Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 6, 1543–1547

0278-7393/10/$12.00 DOI: 10.1037/a0020604

The Production Effect in Memory: Evidence That Distinctiveness Underlies the Benefit

Jason D. Ozubko and Colin M. MacLeod
University of Waterloo

The production effect is the substantial benefit to memory of having studied information aloud as opposed to silently. MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010) have explained this enhancement by suggesting that a word studied aloud acquires a distinctive encoding record and that recollecting this record supports identifying a word studied aloud as "old." This account was tested using a list discrimination paradigm, where the task is to identify in which of 2 studied lists a target word was presented. The critical list was a mixed list containing words studied aloud and words studied silently. Under the distinctiveness explanation, studying an additional list all aloud should disrupt the production effect in the critical list because remembering having said a word aloud in the critical list will no longer be diagnostic of list status. In contrast, studying an additional list all silently should leave the production effect in the critical list intact. These predictions were confirmed in 2 experiments.

Keywords: memory, production, distinctiveness, recognition, aloud

The category of encoding techniques that substantially benefit memory is disappointingly small. Of course, there is rehearsal (e.g., Rundus, 1971), the most intuitive way to increase memory. And every mnemonics book emphasizes imagery, which dramatically improves memory (e.g., Paivio, 1971). Since Craik and Lockhart's (1972) introduction of levels of processing, semantic elaboration has also been well recognized, as has generation since the seminal article of Slamecka and Graf (1978). A few other candidates might be nominated, but the list would remain short. So why would we ignore a potential member of this set?

In 1972, in a rarely cited article, Hopkins and Edwards reported a remarkably simple mnemonic phenomenon that they referred to as a pronunciation effect. Subjects studied a list of words and later performed a recognition test. During study, half of the words were underlined and half were not. In the between-subjects conditions, subjects were to ignore the underlining and either read all of the words aloud or read all of them silently. Recognition was unaffected. However, in the within-subjects condition, where subjects were to read one set aloud (e.g., underlined) and the other set silently (e.g., nonunderlined), words studied aloud were recognized about 10% better than those studied silently.

This article was published Online First August 30, 2010.

Jason D. Ozubko and Colin M. MacLeod, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

This research was supported by an Alexander Graham Bell Canada Graduate Scholarship and by Discovery Grant A7459, both from the Natural Sciences and Engineering Research Council of Canada. We thank Emily Bryntwick and Molly Pottruff for their assistance with the data collection and Ian Dobbins for helpful comments during the review process.

Correspondence concerning this article should be addressed to Jason D. Ozubko or to Colin M. MacLeod, Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1. E-mail: jdozubko@uwaterloo.ca or cmacleod@uwaterloo.ca

For purposes not having to do with the pronunciation enhancement per se, this advantage of saying some words aloud has been replicated a few times in the ensuing almost 40 years (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; MacDonald & MacLeod, 1998) but has left little impression in the literature. In reintroducing and more thoroughly delineating this phenomenon, MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010) have rechristened it with a more general name: the *production effect*.

One thing that makes the production effect noteworthy is its similarity to the generation effect (Slamecka & Graf, 1978), one of the most widely used manipulations in memory research (see Slamecka, 1992). The generation effect has been studied directly or indirectly in hundreds of research articles. In a recent meta-analytic review, Bertsch, Pesta, Wiscott, and McDaniel (2007) suggested that the average improvement that generation provides on a recognition test is roughly 10%. Remarkably, despite its simplicity, the published evidence suggests that the production effect is similar in magnitude to the generation effect.

Both techniques involve producing a word. However, generation appears to require more cognitive effort and semantic analysis because subjects must actually retrieve words, as opposed to the production effect where words are explicitly provided. If the generation effect is taken as the hallmark of an excellent way to increase memorability, then it is indeed surprising to find that a simpler manipulation that does not necessarily lead to deeper semantic processing can improve memory so much. Intriguingly, the production effect even confirms the common intuition that studying aloud improves retention. For these reasons, the production effect warrants more investigation and greater visibility.

What Causes the Production Effect?

