Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale

1196-1961/13/$12.00 DOI: 10.1037/cep0000009

The Offline Production Effect

Randall K. Jamieson and Jackie Spear
University of Manitoba

People remember words they say aloud better than ones they do not, a result called the production effect. The standing explanation for the production effect is that producing a word renders it distinctive in memory and thus memorable at test. Whereas it is now clear that motoric production benefits remembering over nonproduction, and that more intense motoric production benefits remembering to a greater extent than less intense motoric production, there has been no comparison of the memorial benefit conferred by motoric versus imagined production. One reason for the gap is that the standard production-by-vocalization procedure confounds the analysis. To make the comparison, we used a production-by-typing procedure and tested memory for words that people typed, imagined typing, and did not type. Whereas participants remembered the words that they typed and imagined typing better than words that they did not, they remembered the words they typed better than the ones they imagined typing; an advantage that was consistent over tests of recognition memory and source discrimination. We conclude that motoric production is a sufficient and facilitative (but not a necessary) condition to observe the production effect. We explain our results by a sensory feedback account of the production effect and sketch a computational framework to implement that approach.

Keywords: production effect, distinctiveness, recognition, source monitoring, typing

People remember words that they have read aloud better than words that they have not. Whereas the memorial advantage for spoken over unspoken words has been known for some time (Hopkins & Edwards, 1972; see also Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988), it has recently been redubbed and reexamined as *the production effect* (Bodner & Taikh, 2012; Forrin, MacLeod, & Ozubko, 2012; Hourihan & MacLeod, 2008; Lin & MacLeod, 2012; MacLeod, 2010, 2011; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; Ozubko, Gopie, & MacLeod, 2012; Ozubko & MacLeod, 2010).

Several features of the production effect are now established. It is stronger and more reliable in within- than between-subjects designs (Fawcett, 2013; Taikh & Bodner, 2013). It is stronger in experiments with mixed- compared with pure-list designs (MacLeod et al., 2010; but also see Forrin, Groot, & MacLeod, 2013). It is observed if participants issue a unique production for each word but not if the same production is issued to all produced words (MacLeod et al., 2010). It does not depend on the availability of semantic information (MacLeod et al., 2010). It is stronger for words that a learner produces than for words produced by a partner (MacLeod, 2011). It is stronger for words spoken aloud than for words that are whispered (Forrin et al., 2012). It is observed in recognition, recall, and source monitoring (Bodner & Taikh, 2012; MacLeod, 2010; Ozubko & MacLeod, 2010).

The standard explanation for the production effect is that production renders an item distinctive in memory, which aids remembering at test. The explanation fits within a proceduralist view of memory (Kolers, 1973), in which the recollection of having produced a word serves as a diagnostic cue that it was in the studied list: "I can recall saying this word aloud, so it must be old" (Hourihan & MacLeod, 2008; MacLeod et al., 2010). It also fits with the distinctiveness-based explanations offered for related results, such as the enactment effect, in which people remember a written instruction better if they enact or imagine to enact it (e.g., Engelkamp, 1995; Engelkamp & Dehn, 2000; Engelkamp, Zimmer, Mohr, & Sellen, 1994; Peterson & Mulligan, 2010), and the generation effect, in which people remember a word better if they generate it than if they read it (e.g., Begg, Vinski, Frankovich, & Holgate, 1991; Johns & Swanson, 1988; Slamecka & Graf, 1978).

Whereas it is now clear that motoric (i.e., online) production can benefit remembering (see Bodner & Taikh, 2012), there are no data asking if imagined (i.e., offline) production confers the same benefit. The omission is notable because the contrast between online and offline production can contribute novel insights into the core phenomenon. If imagined production does not benefit remembering, we can conclude that motoric production is a necessary condition of the production effect. If imagined production does benefit remembering, we can conclude that motoric production is a sufficient, but not a necessary, condition for observing the production effect. If both motoric and imagined production benefit remembering, any difference (or lack thereof) in the magnitude of benefit can be used to estimate the value of motoric over imagined production. Nevertheless, the comparison has not been made.

