalled seven items already then the serial position of the eighth item would be determined because it would be the last possible item in the list.

Table 1
Example of a Study List (Input) Sequence and a Free Recall Test (Output) Sequence

Study sequence	Recall test sequence
1. Coat	1. Flower (3)
2. Mug	2. Cushion (4)
3. Flower	3. Kitten (7)
4. Cushion	4. Auto (6)
5. Friend	
6. Auto	
7. Kitten	
8. Paper	

cushion, cushion–kitten), but the last pair was recalled in a backward direction (i.e., kitten–auto). Thus, the proportion of pairs recalled in a forward direction in this example is .67. A higher score means a higher match between the studied order and the order of recall; in other words, there is greater input–output correspondence. Overall, the proportions of pairs with sequence information recalled in the forward direction differed significantly for pure silent (M = .72), pure aloud (M = .62), and mixed (M = .61) lists, $F(2, 94) = 8.95$, $MSE = .02$, $p < .001$, $\eta_p^2 = .16$. Follow-up analyses revealed that the average input–output correspondence between words recalled for silent lists differed from that of both mixed lists and aloud lists, $t(47) = 3.72$, $SE = .03$, $p < .01$, $d = 0.54$, and $t(47) = 3.28$, $SE = .03$, $p < .01$, $d = 0.47$, respectively, but that mixed lists and aloud lists did not differ from each other, $t(47) = 0.28$, $SE = .03$, $p = .78$. Thus, when recalling words from pure silent lists, participants were more likely to recall in a forward direction than when recalling from pure aloud lists or mixed lists.

These two follow-up analyses revealed that, although there were no differences in the *number* of items recalled for pure silent lists versus pure aloud lists in this experiment, there were clear differences in the *patterns* of recall. When performing free recall, participants were more likely to retain the relative order of studied items from pure silent lists than from either of the other list types. Although they showed no differences in mean number of items recalled for pure silent lists and pure aloud lists, their retention of order and sequence information was superior for pure silent lists. This suggests that participants use order information to guide their recall from pure silent lists, whereas they likely rely on item-specific information when recalling from pure aloud lists. This finding is entirely consistent with the item-order account, and it confirms that the production effect fits with McDaniel and Bugg's (2008) class of encoding techniques.

General Discussion

In two studies, we examined memory for order information when items were read aloud or silently, in both pure and mixed lists. In both experiments, participants' reconstructions of the studied order of pure silent lists were superior to their reconstructions of pure aloud or mixed lists. In other words, we found a reverse production effect for memory for the studied order in pure lists. We also found that participants were more likely to recall together items that were studied in close proximity in pure silent lists compared to pure aloud lists and mixed lists. Additionally, when analyzing the retained relative order of each successive pair of items recalled, we found higher input–output correspondence for pure silent lists than for pure aloud lists or mixed lists.

Taken together, this pattern of results suggests that the production effect belongs to McDaniel and Bugg's (2008) class of encoding techniques. Indeed, the production effect functions much like the generation effect and the enactment effect, among others, on tests of recall and of memory for order. In Experiment 2, the recall of commonly processed items from mixed lists was much poorer than recall of these items from pure lists, and the recall of unusually processed items from mixed lists was much better than recall of these items from pure lists. In other words, mixed-list recall sustains both a cost and benefit relative to pure lists (cf. Bodner et al., 2013). According to the item-order account, the commonly processed items incur a cost in mixed lists because the presence of the unusually processed items disrupts the encoding of relational information. Indeed, analyses both of the distance between sequentially recalled items and of the input–output correspondence demonstrated that participants better retained the order of silent items from pure lists than of aloud items from pure lists or of all items from mixed lists. Furthermore, the aloud items may benefit in mixed lists because, relative to the pure lists, there are fewer aloud items in mixed lists, so the set size of these items with item-specific information is smaller, making it easier to remember them.

An interesting feature of our findings is that we observed no pure-list production effect. To our knowledge, this is the first report of a pure-list examination of the production effect using a recall test; all previous explorations of the production effect in pure lists have used only recognition tests. In our Experiment 2, which involved recall, there was no evidence of a pure-list production effect. This result can be contrasted with the results of Gathercole and Conway (1988) and of the meta-analysis by Fawcett (2013); these cases suggest that there might be a small pure-list production effect in recognition. The existence of a small between-subjects production effect in recognition accords with the predictions of the item-order account. According to the account, commonly processed items benefit from relational information during recall, but such structural information is either not useful or less useful during a randomly ordered recognition test. Unusually processed items, on the other hand, benefit from the encoded item-specific information in both recall and recognition. Therefore, the item-order account predicts that unusually processed items will benefit from item-specific information irrespective of test type, whereas commonly processed items will benefit from relational information only on recall tests. The consequence of this is that of a between-subjects production effect in recognition but not in recall, which agrees with our findings as well as with those of Fawcett.