To explain the production effect, MacLeod et al. (2010; see also Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988) have championed a distinctiveness account—that production at study makes words distinctive and that this distinctiveness can be used heuristically at test: "I remember saying that aloud so it must be old." Subjects cannot use the opposite strategy to identify silently studied words ("I remember *not* saying that word aloud"), because the distracters at test were not said aloud during study either. MacLeod et al. cast this account in terms of the replaying of an encoding record, in line with the proceduralist view of memory (Kolers, 1973; Kolers & Roediger, 1984). In support of this distinctiveness account, MacLeod et al. observed that the effect persisted when subjects mouthed (but did not speak) words at study, because mouthing still provides distinctive responses to each word. However, when subjects repeatedly pressed a key or repeatedly said "yes" to some words at study, no memory advantage was observed.

Perhaps most compelling, MacLeod et al. (2010) confirmed that the production effect occurs within subjects but not between subjects (see also Dodson & Schacter, 2001; Hopkins & Edwards, 1972). This finding is in line with other effects believed to be driven by distinctiveness. For example, the orthographic distinctiveness effect—the finding that orthographically atypical words are remembered better than less atypical words—occurs only in mixed lists, not in pure lists (Hunt & Elliot, 1980; Hunt & Mitchell, 1982). Similarly, superior memory for bizarre sentences over common sentences, believed to be driven by the distinctiveness of the bizarre sentences, occurs only in mixed lists, not in pure lists (Waddill & McDaniel, 1998). McDaniel and Bugg (2008) have recently presented a thorough analysis of this design difference, which encompasses many well-established encoding effects, notably, generation, enactment, word frequency, perceptual interference, and bizarreness.

Despite the similarities between the production effect and other distinctiveness-driven phenomena, a direct test of the idea that production relies on distinctiveness has not yet been conducted. Our goal here was to undertake that test. To do so, we sought to disrupt the production effect in a manner that is clearly predicted by the distinctiveness account. If this account is correct, then production will lead to superior recognition at test because a distinctive element forms part of the encoding record of each word studied aloud, an element not shared by the silent words that were also studied or by the distracters that appeared only on the test. Logically then, if the distracters could be made to share this element with the aloud words, this should undermine or even eliminate the distinctiveness that ordinarily causes the production effect, causing the effect to diminish or disappear.

Experiment 1

Our goal was to test the production effect in a paradigm where the distinctiveness account predicts that the effect should be disrupted. Subjects studied two separate lists: a critical mixed list and a distracting pure list. The critical list always included half words to be spoken aloud and half words to be read silently. The key manipulation was on the distracting list, where the words could be studied either all aloud or all silently. At test, subjects were required to discriminate whether each word came from the critical mixed list or from the distracting pure list. Thus, all of the test words had been studied: The test was a list discrimination task (see, e.g., Hintzman, Caulton, & Levitin, 1998; Hintzman & Waters, 1970), not a yes–no recognition task.

Under the distinctiveness account, on a standard recognition test, recollecting that a word was read aloud ordinarily is useful for identifying it as studied because the distracters were not read aloud. In list discrimination, however, because words from both the critical mixed list and the all aloud distracting pure list were read aloud at study, recollecting that a word was read aloud should not assist in identifying the word's source. Furthermore, failing to recall that a word was read aloud should not be particularly diagnostic, as this could indicate that the word was read silently or was read aloud but not encoded well. In other words, the absence of a feature at retrieval is not particularly strong evidence that a feature was absent at encoding (although a small effect might conceivably be observed here because this is somewhat informative). However, crucially, when the distracting pure list is read silently, a production effect should be observed because now recollecting that a word was read aloud at study is useful for list discrimination. Any other data pattern would be inconsistent with the distinctiveness account.

Method

Subjects. Thirty-six University of Waterloo students participated for course credit. For the between-subjects distracting pure list manipulation, 18 students were randomly assigned to the all aloud condition and 18 to the all silent condition.

Stimuli and apparatus. A pool of 120 words was taken from the Appendix in MacDonald and MacLeod (1998). The words were all nouns from 5 to 10 characters long with frequencies greater than 30 per million (Thorndike & Lorge, 1944). The experiment was programmed in E-Prime Version 1.2 (www.pstnet.com) and was carried out on IBM PC-compatible computers with 15-in. color monitors.

