Mem Cogn (2012) 40:1046–1055
DOI 10.3758/s13421-012-0210-8

Widening the boundaries of the production effect

Noah D. Forrin · Colin M. MacLeod · Jason D. Ozubko

Published online: 21 April 2012

Abstract Words that are read aloud are more memorable than words that are read silently. The boundaries of this *production effect* (MacLeod, Gopie, Hourihan, Neary, & Ozubko, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671–685, 2010) have been found to extend beyond speech. MacLeod and colleagues demonstrated that mouthing also facilitates memory, leading them to speculate that any distinct, item-specific response should result in a production effect. In Experiment 1, we found support for this conjecture: Relative to silent reading, three unique productions—spelling, writing, and typing—all boosted explicit memory. In Experiment 2, we tested the sensitivity of the production effect. Although mouthing, writing, and whispering all improved explicit memory when compared to silent reading, these other production modalities were not as beneficial as speech. We argue that the enhanced distinctiveness of speech relative to other productions—and of other productions relative to silent reading—underlies this pattern of results.

Keywords Memory · Word production · Word recognition · Reading

The *production effect* is the finding that people have better explicit memory for words that they read aloud relative to words that they read silently (Hourihan & MacLeod, 2008; MacLeod, 2011; MacLeod, Gopie, Hourihan, Neary & Ozubko, 2010; Ozubko & MacLeod, 2010). Although the benefit of vocalization for long-term memory had received periodic research attention (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Gregg & Gardiner, 1991; Hopkins & Edwards, 1972; Kurtz & Hovland, 1953; MacDonald & MacLeod, 1998; Rosenbaum, 1962), MacLeod and colleagues have recently brought this phenomenon to the fore. They have reported that "production" is a robust mnemonic that enhances both recognition (MacLeod et al., 2010) and recall (Lin & MacLeod, in press; see also Conway & Gathercole, 1987, Exp. 3) and that compares favorably, in terms of its benefits, to established techniques such as generation (Slamecka & Graf, 1978) and enactment (Engelkamp & Krumnacker, 1980).

In explaining the production effect, MacLeod and colleagues (2010; MacLeod, 2011; Ozubko, Gopie & MacLeod, 2012; Ozubko & MacLeod, 2010) have posited a distinctiveness account (inspired by the *relational-distinctiveness* account of Conway & Gathercole, 1987). Relative to silent reading, reading a word aloud involves the encoding of an additional dimension that stands out as distinct—speech. According to the proceduralist framework (Kolers, 1973; Kolers & Roediger, 1984), the process of vocalizing at study will be retained in a record of that processing. In an explicit memory test, participants can then retrieve this distinctive speech information to determine whether a word was studied. In short, remembering having said a word aloud provides confirmatory evidence that it was studied ("I remember saying that word out loud, so I must have studied it").

Consistent with this distinctiveness account, MacLeod and colleagues (2010) found that the production effect was

N. D. Forrin (✉) · C. M. MacLeod · J. D. Ozubko
Department of Psychology, University of Waterloo,
200 University Avenue West,
Waterloo, Ontario N2L 3G1, Canada
e-mail: nforrin@uwaterloo.ca

C. M. MacLeod
e-mail: cmacleod@uwaterloo.ca

J. D. Ozubko
e-mail: jdozubko@rotman-baycrest.on.ca

limited to within-subjects, mixed-list designs, in which both spoken and silent study items occurred. Distinctiveness is relative (Hunt, 2006): Without silent items in the same list, the process of vocalizing would not stand out as distinct. Indeed, when MacLeod et al. used a between-subjects design featuring "pure" lists of either spoken or silent items, there was no longer a reliable memory advantage for words studied aloud (see also Dodson & Schacter, 2001; Hopkins & Edwards, 1972).[1]

A second boundary condition identified by MacLeod et al. (2010) is that the "produced" responses must uniquely identify studied items (as is the case when words are read aloud). MacLeod et al. found that producing an unrelated response to study items—either by saying "yes" or by pressing an arbitrary key—did not bolster memory for these items relative to silent reading (their Exp. 4). Importantly, words do not have to be read aloud to benefit from production, since mouthing words also resulted in a robust production advantage (their Exp. 5). These findings led MacLeod et al. to conclude that the key mechanism underlying the production effect is distinct, item-specific responding—regardless of whether that response is acoustic.

The primary aim of the present research was to explore further the boundaries of the production effect. In this regard, we are taking up where MacLeod et al. (2010) left off. In Experiment 1, we tested the boundaries of their finding that distinct, item-specific productions—including nonvocal productions—are more memorable than silently read items. We examined three such item-specific nonvocal productions—spelling (Exp. 1A), writing (Exp. 1B), and typing (Exp. 1C)—using "two-level" mixed-list designs. In line with the distinctiveness account (Conway & Gathercole, 1987; MacLeod et al., 2010), we hypothesized that each of these productions should lead to unique, item-specific responses at study, thereby making the produced items more memorable at test than words that were not produced (i.e., that were simply read silently).