The production effect is an important addition to McDaniel and Bugg's (2008) class of encoding techniques because it is very simple to carry out, unlike the encoding techniques listed by McDaniel and Bugg. For example, generation requires a unique cue, making it difficult to develop a large stimulus set. Similarly, it is difficult to develop a large set of enactment stimuli or bizarre images. Furthermore, it is unclear whether items that are not generated successfully should be removed from analyses. If they are, the consequence is a smaller set size of generated items compared to read items. If they are not, they might artificially lower the later recall rate because participants are expected to recall those items even though they were never actually exposed to the correct target. The production effect is not subject to these pitfalls. It can be employed using arbitrarily large lists, participants rarely—if ever—make errors, and the aloud/silent manipulation can easily be randomly assigned.

Although the predictions of the item-order account were clearly laid out by McDaniel and Bugg (2008), there surely will be important research questions that arise from this account. For example, there is a more general paradox for the item-order account that will have to be resolved: Typically, the evidence for the item-order account comes from studies using short lists, but the argument for the key role of item order in design-driven patterns has been applied to studies that use long lists. This concern was originally raised by Engelkamp, Selier, and Zimmer (2004) for action memory and was followed up using verbal and pictorial stimuli by Mulligan and Lozito (2007). It is our hope that future investigations will shed light on this paradox, and the production manipulation provides a straightforward testing ground for further exploration of the item-order account.

To date, the dominant theoretical explanation of the production effect has relied on a *distinctiveness* mechanism. A distinctiveness account emphasizes uniqueness at the item level, relative to all other items in a context (see Hunt, 2013). For example, a single red fish in an aquarium might be distinct when all other fish are blue because it has the unique feature of *red* in the context of *blue.* This uniqueness benefits memory for the distinct item. Intriguingly, a distinctiveness account has been used to address a number of other memory phenomena that appear to be robust for mixed study lists but not for pure study lists (e.g., the bizarreness effect; McDaniel & Einstein, 1986) by emphasizing the uniqueness of the unusually processed items. Such an account can explain the recall advantage for the aloud items in mixed lists but struggles to explain the differences in memory for word order in pure lists. Thus, as McDaniel and Bugg (2008) discussed, a distinctiveness mechanism does not appear to be sufficient for explaining the recall patterns observed for their class of encoding techniques. This seems to be the case with the production effect as well. The present results suggest that a distinctiveness account might be complemented by the item-order account or even that the item-order account could provide an entirely new framework for a better conceptualization of the production effect. In either case, our results clearly demonstrate that order retention plays a role in the production effect just as it does in a number of other extensively explored memory phenomena and therefore that accounts of the production effect must acknowledge the role of memory for order in recall.

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 and Applied, 54,* 3–31. doi:10.1080/00223980.1962.9713093

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2013). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-013-0485-1

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition, 41,* 28–35. doi:10.3758/s13421-012-0249-6

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22,* 267–281. doi:10.1111/j.1467-9450.1981.tb00402.x

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33*(A), 497–505.

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

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 671–682. doi:10.1037/0278-7393.26.3.671

Engelkamp, J., Selier, K. H., & Zimmer, H. D. (2004). Memory for actions: Item and relational information in categorized lists. *Psychological Research, 69,* 1–10. doi:10.1007/s00426-003-0160-7

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

Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2013). Production improves memory equivalently following elaborative vs. non-elaborative processing. *Memory*. Advance online publication. doi:10.1080/09658211.2013.798417

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

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior, 11,* 534–537. doi:10.1016/S0022-5371(72)80036-7

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 62,* 242–246. doi:10.1037/1196-1961.62.4.242

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195169669.003.0001

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science, 22,* 10–15. doi:10.1177/0963721412463228

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62,* 227–239. doi:10.1016/j.jml.2009.11.010

Lin, O. Y. H., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66,* 212–216. doi:10.1037/a0028309

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98,* 291–310. doi:10.1016/S0001-6918(97)00047-4

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review, 18,* 1197–1202. doi:10.3758/s13423-011-0168-8

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

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15,* 237–255. doi:10.3758/PBR.15.2.237

McDaniel, M. A., Cahill, M., Bugg, J. M., & Meadow, N. G. (2011). Dissociative effects of orthographic distinctiveness in pure and mixed lists: An item-order account. *Memory & Cognition, 39,* 1162–1173. doi:10.3758/s13421-011-0097-9

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

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, and Cognition, 25,* 54–69. doi:10.1037/0278-7393.25.1.54

Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition, 30,* 850–861. doi:10.3758/BF03195771

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. doi:10.1080/17470210600785141

Nairne, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 248–255. doi:10.1037/0278-7393.14.2.248

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 702–709. doi:10.1037/0278-7393.17.4.702

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40,* 326–338. doi:10.3758/s13421-011-0165-1

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

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1543–1547. doi:10.1037/a0020604

Ozubko, J. D., Major, J., & MacLeod, C. M. (in press). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*.

Peterson, D. J., & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1287–1293. doi:10.1037/a0031337

Quinlan, C. K., & Taylor, T. L. (in press). Enhancing the production effect in memory. *Memory*.

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21,* 34–40. doi:10.3758/BF03211162

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

Received July 2, 2013
Revision received September 5, 2013
Accepted September 23, 2013 ■