We surmise that no one has published a comparison of online versus offline production because the standard production-by-vocalization prevents it. To explain, consider the standard task in which production is accomplished by reading a word aloud and nonproduction is accomplished by reading a word silently. Whereas those experimental manipulations would seem to distin-

Randall K. Jamieson and Jackie Spear, Department of Psychology, University of Manitoba.

The research was supported by a Discovery Grant from the National Sciences and Engineering Research Council of Canada to Randall K. Jamieson.

Correspondence concerning this article should be addressed to Randall K. Jamieson, P404 Duff Roblin Building, Department of Psychology, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2. E-mail: randy.jamieson@umanitoba.ca

guish production from nonproduction, they do not. First, silent reading compels subvocal articulation: subtle movements in the larynx and other muscles involved in the articulation of speech (e.g., Haber & Haber, 1982; McGuigan, 1970; Tanenhaus, Flanigan, & Seidenberg, 1980). Consequently, silent reading (i.e., nonproduction) compels motoric production that is consistent with the same motoric production involved in speaking aloud (i.e., production). Second, silent reading compels inner speech and, consequently, is better construed as imagined speech than nonproduction (e.g., Haber & Haber, 1982; Oppenheim & Dell, 2010). Finally, verbal recoding (i.e., translation from the orthographic to phonologic code) is so fast and compulsive that it is unreasonable to expect that participants can suppress inner speech and subvocal articulation to force a wishful experimental distinction between production and nonproduction (MacLeod, 1991; Rubenstein, Lewis, & Rubenstein, 1971).

Whereas the overlap in production and nonproduction is a problem in general, it is particularly problematic for addressing the distinction between online and offline production. First, because both production and nonproduction invoke a shared motor program, production-by-vocalization cannot be used to test if motoric production is critical to the production effect. Second, because reading compels inner (i.e., imagined) speech, it is not clear how one might design an experimental manipulation so that imagined production differs meaningfully from nonproduction. To solve the problems, a different experimental method is needed, one that can be used to assess the offline production effect and enables an experimental comparison of the memorial benefits derived from online versus offline production.

In the work that follows, we compare the memorial benefit of online production, offline production, and nonproduction using a variation on Forrin et al.'s (2012) production-by-typing protocol (see also Spear & Jamieson, 2011). We use the production- by-typing task because it solves the problems associated with the production-by-vocalization procedure. First, production-by-typing divorces reading from production, allowing us to better differentiate nonproduction (i.e., silent reading and no typing) from both online (i.e., silent reading and motoric typing) and offline (i.e., silent reading and imagined typing) production. Second, whereas reading compels subvocal articulation and inner speech, reading does not compel typing or imagined typing. Thus, we can distinguish production and imagined production from nonproduction and resolve the problems associated with relying on participants' skill at suppressing their reflexive language and motor processes during study. In summary, the production-by-typing task affords a degree of experimental control that the production-by-vocalization task does not. In the experiments described here, we exploit the added experimental control to answer questions about a potential offline production effect and its magnitude relative to its online counterpart.

We will adopt the following strategy in the remainder of the paper. In Experiment 1, we examine the online and offline production effects in recognition memory. In Experiment 2, we verify our interpretation of results from the recognition test in Experiment 1. In Experiment 3, we use a source discrimination test to develop converging evidence for our conclusions from the recognition memory results. Although we will use a production-by-typing task, we will design our experiments to be consistent with previous work.

Experiment 1

Experiment 1 was conducted to examine the benefit of imagined production on recognition. In the study phase, participants will type some words, imagine typing other words, and read the remainder. Afterward, they will be tested for recognition of the words they studied.

We predict that participants will recognise the words that they typed better than the words they read: an online production effect. We also predict that participants will remember the words that they imagined typing better than the words they read: an offline production effect. We will compare recognition for words that participants typed and imagined typing to assess the relative benefit of online versus offline production. If participants recognise words that they imagined typing better than words they read, we will conclude that imagined production is a sufficient condition for observing the production effect and that motoric production is not critical. If, in addition, participants recognise words that they typed better than words they imagined typing, we will conclude that motoric production benefits recognition over and above the benefit derived from imagined production.