Procedure. For each subject, the 120 words were randomly ordered and two sets of 32 words were selected to form the critical mixed list and the distracting pure list. For the critical mixed list, half of the words were randomly selected to appear in blue and half in white, with color order random through the list. For the distracting pure list, the words appeared either all in blue (aloud) or all in white (silent). Subjects were told to read blue words aloud and white words silently.

During study, words were presented individually at the center of the screen for 2 s with a 0.5-s interstimulus interval. Prior to the test phase, subjects were instructed that all of the words had been studied (i.e., that all words were "old") and that their task was to identify which list each word came from. To identify a word as belonging to the first (mixed) list, subjects pressed *c*; to identify a word as belonging to the second (pure) list, they pressed *m*.

Results

The results are shown in the top portion of Table 1. An alpha level of .05 was applied throughout the analyses. Performance on the distracting pure list did not differ as a function of whether it was all aloud or all silent, $t(34) = 0.79$, $p = .44$, $d = 0.31$. This is consistent with the absence of a between-subjects production effect previously reported by Hopkins and Edwards (1972), Dodson and Schacter (2001), and MacLeod et al. (2010).

Table 1
Results of Experiments 1 and 2

	Mixed list				Pure list	
	Aloud		Silent			
Pure list type	*M*	*SE*	*M*	*SE*	*M*	*SE*
Experiment 1: Critical mixed list first						
All silent	.76	.05	.65	.05	.61	.04
All aloud	.56	.04	.62	.04	.65	.02
Experiment 2: Critical mixed list second						
All silent	.77	.04	.50	.04	.67	.03
All aloud	.62	.03	.64	.04	.65	.03

Note. Mean proportions (and standard errors) of aloud versus silent studied words correctly assigned to the critical mixed list are shown separately as a function of the type of distracting pure list (all aloud vs. all silent). Also shown are the mean proportions (and standard errors) of words correctly assigned to the distracting pure lists.

Our focal analysis was a 2 (mixed list word type: aloud vs. silent; within subjects) $\times$ 2 (pure list word type: all aloud vs. all silent; between subjects) mixed analysis of variance (ANOVA) conducted on the accuracy scores for the critical mixed list words. Overall recognition of aloud and silent critical mixed list words was equivalent, $F(1, 34) < 1$. When the distracting pure list contained only aloud words, accuracy for the critical mixed list words was lower than when the distracting pure list contained only silent words, $F(1, 34) = 7.09$, $MSE = 0.04$, $p\eta^2 = .17$.

Critically, consistent with the distinctiveness account, the interaction was significant, $F(1, 34) = 5.56$, $MSE = 0.02$, $p\eta^2 = .14$. As Table 1 clearly shows, there was a reliable production effect for the critical mixed list only when all words on the distracting pure list were studied silently, $t(17) = 2.13$, $d = 0.50$, with 14 of the 18 subjects (78%) showing the effect. The production effect disappeared when all words on the distracting pure list were studied aloud, $t(17) = 1.24$, $d = 0.30$, with only five of the 18 subjects (28%) showing the effect.

Discussion

Consistent with the prediction of the distinctiveness account, the production advantage usually seen for a mixed list of aloud and silent words was eliminated when the words in an additional distracting list were all read aloud. The production effect was, however, robust (11%) when the words in the distracting list were all read silently. This fits with the idea that, at test, subjects are recollecting having read a word aloud during study and are using this processing record to improve memory accuracy for the aloud words. Such recollection succeeds because of the distinctiveness of the aloud words in a mixed list only when no other aloud words are present, as in the all silent distracting list. But when other words are said aloud during study, as in the all aloud distracting list, such recollection is not diagnostic of list status, and consequently the production effect disappears.

Experiment 2

Experiment 2 constituted a conceptual replication of Experiment 1. Only one change was made: This time, the distracting pure list preceded the critical mixed list, representing a situation of proactive influence. The same prediction was made as for Experiment 1.

Method

Subjects. Thirty-six students from the same pool took part for the same credit. Half were randomly assigned to the all aloud and half to the all silent pure list condition.