In Experiment 2, we tested the sensitivity of the production effect by adding a third "level" to the mixed-list design. The purpose of this experiment was to investigate how effective three different productions—writing (Exp. 2A), mouthing (Exp. 2B), and whispering (Exp. 2C)—would be relative to speech in terms of boosting explicit memory for words. Although mouthing words has been shown to result in a production effect (MacLeod et al., 2010), it is not clear whether mouthing (or other nonvocal productions) are as memorable as speech. This was examined by including two different types of production—speech and either writing, mouthing, or whispering—in a mixed-list design that also included unproduced (i.e., silently read) items.

The distinctiveness framework (Conway & Gathercole, 1987; MacLeod et al., 2010) again guided our predictions. In a mixed-list design featuring two different modes of production, the production that includes a greater number of distinct, nonoverlapping processes should provide a larger boost to explicit memory. The more processes involved in the encoding, the more that can later be replayed to aid recognition—and distinctive processes should be particularly memorable because they are deeply encoded at study. Consistent with this account, MacLeod (2011) found that spoken words were more memorable than heard words, which were, in turn, more memorable than silently read words (see also Conway & Gathercole, 1987). Relative to silent reading, memory for heard words benefits from distinct auditory processing. Speaking, however, involves *two* distinct processing dimensions relative to silent reading: both audition and articulation.

We expected to find a congruent pattern of results for three-level mixed-list designs that feature writing (Exp. 2A) and mouthing (Exp. 2B); namely, we predicted that writing and mouthing would enhance memory relative to silent reading because each of these productions involves a distinct motor process—manual movement and articulation, respectively. We also predicted that speaking would be a more memory-enhancing production than either writing or mouthing, for speech involves two distinct processes relative to silent reading—articulation and audition—and the latter also stands out as distinct relative to writing and mouthing. Thus, spoken words should be processed more distinctively than these nonvocal productions, resulting in a larger boost to explicit memory.

The predictions regarding a three-level design that included whispering (Exp. 2C) were less straightforward. We certainly expected that whispered words would be more memorable than silently read words, because whispering involves two additional, distinct processes (articulation and audition) relative to reading silently. Any possible difference between whispering and speaking, however, was harder to anticipate. It would appear that speaking consists of the same processing as whispering, but that the auditory component is sharply reduced for whispering. This suggested to us that spoken words would still be more memorable than whispered words, but it was also possible that the two vocalized conditions would not differ.

Experiment 1

In Experiment 1, we explored the boundaries of the production effect by testing whether a set of item-specific productions would each be more memorable than silent reading.

[1] As reviewed by McDaniel and Bugg (2008), several other encoding techniques that are thought to be driven by distinctiveness operate primarily in a mixed-list design, including enactment (e.g., Engelkamp & Zimmer, 1997), generation (e.g., Begg & Snider, 1987), and bizarreness (e.g., McDaniel & Einstein, 1986).

These productions were chosen to be very common kinds of productions in everyday experience, permitting generalization of the production effect. Our participants produced words by spelling (Exp. 1A), writing (Exp. 1B), or typing (Exp. 1C). Because each modality involves a unique, item-specific type of production, we predicted that each would result in superior recognition relative to silently reading words. In each case, a mixed-list design was used, in which study words were randomly presented either in blue or in white print; the blue words were produced by the participants (spelled, written, or typed), and the white words were studied silently.

Experiment 1A: Spelling

In this initial experiment, we investigated whether the production effect would be observed for words that were spelled aloud. Although spelling words naturally involves the reuse of letters across words, each word contains a unique combination and order of letters. We therefore expected that spelling words aloud would be a distinctive—and consequently, memorable—process that could later be retrieved during the recognition test. To better equate the two study conditions, the participants were instructed that words in the silent condition were to be spelled silently, so that what differed was whether processing was done with or without production.

Method

Participants A group of 32 undergraduate students at the University of Waterloo participated in the experiment and were reimbursed with bonus course credit. In every experiment reported here, each participant was tested individually.

Stimuli and apparatus An item pool of 120 words was taken from the Appendix of MacDonald and MacLeod (1998). The words were nouns, five to ten letters long, that had frequencies greater than 30 per million (Thorndike & Lorge, 1944). A PC-compatible computer with a 17-in. color monitor was used for testing, and the controlling program was written in E-Prime version 1.2. The words were presented in 16-point Courier New font, in lowercase against a black background.

Procedure The participants were told that they would be presented with a list of words, some printed in blue and others in white, followed by an unspecified memory test. The study phase consisted of 80 words, 40 printed in blue and 40 in white, which the controlling program randomly selected from the 120-item pool. These study words were randomly intermixed and presented individually in the center of the screen. The participants were instructed to spell blue words out loud (e.g., "T-A-B-L-E") and to spell white words silently to themselves, without moving their lips or saying anything aloud. Each word stayed on the screen for 4,000 ms and was offset from adjacent words by a 500-ms blank screen. This presentation time, which is longer than has been used in past production studies, was adopted to permit spelling, which we expected to be slower than the usual reading.