Method

Participants. Sixty undergraduate students from the University of Manitoba undergraduate participant pool took part in the study. All participants reported normal or corrected-to-normal vision, and all participants reported that they were capable touch typists.

Apparatus. The experiment was administered on personal computers, each equipped with a 21.5-in. wide-screen monitor, a standard keyboard, and a standard mouse.

Materials. Each participant was presented with a random sample of 90 words taken from the Toronto Word Pool, a database that includes over 1,000 words falling within regular ranges for print frequency, imageability, and concreteness (Friendly, Franklin, Hoffman, & Rubin, 1982). We excluded the word IMAGE from the word pool because it had potential to be confused with the "imagine" instruction in the experimental procedure.

Procedure. Three to seven participants were tested at a time in the same room, each at a different computer terminal.

Prior to the experiment, participants were told that they would be shown 45 words, each of which would be followed by an instruction to type the word, to imagine typing the word, or to do nothing. Participants were told to keep their fingers on the home row during the entire study phase and that they should not move their fingers from home row over the study phase, including when they imagined typing a word. The experimenter was present during the experimental sessions to monitor participants' compliance with the experimental instructions.

The participant initiated the training phase by clicking on the message "Start," at which point the computer screen was cleared and an instruction "Press the spacebar" was presented at the vertical and horizontal midpoint of the screen. When the participant pressed the spacebar, the screen was cleared; 1,000 ms later, a study word was presented in all-caps and in 40-point Arial font. The word remained on the screen for 2,000 ms, after which the screen was cleared; 100 ms later, one of three instruction words was presented— "Type," "Imagine," or "Nothing." The instruction word was presented at the same location on the computer screen where the study word had been presented.

On type trials, the participant typed the word on the computer keyboard and then pressed the return key (the word “Type” remained on the screen until the participant pressed the return key, at which point the screen was cleared); keystrokes were not echoed to the screen and no feedback was provided about typing accuracy. On imagine trials, the participant imagined typing the word on the computer keyboard and then pressed the return key (the word “Imagine” remained on the screen until the participant pressed the return key, at which point the screen was cleared). On nothing trials, the subject waited 2,000 ms (the word “Nothing” remained on the screen for 2,000 ms, at which point the screen was cleared).

Participants studied 45 words in total: 15 of which they typed, 15 of which they imagined typing, and 15 of which they only read. The order of the 15 type, 15 imagine, and 15 read trials was randomized for each participant.

Following the training phase, instructions for the test phase were presented on the screen. Participants were told that they would be presented with 90 words, the 45 words they studied plus 45 words they had not, and that it was their task to discriminate the two kinds. A button marked “Begin” was presented below the instructions. When the participant clicked on the button, the test phase began.

On each recognition trial, a word was presented at the centre of the screen in all-caps and in 40-point Arial font. Two tick boxes were displayed beneath the presented word. One tick box was labelled “Old” and the other labelled “New.” A button marked “OK” was displayed beneath the tick boxes. Participants recorded their recognition decision for a word by clicking one of the tick boxes, and then clicking the “OK” button. When the participant clicked on the “OK” button, the screen was cleared; 1,000 ms later, the next test word was presented. This cycle continued until all of the 90 test words had been presented and the participant had provided a response for each one.

Results and Discussion

Figure 1 shows the mean percentages of “Old” responses that participants issued to the four classes of test items. Means for the typed, imagined, and read words are hit rates; the mean for foils is a false alarm rate. Whiskers indicate one standard error of the mean.

We analysed the data in a single-factor repeated-measures ANOVA, with probe type as the repeated factor. The omnibus test was significant, $F(3, 177) = 321.24$, $MSE = 153.40$, $p < .0001$, indicating at least one statistically significant difference amongst the four means. We pulled the differences apart in three planned orthogonal contrasts. First, we compared the mean percentage of old responses for studied words (i.e., typed, imagined, and read) against the mean percentage of old responses for unstudied words (i.e., foils). The comparison was statistically significant and confirmed that participants discriminated studied from unstudied words, $F(1, 59) = 775.59$, $MSE = 187.15$, $p < .0001$. Second, we compared the mean percentage of old responses for the words that participants produced (i.e., typed and imagined) to the mean percentage of old responses for words that they read but did not produce (i.e., read). The comparison confirmed that participants recognised the words that they produced at study better than the ones they read, $F(1, 59) = 14.39$, $MSE = 132.00$, $p = .0004$: an omnibus production effect. We also assessed the online and offline