Stimuli and apparatus. The same stimuli and apparatus were used as in Experiment 1.

Procedure. The procedure was identical to that of Experiment 1 except that the distracting pure list now preceded the critical mixed list, rather than following it.

Results

The results are shown in the bottom half of Table 1. As in Experiment 1, performance on the all aloud versus all silent pure lists did not differ between subjects, $t(34) = 0.55$, $p = .58$, $d = 0.19$. This again is consistent with the uniform absence of a between-subjects production effect in prior studies (Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010).

The same 2 $\times$ 2 mixed ANOVA was used as in Experiment 1. Again, there was no overall difference in critical mixed list accuracy when the distracting pure list was all aloud versus all silent, $F(1, 34) < 1$. This time, however, aloud study words were recognized significantly better overall than were silent study words, $F(1, 34) = 14.41$, $MSE = 0.02$, $p\eta^2 = .30$, likely because the critical mixed list was studied immediately before the test.

Most important, as in Experiment 1, critical mixed list word type interacted with distracting pure list type, $F(1, 34) = 20.53$, $MSE = 0.02$, $p\eta^2 = .38$. There was a large and reliable production effect when the distracting pure list was all silent, $t(17) = 6.10$, $d = 1.44$, but no production effect when the distracting pure list was all aloud, $t(17) = 0.50$, $d = 0.12$. The production advantage was present for 16 of 18 subjects (89%) when the pure list was all silent but for only seven of 18 subjects (39%) when the pure list was all aloud.

Discussion

In sum, the results of Experiment 2 very closely paralleled those of Experiment 1. Again, a robust production effect (27%) occurred in the critical mixed list only when the distracting pure list comprised all silent words; it was eliminated when the distracting pure list comprised all aloud words. A 2 × 2 × 2 ANOVA incorporating experiment as a variable further demonstrated the consistency of the experiments, with no three-way interaction ($p > .22$). The results of the two experiments are, therefore, thoroughly consistent with each other and with the predictions of the distinctiveness account.

General Discussion

The production effect is a simple yet effective means to improve memory, as MacLeod et al. (2010) have argued. Indeed, it evidently leads to substantial improvement, so its virtual invisibility in the literature is surprising. To date, the few published studies examining the benefit of production (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972; Hourihan & MacLeod, 2008; MacDonald & MacLeod, 1998; MacLeod et al., 2010) have produced data largely consistent with a distinctiveness account but have not directly tested the main mechanism underlying that account. Our goal was therefore to interfere with the production effect in a manner that should undermine the distinctiveness account's driving mechanism.

Using a list discrimination test, Experiment 1 demonstrated that when a second list is studied after the critical mixed list, the production effect for the critical mixed list is eliminated when the words from that second list are all said aloud but remains robust when they are all read silently. Experiment 2 confirmed this pattern when the words in the extra list were instead studied before the critical mixed list. These results are thoroughly consistent with a distinctiveness account: Only when recollection that a word was said aloud is diagnostic of list status should a production effect emerge.

These experiments confirm that recollective distinctiveness—and hence memory—can be improved via production. Production provides extra distinctive information which can be used heuristically at test to improve performance. This phenomenon bears some similarity to other recent work that emphasizes the heuristic value of using distinctive information to reject distracters—the *recall-to-reject* strategy (see, e.g., Clark & Gronlund, 1996; see also the idea of *criteria recollection*, e.g., Gallo, Weiss, & Schacter, 2004). For example, using a contextual encoding manipulation, Dobbins, Kroll, Yonelinas, and Liu (1998) found that items encoded in more distinctive contexts were better remembered than items encoded in less distinctive contexts but that this benefit arose primarily from participants' ability to reject distracters. In essence, distinctive information did not benefit recall or hit rates but did help participants reject novel test probes.

Generally, then, studies that find that participants use discriminative information at test to strategically reject distracters also find that distinctive information does little to boost hit rates. It is important to realize, though, that both benefits of distinctiveness—more hits and more correct rejections—have been previously reported, notably by Hunt (2003). Because in traditional recognition paradigms, production seems primarily to improve memory for target items (see MacLeod et al., 2010), production provides an instance where distinctive information is used to increase hit rate, whereas the recall-to-reject work focuses more on instances where distinctive information is used to decrease false alarm rate.