A recognition test immediately followed the study phase. This recognition test consisted of 20 of the words that had been spelled aloud and 20 of the words that had been spelled silently during study, along with 20 new words.[2] All words were individually presented in the center of the screen and were printed in yellow font to avoid color overlap between study and test. The participants were asked to press the "m" key, if they recognized the word from the study phase, or to press the "c" key, if they thought that the word was new.

Results

Table 1 presents the recognition data in terms of proportions of *yes* responses, along with their respective standard errors. These are the hit rates for the studied words (spelled either aloud or silently) and the false alarm rates for the unstudied words. Table 1 also includes the hit rates and false alarm rates for Experiment 1B (writing) and Experiment 1C (typing), as well as the results for all three experiments combined.

A one-way analysis of variance (ANOVA) comparing *yes* responses for words spelled aloud, words spelled silently, and unstudied words showed a reliable overall effect, $F(2, 62) = 193.43$, $MSE = .013$, $p < .001$, $\eta_p^2 = .86$. The first planned comparison showed good memory for the studied words; we found a significant contrast between studied words (both spelled aloud and spelled silently) and unstudied words (new), $F(1, 23) = 321.81$, $MSE = .011$, $p < .001$, $\eta_p^2 = .91$. Most important, the second planned comparison revealed a reliable production effect, with a higher proportion of hits for words that were spelled aloud than for those that were spelled silently, $F(1, 23) = 19.16$, $MSE = .011$, $p < .001$, $\eta_p^2 = .38$.

Experiment 1B: Writing

Experiment 1A suggested that spelling aloud, much like reading aloud, is a unique and distinctive process—the

[2] Following the procedure of MacLeod et al. (2010), the other half of the studied words (20 blue and 20 white) appeared on a speeded reading implicit memory test, along with the remaining 20 words in the pool as unstudied baseline items. As in MacLeod et al. (2010, see their Appendix), there was no production effect on this implicit measure (i.e., words that had been spelled aloud at study were not faster to pronounce than words that had been spelled silently), so these data are not reported here.

Table 1 Proportions of *yes* responses in recognition (shown with standard errors of the respective means) as a function of study condition for Experiments 1A, 1B, 1C, and for all three experiments combined

	Spell/Write/Type		Read/Spell Silently		Not Studied	
Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*
Exp. 1A: Spell	.759	.026	.644	.024	.223	.022
Exp. 1B: Write	.795	.021	.665	.029	.193	.034
Exp. 1C: Type	.799	.022	.685	.030	.181	.026
Combined	.782	.014	.663	.016	.201	.015

memory for which can be used at test to discriminate between old and new study words. As in a "typical" production study, the distinctive process in Experiment 1A was vocal; we expected, however, that distinctive item-specific nonvocal productions should also be effective in conferring a memory advantage, as MacLeod and colleagues (2010) demonstrated with mouthing. In Experiment 1B, therefore, we shifted our focus from vocal production to orthographic production (i.e., writing). Although research by Conway and Gathercole (1987, 1990) did not show a clear memory advantage for written over silently read study words, this may have been due to the small sample sizes that they used (and, consequently, to limited power). We predicted that writing would be more memorable than silent reading because writing, like vocalizing, provides a distinct cue that has discriminative value at test.

Method

Participants A group of 24 students from the same source as in Experiment 1A participated for bonus course credit.

Stimuli and apparatus These were the same as in Experiment 1A. Ten words were added for a practice phase (see below).

Procedure The participants were instructed to silently read white words and to write down blue words using a handheld Fisher Price "Magic Erase Board" (a miniature whiteboard). The board came equipped with a dry-erase marker for writing words and a sliding bar that conveniently erased them. To prevent continued study or rehearsal of prior items, the participants were asked to immediately erase each word after writing it.

Prior to study, we presented a brief practice phase consisting of ten words (common male names) individually presented in the center of the screen, five printed in blue and five in white. The participants had unlimited time to respond to each word and to hit the spacebar to advance to the next trial. Following the practice phase, participants were instructed that the study phase was about to begin and to "try not to fall behind," because study words would now advance automatically. As in Experiment 1A, the study phase consisted of 40 blue and 40 white words randomly presented in the center of the screen. Both blue and white words were presented for 4,000 ms, offset by 1,000-ms blanks (this longer interstimulus interval, relative to the previous subexperiment, was implemented to give participants sufficient time to erase each word). The recognition test was identical to that used in Experiment 1A.