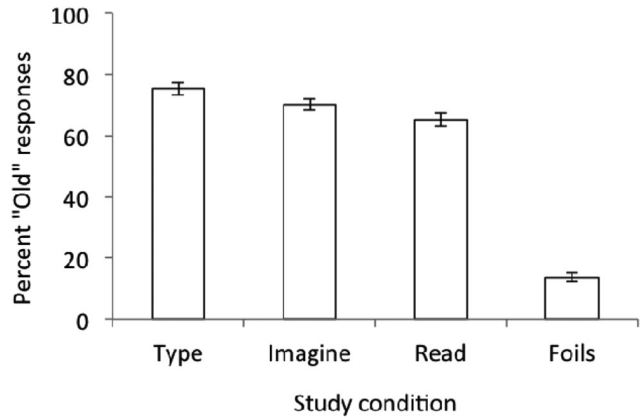

Figure 1. Percentages of old responses given to items that participants (a) read and typed at study, (b) read and imagined typing at study, (c) read at study, and (d) did not study.

production effects separately by comparing recognition for the words that participants typed against recognition for the words they read, $t(59) = 4.53$, $p < .0001$, and recognition for the words that participants imagined typing to recognition for the words they read, $t(59) = 2.03$, $p = 0.471$. Both comparisons were statistically significant and provided confirmation of independent online and offline production effects. Finally, we compared the mean percentage of old responses for the words that participants typed at study versus the mean percentage of old responses for the words that they imagined typing at study. The comparison was statistically significant, confirming that participants recognised the words that they typed better than the words they imagined typing, $F(1, 59) = 5.57$, $MSE = 141.07$, $p = .0216$: evidence that online production benefitted remembering more than offline production.[1]

Finally, we examined production times to evaluate the possibility of a confound between recognition performance and the time that participants took to type and imagine typing words in the study phase. The analysis provided no evidence of contamination. Participants took approximately the same amount of time to type ($M = 1,888$ ms, $SEM = 73$ ms) and imagine typing ($M = 1,947$ ms, $SEM = 101$ ms) words in the study phase, $t(59) = 0.62$, $p = .5393$.

In summary, we observed a memorial benefit of both online and offline production. However, online production benefitted recognition more than offline production. We conclude that both online and offline production benefit remembering, but online production benefits remembering more than offline production.

Although our data are clear, our interpretation of them is clouded (Bodner & Taikh, 2012). To explain, consider our data from Experiment 1. Participants remembered the words they produced better than the words they did not, and we concluded that production benefitted remembering. However, the difference can be interpreted two other ways. Perhaps the data reflect a cost of nonproduction: a reversed production effect (Bodner & Taikh, 2012). Or, perhaps the data reflect a mixture of results: a benefit of production and a cost of nonproduction.

[1] We do not report a signal detection analysis of the data because statistics for typed, imagined, and read hit rates would need to be derived from the same false alarm rate (Green & Swets, 1966).

Fortunately, we can resolve the ambiguity empirically. In Experiment 2, we test participants recognition of words, all of which they silently read at study. We will interpret the hit and false-alarm rate from that procedure as a benchmark for recognition performance in Experiment 1. If the hit rate for produced words from Experiment 1 is greater than the hit rate for read words in Experiment 2, we will conclude that production benefitted performance in Experiment 1. If the hit rate for read words in Experiment 1 is less than the hit rate for read words in Experiment 2, we will conclude that nonproduction impaired performance in Experiment 1. If both, we will conclude that the production effect observed in Experiment 1 reflects a combination of production benefits and nonproduction costs.