Our work also bears on simple strength accounts of production. Recall that our distinctiveness account predicted that a production effect should emerge only when recollection that a word was said aloud is diagnostic of list status. A straightforward strength account (see, e.g., Hovland, 1951; Wickelgren, 1969) would not readily make this prediction. Instead, appealing to the idea of repetition incrementing strength (see Murdock, 1989, for a review), a strength account would say that words said aloud are encoded more strongly than those said silently. If so, then the nature of the distracting pure list should not selectively weaken the stronger aloud words more than the weaker silent words in the critical mixed list. Indeed, a strength account would also predict that if the encoding of aloud words is stronger than that of silent words this should be true even in the distracting pure list situation. But studies have repeatedly shown that there is no production effect between subjects, a finding which was confirmed here in that the all aloud and all silent distracting pure list items were equally well discriminated. The strength view is therefore wrong on both counts.

A major empirical appeal of the production effect is its simplicity. Consequently, it can easily be used in diverse situations. For example, as a quick method for increasing the distinctiveness of some items over others, the production effect is far easier to use than the generation effect. Generation requires the creation of extra cues from which the items will be generated. Further, in generation, subjects occasionally make "mistakes," generating an item other than the intended one. None of these issues exist for production because the word being studied is its own cue. The production effect may therefore prove to be a useful technique in study situations, or whenever relatively quick and easy memory enhancement is desired. As noted earlier, the existence of the production effect fits with the common notion that information studied aloud is better remembered.

In this article, we sought to examine the theoretical basis of the production effect. Our findings are entirely consistent with a distinctiveness account: Production makes items stand out at study, and recollection of this is useful at test. When distinctiveness is undermined, the production effect vanishes. Clearly, any adequate theory needs to successfully predict not only when an effect will be present but also when it will be disrupted or eliminated. So far, the distinctiveness account has met this requirement for the production effect.

References

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35,* 201–210.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684. doi:10.1016/S0022-5371(72)80001-X

Dobbins, I. G., Kroll, N. E. A., Yonelinas, A. P., & Liu, Q. (1998). Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. *Journal of Memory and Language, 38,* 381–400. doi:10.1006/jmla.1997.2554

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8,* 155–161.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51,* 473–493. doi:10.1016/j.jml.2004.06.002

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

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition, 26,* 449–462.

Hintzman, D. L., & Waters, R. M. (1970). Recency and frequency as factors in list discrimination. *Journal of Verbal Learning and Verbal Behavior, 9,* 218–221. doi:10.1016/S0022-5371(70)80054-8

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and 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, 62,* 242–246. doi:10.1037/1196-1961.62.4.242

Hovland, C. I. (1951). Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 613–689). Oxford, England: Wiley.

Hunt, R. R. (2003). Two contributions of distinctive processing to accurate memory. *Journal of Memory and Language, 48,* 811–825. doi:10.1016/S0749-596X(03)00018-4

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General, 109,* 49–74. doi:10.1037/0096-3445.109.1.49

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 81–87. doi:10.1037/0278-7393.8.1.81

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition, 1,* 347–355.

Kolers, P. A., & Roediger, H. L., III. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 425–449. doi:10.1016/S0022-5371(84)90282-2

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., 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

Murdock, B. B., Jr. (1989). Learning in a distributed memory model. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree Symposium on Cognition* (pp. 69–106). Hillsdale, NJ: Erlbaum.

Paivio, A. (1971). *Imagery and verbal processes.* Oxford, England: Holt, Rinehart & Winston.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77. doi:10.1037/h0031185

Slamecka, N. J. (1992). Delineation of the generation effect—A citation-classic commentary on "The generation effect: Delineation of a phenomenon," by Slamecka, N. J. & Graf, P. *Current Contents/Social & Behavioral Sciences, 18,* 10.

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

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York, NY: Columbia University Teacher's College.

Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition, 26,* 108–120.

Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology, 6,* 13–61. doi:10.1016/0022-2496(69)90028-5

Received August 11, 2009
Revision received May 11, 2010
Accepted May 26, 2010 ■