Results

A one-way ANOVA comparing *yes* responses for written, silent, and unstudied words showed a significant overall effect, $F(2, 46) = 182.74$, $MSE = .013$, $p < .001$, $\eta_p^2 = .89$. The first planned comparison showed good memory for the studied words; we found a significant contrast between studied words (written plus silent) and unstudied words (new), $F(1, 23) = 223.18$, $MSE = .015$, $p < .001$, $\eta_p^2 = .91$. As before, the second planned comparison revealed a reliable production effect, with a higher proportion of hits for written words than for silently read words, $F(1, 23) = 35.50$, $MSE = .006$, $p < .001$, $\eta_p^2 = .61$.

Experiment 1C: Typing

In Experiment 1C, we investigated whether typing study words would also lead to a production effect. We also sought to address a potential confound of Experiment 1B—that participants saw *two* versions of each written study word (the original version and the copy that they wrote), whereas they only saw one version of each silently read word. Although the participants erased their words immediately after writing them, it was still possible that the written words were more memorable, in part, because their frequency of appearance was doubled relative to silently read words.

Thus, an aim of the present study was to factor out the influence of repetition on manual production—in this case, typing. We controlled for repetition across study conditions by showing participants two copies of each silently read study word, one below the other; in that way, each silent word and each typed word would appear twice, with the second typed word being produced by the participant and echoed below the presented word.

Method

Participants A group of 26 students from the same source as in the previous experiments received bonus course credit for taking part.

Stimuli and apparatus These were the same as in the previous experiments.

Procedure The study phase again consisted of 80 words (40 blue and 40 white) that were randomly intermixed. The participants typed blue words using the keyboard. They could see their own typing, which appeared in blue font and was positioned three-quarters of the way down the screen, below each study word. The participants read white words silently without moving their lips. A duplicate of each white study word was positioned three-quarters of the way down the screen. The inclusion of this "echo" was designed to control for the fact that, when participants typed, they saw a duplicate of each blue study word in this same position.

All of the white study words appeared on the screen for 4,000 ms. After typing a blue study word, however, participants immediately pressed Enter to advance to the next word (in both conditions, successive words were offset by 250-ms blanks). If the Enter key was not pressed within 4,000 ms, the next study word would appear automatically. Participants did not find it difficult to type the words within this time limit; they were unable to complete doing so on only 0.4 % of the study trials. Consequently, the average presentation time for blue study words (M=2,469 ms) was shorter than the fixed 4,000-ms presentation time for white study words, a difference that worked against the expected memory advantage for typing over silent reading (analogous to the procedure used by MacLeod et al., 2010). The recognition test was identical to that used in the previous experiments.

Results

A one-way ANOVA comparing *yes* responses for typed, silent, and unstudied words showed a significant overall effect, $F(2, 50) = 299.35$, $MSE = .009$, $p < .001$, $\eta_p^2 = .92$. The first planned comparison showed good memory for the studied words; we found a significant contrast between studied words (typed plus silent) and unstudied words (new), $F(1, 25) = 487.44$, $MSE = .008$, $p < .001$, $\eta_p^2 = .95$. Once again, the second planned comparison was the key one, and it revealed a reliable production effect, with a higher proportion of hits for typed than for silently read words, $F(1, 25) = 22.15$, $MSE = .008$, $p < .001$, $\eta_p^2 = .47$.

Combined results

Because the designs of Experiments 1A, 1B, and 1C were almost identical, we conducted a 2×3 ANOVA, treating study condition (produced vs. unproduced) as a within-subjects variable and experiment as a between-subjects variable. Unsurprisingly, neither the main effect of experiment nor the Study Condition×Experiment interaction was reliable (Fs<1). Only the main effect of study condition was significant, $F(1, 79) = 68.56$, $MSE = .001$, $p < .001$, $\eta_p^2 = .47$, confirming that the same production effect pattern occurred in all three experiments.

Discussion

Experiment 1 demonstrated that a wide variety of different productions boost explicit memory for words. Spelling aloud, writing, and typing each improved memory by a remarkably consistent 11 %–13 % (see Table 1). The results of Experiment 1, therefore, strengthen the case that nonvocal productions can improve memory. Previously, MacLeod et al. (2010, Exp. 5) had found an advantage for mouthing, and now writing and typing can be added to the list of nonvocal productions that facilitate memory for words. Experiment 1 also showed that a vocal response other than reading—in this case, spelling aloud—causes a production advantage.

Thus, the results of Experiment 1 are consistent with MacLeod and colleagues' (2010) claim that the production of a unique, item-specific response at study underlies enhanced recognition at test. All three of the productions examined—spelling aloud, writing, and typing—involved unique responses relative to silently studied words, whether the production was acoustic and motor (spelling aloud) or orthographic and motor (writing and typing). Importantly, these unique responses conferred distinctiveness upon words at study, which then enhanced participants' ability to recognize them at test. In other words, the process of spelling aloud or writing or typing served as a distinct cue that a word had been studied (e.g., "I remember spelling that out loud"), which participants could use to assist their remembering on the recognition test.