Experiment 2

Experiment 2 was conducted to disambiguate the interpretation of data from Experiment 1. Participants were tested for recognition of 45 words, all of which they silently read at study. We expect participants will recognise the words they read as well as participants did in Experiment 1: approximate hit and false alarm rates of 65% and 15%, respectively. If so, we will conclude that both typing and imagined typing benefitted recognition in Experiment 1 relative to reading.

Method

Participants. Sixty undergraduate students from the University of Manitoba undergraduate participant pool took part in the study. All participants reported normal or corrected-to-normal vision.

Apparatus and materials. The experimental apparatus and word pool used were the same as in Experiment 1.

Procedure. The experimental protocol was nearly identical to that of Experiment 1. The only difference was that participants read all of the study words (i.e., they were not instructed to produce any of the words).

Results and Discussion

The data were consistent with our predictions. Participants recognised 64.0% (SEM = 2.1%) of the words they studied and false alarmed to 16.3% (SEM = 1.5%) of the words they did not. A dependent samples t test confirmed a difference between the two means, $t(59) = 17.46$, $p < .0001$: participants discriminated studied from unstudied test words. A signal detection analysis corroborated the conclusions. Mean discrimination was better than chance ($d' = 1.51$, $SEM = 0.10$), $t(59) = 15.28$, $p < .0001$; participants exhibited a conservative decision bias ($C = 0.35$, $SEM = 0.04$), $t(59) = 8.25$, $p < .0001$.

Next, we compared the results of the present experiment against the results from Experiment 1 (see Figure 1). The hit rates for read words, $t(118) = 0.57$, $p = .5685$, and the false alarm rates for unstudied words, $t(118) = 1.34$, $p = .1840$, were statistically equal in the two experiments. We conclude that performance for read words was equivalent in the two experiments.

Next, we compared performance for the words that people read in the present experiment against performance for the words that participants typed and imagined typing in Experiment 1. Participants in Experiment 1 recognised words that they typed, $t(118) = 3.86$, $p = .0002$, and imagined typing, $t(118) = 2.12$, $p = .0359$, better than participants in Experiment 2 recognised words that they read. On the basis of this comparison, we conclude that memory performance of participants in Experiment 1 benefitted from both typing and imagined typing relative to control. The data reinforce our interpretation of results from Experiment 1: Both online and offline production benefit remembering, but online production benefits remembering more than offline production.

A compelling aspect of the production effect is that it emerges in an array of memory measurements. For example, Ozubko and MacLeod (2010) showed that production benefits performance in source monitoring as well as recognition (see also Bodner & Taikh, 2012). In their experiment, participants studied two successive word lists. For the first list, participants read half of the words silently and half of the words aloud. For the second list, participants read all of the words silently (or all of the words aloud). Following study, participants were asked to identify which of the two lists each studied word was from (i.e., list one or list two). When participants read some words aloud and others silently in List 1 and then read all of the words silently in List 2, discrimination of items in list one was better for the spoken than for the silently read words: a production effect. However, when participants read half of the words aloud and half silently in list one and then read all words aloud in list two, the production effect was eliminated. Ozubko and MacLeod explained the difference as a corollary of distinctiveness in memory. Reading all of the words silently in list two makes memory of having produced a word at study diagnostic of its list one membership. In contrast, reading all of the words aloud in list two renders memory of having produced a word aloud in list one ambiguous (see Bodner & Taikh, 2012, for a different explanation).

In Experiment 3, we build upon Ozubko and MacLeod's (2010) examination of the production effect in source monitoring using the production-by-typing protocol. The experiment asks if the memorial advantage for motoric over imagined production observed in recognition generalizes when memory for produced and unproduced words is tested by source discrimination.

Experiment 3

Participants were assigned to one of two treatment conditions. Participants in a *type/read* condition typed some words and read the remainder. Participants in an *imagine/read* condition imagined typing some words and read the remainder. Following the study phase, participants in both conditions were presented with each of the words they studied and were invited to remember how they processed it at study: "Did you produce or did you read this word at study?" If we replicate the advantage for online over offline production from Experiment 1, we will observe better discrimination of produced from read words in the type/read than in the imagine/read condition.