Experiment 2

Experiment 1 revealed that the boundaries of the production effect may be very wide indeed, with several unique productions at study leading to better memory than silent reading. Yet even though that experiment demonstrated that vocalization is not an essential component of the production effect, it is still possible that vocal productions benefit memory more than nonvocal productions. Our main focus in Experiment 2, therefore, was to compare participants' memory for words that they had read out loud to their memory for words that they had produced using other unique responses: writing (Exp. 2A), mouthing (Exp. 2B), and whispering (Exp. 2C). Silently read items were also included, resulting in a three-level mixed-list design in each subexperiment.

Once again, the distinctiveness account (Conway & Gathercole, 1987; MacLeod et al., 2010) drove our predictions. First, we anticipated that participants would have better memory for words that they read aloud relative to words that they wrote or mouthed, because speech involves distinct auditory processing that is absent in these other two production modes. Whispering, however, also includes auditory processing; thus, we wanted to know whether participants would remember more words when they were spoken in a normal voice than when they were whispered.

Although we expected speech to emerge as the most memory-enhancing production, we also anticipated that the other three productions would boost memory relative to silent reading, consistent with our results from Experiment 1. Memory for written and mouthed words would both be enhanced by the addition of distinct motor processing—manual and articulatory, respectively—that is absent in silent reading. Memory for whispered words, moreover, would benefit from the addition of both motor *and* auditory processing, and might therefore be particularly memorable relative to silently read words, although perhaps not as memorable as words spoken in a normal voice, because of the reduced volume.

Experiment 2A: Writing

In Experiment 2A, the participants studied a mixed list of words that they either read aloud, wrote, or read silently. Again, we turned to the distinctiveness account (Conway & Gathercole, 1987; MacLeod et al., 2010) to inform our predictions. In a mixed study list, the more distinct processes that are involved in a production, the more effective that production should be at facilitating memory. Writing involves one distinct motor process (manual), as compared to the other study conditions. Speaking, however, involves *two* distinct processes (audition and articulation). Therefore, we expected that participants' memory for words would conform to a read aloud > write > read silently empirical gradient.

Method

Participants A group of 21 students from the same source as in the previous experiments participated. Eleven of the students were given bonus course credit, and the other ten were paid $5 each for taking part.

Stimuli and apparatus An item pool of 180 words was used, extending the 90 words used by Masson and MacLeod (1992) with a further 90 words taken from the MRC Psycholinguistic Database (www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). The controlling program was written in E-Prime, and the apparatus was the same as in Experiment 1.

Procedure The participants studied a list of 90 words randomly selected from the pool of 180. The words were individually presented in the center of the screen for 5,000 ms and were separated by 250-ms blank screens. The study conditions were indicated using icons (taken from a list of Microsoft "Webdings") that appeared immediately above each word. The participants were instructed to say the word aloud if they saw a mouth icon, to write the word if they saw a writing icon, and to read the word silently if they saw an eye icon. Thirty of the words were assigned to each condition and were presented in a random order. As was the case in Experiment 1B, the participants were given a hand-held Fisher Price Magic Erase Board to write words and were asked to erase each word immediately after writing it.

Following the study phase, the participants completed a recognition test that was identical to Experiment 1, except for the number of items. A total of 180 recognition items were presented (30 of which had been read aloud, 30 written, 30 read silently, and 90 new); thus, there were equal numbers of old and new items, and the old items comprised equal numbers of read-aloud, written, and silent words.

Results

Table 2 presents the recognition data in terms of proportions of *yes* responses, along with their respective standard errors. These are the hit rates for the studied words (either read aloud, written, or read silently) and the false alarm rates for unstudied words. Table 2 also includes the hit rates and false alarm rates for Experiment 2B (mouthing) and Experiment 2C (whispering), as well as for all three experiments combined.

The three study conditions differed reliably overall, $F(2, 40) = 24.26$, $MSE = .009$, $p < .001$, $\eta_p^2 = .55$. A clear pattern of results emerged: More words that were read aloud were recognized than words that were written, $t(20) = 5.00$, $p < .001$, and more words that were written were recognized than words that were read silently, $t(20) = 2.30$, $p < .05$.[3]

Experiment 2B: Mouthing

In Experiment 2B, mouthing took the place of writing in the three-level mixed study list. Both speaking and mouthing should result in superior memory relative to silent reading (MacLeod et al., 2010)—but would mouthing be as memorable as speaking? Once again, the distinctiveness account led us to expect an empirical gradient in which spoken would be remembered better than mouthed words which, in turn, would be remembered better than silently read words. Spoken words

[3] This pattern of results was replicated in a follow-up experiment in which participants wrote each blue word on a cue card, as opposed to using the Magic Erase Board. After writing each word, the participants immediately flipped over the cue card and placed it face down on the table. All other aspects of the design were identical. Once again, more words were recognized when they were read aloud (M=.79, SE=.022) than when they were written (M=.63, SE=.033), $t(20)$=5.00, p<.001, and more words were recognized when they were written than when they were read silently (M=.53, SE=.039), $t(20)$=2.30, p<.05. The false alarm rates were comparably low (M=.15, SE=.026). Put simply, this provided a straightforward replication.

Table 2 Proportions of *yes* responses in recognition (shown with standard errors of the respective means) as a function of study condition for Experiments 2A, 2B, 2C, and for all three experiments combined

	Read Aloud		Write/Mouth/Whisper		Read Silently		Not Studied	
Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Exp. 2A: Write	.764	.035	.631	.043	.569	.045	.175	.032
Exp. 2B: Mouth	.718	.034	.561	.031	.433	.037	.133	.016
Exp. 2C: Whisper	.661	.027	.567	.032	.428	.032	.173	.024
Combined	.705	.018	.581	.020	.465	.022	.161	.014

have both motor and auditory distinctiveness relative to silently read words, whereas mouthed words only have motor distinctiveness. Moreover, spoken words also have auditory distinctiveness relative to mouthed words.

Thus, we predicted that spoken words would be processed more distinctively than mouthed words, and mouthed words would be processed more distinctively than silently read words, leading to a spoken > mouthed > silently read pattern for recognition.

Method

Participants A group of 27 students from the same source were given bonus course credit for taking part in the experiment.

Stimuli and apparatus An item pool of 348 words was selected from the MRC Psycholinguistic Database. The words were five to ten letters long and had frequencies of greater than 30 per million (Thorndike & Lorge, 1944). The apparatus was the same as in Experiment 1.

Procedure The participants studied a list of 90 words randomly selected from the item pool and printed in three different colors: blue, red, and white. The participants read blue words aloud, mouthed red words (without making any noise), and read white words silently. Thirty words of each color were randomly intermixed and presented individually in the center of the computer monitor for 2,000 ms each, separated by 250-ms blank screens. This shorter presentation time was implemented because mouthing is a quicker production than writing, and we did not want to give participants additional time to silently rehearse each word. The recognition test format was identical to that used in Experiment 2A.

Results

The three study conditions (read aloud, mouthed, and read silently) differed reliably overall, $F(2, 52) = 45.47$, $MSE = .012$, $p < .001$, $\eta_p^2 = .64$. A pattern of results consistent with the pattern in Experiment 2A was found: More words were recognized when they were read aloud than when they were mouthed, $t(26) = 5.74$, $p < .001$, and more words were recognized when they were mouthed than when they were read silently, $t(26) = 4.53$, $p < .001$.

Experiment 2C: Whispering

In Experiment 2C, we investigated whether participants' memory for whispered words would be intermediate to their memory for normally spoken and silent words—the pattern found in the previous two experiments—with mouthing and writing both occupying the middle position of the empirical gradient. That said, similar processing would seem to underlie whispering and normal speech, in that both involve articulation and audition. Distinctiveness theory, therefore, would not predict a clear-cut advantage for normal speech over whispering, but it would predict a production effect for both. The empirical question then becomes whether the different volumes of whispering and ordinary speech would contribute to the magnitude of the production effect.

Method

Participants A group of 36 students from the same source participated in exchange for bonus course credit.

Stimuli and apparatus These were the same as in Experiment 2B.

Procedure The procedure was identical to that of Experiment 2B, except that the whisper study condition took the place of the mouthing study condition—a simple change of the instructions.

Results

The three study conditions differed reliably overall, $F(2, 70) = 43.49$, $MSE = .011$, $p < .001$, $\eta_p^2 = .55$. Once again, words that participants read aloud were the most memorable. More words were recognized when they were read aloud than when they were whispered, $t(35) = 3.98$, $p < .001$, and more words were recognized when they were whispered than when they were read silently, $t(35) = 5.37$, $p < .001$.

Combined results

Because the designs of Experiments 2A, 2B, and 2C were virtually identical, we conducted a 3 × 3 ANOVA, treating study condition as a within-subjects variable and experiment as a between-subjects variable. As we expected on the basis of the individual subexperiments, the main effect of condition was significant, $F(2, 162) = 103.80$, $MSE = .011$, $p < .001$, $\eta_p^2 = .56$. The main effect of experiment was marginal, $F(2, 81) = 2.84$, $MSE = .077$, $p = .06$, $\eta_p^2 = .066$, and was likely due to the somewhat better overall performance in Experiment 2A, the subexperiment involving writing. Most importantly, the Condition × Experiment interaction was not reliable, $F(4, 162) = 1.84$, $MSE = .011$, $p > .10$, $\eta_p^2 = .044$, which confirmed that the pattern of results over conditions was the same in all three experiments: Namely, words that were read aloud were remembered better than words produced by other unique responses (writing, mouthing, and whispering), $t(83) = 8.21$, $p < .001$, and words produced by these other responses were, in turn, remembered better than words that were read silently, $t(83) = 7.24$, $p < 001$.