Method

Participants. One hundred thirty-one undergraduate students from the University of Manitoba undergraduate participant pool took part in the study. Seventy of the students were assigned to the type/read condition and 61 were assigned to the imagine/read condition: The difference of nine subjects in the two conditions reflects the outcome of random assignment to conditions. All participants reported normal or corrected-to-normal vision and all participants reported being capable touch typists.

Apparatus and materials. The experimental apparatus and word pool used were the same as in Experiment 1. However, we sampled 40, rather than 90, words from the Toronto word pool for each subject and presented all 40 words to that subject in both the study and test phases (i.e., there were no unstudied foils).

Procedure. The experimental protocol was very similar to that of Experiment 1, except for three critical changes. First, participants studied 40, rather than 45, words, half of which they produced (i.e., typed or imagined typing depending on the assigned condition) and half of which they did not (i.e., read). Second, memory was evaluated at test by asking participants to indicate whether they did or did not produce the word at study (i.e., instead of asking if they recognised the word). Third, tick boxes in the test phase were labelled "Typed" and "Read" in the type/read condition, and labelled "Imagined" and "Read" in the imagine/read condition.

Results and Discussion

Figure 2 shows the percentages of "Produced" responses that participants issued to words that they did (i.e., typed or imagined typing) and did not (i.e., read) produce at study. Thus, filled bars in Figure 2 indicate correct decision rates (i.e., hits) and open bars indicate incorrect decision rates (i.e., false alarms). Whiskers indicate one standard error of the mean.

The means in Figure 2 offer an objective picture of participants' judgments. However, it would be misleading to directly compare means in the two conditions because they might contain an unknown and uneven mixture of response bias. To unconfound discrimination from bias, we analysed performance using the d' statistic from signal detection theory.

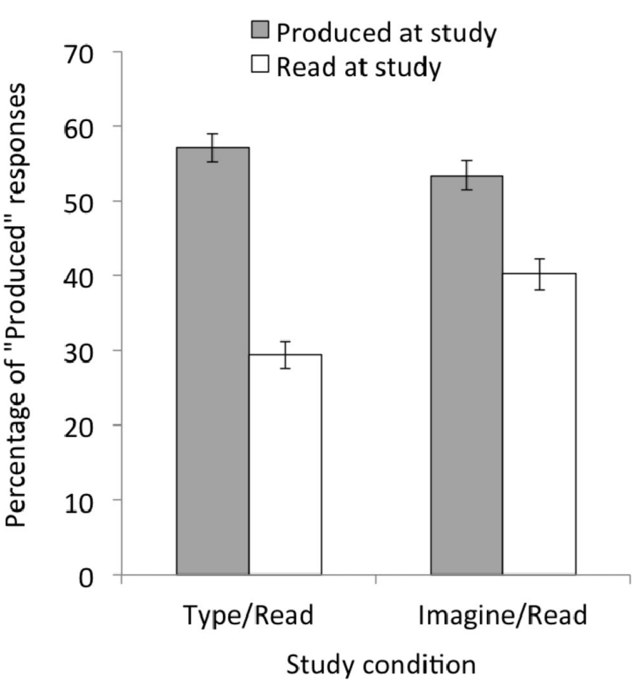

Figure 2. Percentages of produced responses to items that were produced at study (i.e., typed or imagined) and items that were not produced at study (i.e., read).

As shown in Figure 2, participants in both the type/read ($d' = 0.81$, $SEM = 0.08$), $t(69) = 10.51$, $p < .0001$, and imagine/read ($d' = 0.38$, $SEM = 0.07$), $t(60) = 5.27$, $p < .0001$, conditions discriminated produced from unproduced words and did so statistically better than chance (i.e., $d' = 0$). We used an independent samples t test to compare d' scores between the two conditions: The test confirmed that participants in the type/read condition discriminated produced from unproduced words better than participants in the imagine/read condition, $t(129) = 4.04$, $p < .0001$. We also analysed response bias in the two experimental conditions using the C statistic from signal detection theory. Participants in both the type/read ($C = 0.20$, $SEM = 0.04$), $t(69) = 4.77$, $p < .0001$, and imagine/read ($C = 0.09$, $SEM = 0.05$), $t(60) = 1.95$, $p = .0564$, conditions responded "read" more often than "typed" or "imagined" (i.e., a conservative bias consistent with data from Experiment 1). An independent samples t test failed to provide evidence of a statistical difference in the magnitude of bias in the two conditions, $t(129) = 1.69$, $p = .0935$.