Discussion

Experiment 2 demonstrated a highly consistent pattern of results. Words that were read aloud were always remembered best, followed by words that were produced in other ways (written, mouthed, or whispered), followed by words that were read silently. It would seem that reading words aloud—despite the fact that it is certainly not the only type of production that benefits memory—is nevertheless the most effective. Although mouthing, writing, and whispering words made them more memorable than did silent reading, all three of these productions were inferior to reading aloud.[4] This pattern of results persisted despite a difference in word presentation times between Experiment 2A (5,000 ms) and Experiments 2B and 2C (2,000 ms).

Importantly, this pattern of results is consistent with distinctiveness theory. Mouthing, writing, and whispering all involve additional processing that stands out as distinct—and therefore memorable—relative to silent reading. Reading aloud, however, benefits from additional processing on two distinct dimensions relative to silent reading: both articulatory and auditory processing. Moreover, the auditory processing is also distinct relative to the other modes of production, with the exception of whispering. Whispering also entails auditory processing, although the acoustic signal is weaker than the signal when reading aloud, due to the lower volume of the response, which may lead to a weaker record of processing. We will consider this puzzle in more detail shortly, in the context of our general account of the production effect.

General discussion

The research reported here was designed to test the boundaries of the production effect. Two experiments demonstrated that a wide variety of productions improve explicit memory. In Experiment 1, three different types of productions—spelling, writing, and typing—benefited memory for words. In Experiment 2, we found that reading out loud (in a normal speaking voice) improved memory for words to a greater extent than did three other production modalities: writing, mouthing, and whispering. Nevertheless, these three types of production did still enhance memory relative to silent reading. These results highlighted a consistent empirical gradient: Words that were read aloud were remembered better than words produced by other unique responses, which were, in turn, remembered better than unproduced (i.e., silently read) words.

These experiments suggest that any mode of production that features unique, item-specific responses will result in a distinctive record in episodic memory that can then be retrieved and used during a memory test. Along these lines, Ozubko et al. (2012) showed that the advantage of reading aloud over reading silently occurred for words for which participants had a recollective experience or for words that were simply deemed to be familiar (regarding recollection, see also Gregg & Gardiner, 1991). The finding of a reliable production effect in recall (Conway & Gathercole, 1987, Exp. 3; Lin & MacLeod, in press) also fits with this observation. As well, Ozubko et al. (2012) and Conway and Gathercole (1987) found that participants were more likely to make correct than to make incorrect modality judgments for words that they read aloud than for words that they read silently. Just as individuals can draw upon retrieval of "I said that aloud" to help recognize a previously studied word, so, too, can they make use of retrieving "I mouthed that," "I wrote that," and so on.

Importantly, productions do not have to be vocal to benefit memory: Typing (Exp. 1C), writing (Exps. 1B and 2A), and mouthing (Exp. 2B) were each found to have a robust and consistent advantage (11 %–13 %) over silent reading. These results bolster MacLeod and colleagues' (2010) claim that the production effect is derived from unique, item-specific responses at study. Seemingly any

[4] Conway and Gathercole (1987) also found evidence of a spoken > mouthed > silent gradient, although their participants' memory was not reliably higher for mouthed than for silent items, nor was it reliably lower than memory for spoken items, likely due to the relatively low power resulting from their small sample size.

mixed-list design featuring "produced" words and "unproduced" (silently read) words will yield a production effect, regardless of the mode of the production—unless the distinctiveness of that production is undermined (see Ozubko & MacLeod, 2010). Indeed, these varied production effects persisted across experiments, despite procedural differences in study list length, presentation time, and old/new ratio on the recognition test. In short, any unique production provides a distinctive cue that participants can use at test to help remember studied words. In line with the proceduralist account, this distinct encoding activity is preserved in the original processing record (Kolers, 1973; Kolers & Roediger, 1984) and can subsequently be replayed to aid retrieval.

As expected, the present results were consistent with the distinctiveness account. Experiment 1 suggested that—like reading aloud—spelling, writing, and typing can also imbue words with distinctiveness (MacLeod et al., 2010). Each of these types of production involves distinct processing (motor, auditory, or both) relative to silent reading, and this distinct processing benefits memory. Words that are distinctively processed at study—regardless of whether they are spoken, mouthed, written, or produced in some other way—retain their distinctiveness at test, particularly against a backdrop of silently read recognition items. Conversely, words that are read silently at study will not be processed distinctively at test—in fact, they will "blend in" with the silently read distractors.

Experiment 2 further extended these boundaries by highlighting the relational nature of distinctiveness (Conway & Gathercole, 1987). Written and mouthed words both have motor distinctiveness relative to silently read words. Spoken words, however, have both motor *and* auditory distinctiveness relative to silently read words, as well as auditory distinctiveness relative to written and mouthed words. Thus, it appears that speaking resulted in a greater memory advantage than did writing or mouthing because speaking involves distinctive processing along an additional dimension. Spoken words benefit from distinct auditory processing that does not overlap with nonvocal productions or silent reading.

We believe that the results of Experiment 2 show a consistent auditory advantage over other means of production because the auditory processing involved in speech stands out as distinct relative to the processing in the other study conditions. However, in mixed lists featuring spoken, heard, and silently read words (MacLeod, 2011), articulation emerged as the distinct process that made spoken words the most memorable. These results suggest that auditory processing is not necessarily the distinct factor that benefits memory for spoken words; in other mixed lists, articulation stands out as distinct.

Somewhat at odds with this distinctiveness account, however, is our finding that words that were read aloud in a normal voice were remembered better than words that were whispered (Exp. 2C), even though reading aloud and whispering involve processing along the same dimensions. Interestingly, Murray (1965a, 1965b, 1966) also demonstrated a spoken > whispered > silent vocalization gradient, in his case in short-term serial recall. Although, like normal speaking, whispering words also entails auditory processing, having read words aloud is more memorable, possibly because normal speech involves the production of a stronger auditory signal than does whispering, which requires more "active" encoding. In keeping with the embodied-cognition perspective (Robbins & Aydede, 2009), MacLeod (2011) speculated that it may be the *action* of speaking that drives the production effect. Normal speaking may be a more memory-enhancing form of production than whispering because normal speaking involves more active—and therefore more distinct—processing. Another reason why speech may be more memory-enhancing than whispering is that speech allows for a greater variety of vocal responses. Whispered responses are less item-specific due to their spectral flatness. This harks back to MacLeod et al.'s (2010) finding that a production needs to be item-specific to be distinct. Along these lines, speaking may confer more distinctiveness than does whispering because spoken words are more varied, and hence more item-specific.

In sum, the present research has shown that a wide variety of item-specific productions can benefit explicit memory, including spelling, writing, typing, mouthing, and whispering. Yet although the boundaries of the production effect appear to be wide, reading aloud has consistently led to the best memory for words, holding an advantage over nonauditory (writing and mouthing), and even over another type of auditory (whispering), production. We argue that reading out loud facilitates memory more than these other productions because the act of speaking is more likely to be incorporated into the processing record, and hence is more distinct. Speaking involves processing along an extra dimension relative to many other productions, processing that is more active and more embodied. Although any item-specific production may facilitate memory, the effect of speaking is, in a nutshell, more pronounced.

Author note This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant A7459. We thank Jingjing Ji, Merrick Levene, and Nivedita Sharma for assistance in collecting the data. J.D.O. is now at the Rotman Research Institute, Baycrest Centre, Toronto, Ontario, Canada.

References

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 13, 553–563. doi:10.1037/0278-7393.13.4.553

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

Conway, M. A., & Gathercole, S. E. (1990). Writing and long-term memory: Evidence for a "translation" hypothesis. *Quarterly Journal of Experimental Psychology, 42A,* 513–527. doi:10.1080/14640749008401235

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. doi:10.3758/BF03196152

Engelkamp, J., & Krumnacker, H. (1980). Imaginale und motorische Prozesse beim Behalten verbalen Materials [Imagery and motor processes in memory of verbal material]. *Zeitschrift für experimentelle und angewandte Psychologie, 27,* 511–533.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica, 96,* 43–60. doi:10.1016/S0001-6918(97)00005-X

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

Gregg, V. H., & Gardiner, J. M. (1991). Components of conscious awareness in a long-term modality effect. *British Journal of Psychology, 82,* 153–162. doi:10.1111/j.2044-8295.1991.tb02390.x

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

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.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition, 1,* 347–355. doi:10.3758/BF03198119

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

Kurtz, K. H., & Hovland, C. I. (1953). The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *Journal of Experimental Psychology, 45,* 157–164. doi:10.1037/h0061022

Lin, O.Y.H., MacLeod, C.M. (in press). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology*

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

Masson, M. E., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General, 121,* 145–176. doi:10.1037/0096-3445.121.2.145

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

Murray, D. J. (1965a). The effect of white noise upon the recall of vocalized lists. *Canadian Journal of Psychology, 19,* 333–345. doi:10.1037/h0082913

Murray, D. J. (1965b). Vocalization-at-presentation and immediate recall, with varying presentation-rates. *Quarterly Journal of Experimental Psychology, 17,* 47–56. doi:10.1080/17470216508416407

Murray, D. J. (1966). Vocalization at presentation and the recall of lists varying in association value. *Journal of Verbal Learning and Verbal Behavior, 5,* 488–491. doi:10.1016/S0022-5371(66)80066-X

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., & 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

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 3–10). New York, NY: Cambridge University Press.

Rosenbaum, M. E. (1962). Effect of direct and vicarious verbalization on retention. *Child Development, 33,* 103–110.

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, Teachers College.