Finally, we examined production times to determine if accuracy was correlated with the amount of time that participants took to produce words. The analysis provided no evidence of contamination: Participants took approximately the same amount of time to type ($M = 1{,}560$ ms, $SEM = 65$ ms) and imagine typing ($M = 1{,}753$ ms, $SEM = 103$ ms) words at study, $t(129) = 1.62$, $p = .1072$.

In summary, participants in both the online and offline production conditions discriminated the words that they produced from the words that they read. However, participants in the online production condition discriminated produced from unproduced words better than those in the offline condition. The results corroborate those from Experiment 1: Both online and offline production benefit remembering, but online production benefits remembering more than offline production.

General Discussion

People remember words they say aloud better than ones they do not, a result called the production effect (MacLeod et al., 2010). We have reported an experimental analysis of the production effect using Forrin et al.'s (2012) production-by-typing procedure to ask two novel and unexamined questions: Do people remember words that they imagine producing better than words they do not produce (i.e., the offline production effect), and, if so, does the memorial benefit gained from imagined production differ in magnitude to the memorial benefit gained from motoric production? Our data provide clear answers to both questions. First, participants remembered words that they imagined typing better than words they read: an offline production effect. Second, participants remembered the words that they typed better than the words that they imagined typing: Online production benefits remembering more than offline production.

The dominant explanation for the production effect is that production renders a word distinctive in memory and that the added distinctiveness benefits performance at test. Working backward from that account, we might conclude that both online and offline production render a word distinctive in memory but that online production renders it more so. But what is distinctiveness and how does it benefit remembering?

One answer is that distinctiveness is related to the vividness of imagined memories. If imagining to type a word leads to a perceptually impoverished and vague representation lacking in details, then one might anticipate better memory for typed words. We think that this basic notion is an important one for understanding the production effect and why the online production effect is stronger than the offline production effect.

In a recent article, Forrin et al. (2012) argued that producing a word yields sensory feedback (e.g., auditory feedback from speech, motoric feedback from mouthing), some of which is integrated into memory for the studied word. They also argued that the greater the number of senses involved in production, the more distinctive the word is rendered in memory, and thus the easier it is to remember at test:

> 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. (p. 1054)

We think Forrin et al.'s (2012) explanation explains our data (as well as their own). However, their explanation could benefit from a formalization that specifies (a) how experimental manipulations give rise to memorial distinctiveness, and (b) how memorial distinctiveness is leveraged to benefit performance at test.

Most formal models of memory imagine memory as a hyperdimensional geometric space. Each studied item is plotted as a coordinate in the space. Distinctive items (i.e., items that differ from other items in memory) are located in sparse or poorly populated regions.

Indistinctive items (i.e., items that do not differ much from other items in memory) are located in dense or well-populated regions. When a probe is presented to memory, it retrieves items near it. Thus, a distinctive probe (i.e., one located in a sparse region of memory) can retrieve itself from memory without retrieving many near neighbours (i.e., because there are few). An indistinctive probe, on the other hand, cannot help but retrieve its near neighbours (i.e., because it has many). By a difference of interference at retrieval, distinctive items are retrieved more clearly than indistinctive items—a factor that benefits recognition. This scheme is well described in Brown, Neath, and Chater's (2007) SIMPLE model of memory and in the models it was developed from (Nairne, 1988, 1990).

But how does a word come to be encoded to a sparse region of memory? Put differently, how does an item achieve distinctiveness? One solution is to treat the stimulus as distinctive. But distinctiveness is not a stimulus property. Rather, distinctiveness is gained through processing. For example, if an item mismatches memory, it achieves distinctiveness by virtue of that mismatch. We note, however, that distinctiveness based on an item's difference to memory of other studied items would not predict our results.

Nairne's (1988, 1990) feature model provides a second and, we think, directly relevant definition of distinctiveness for our experiments and for the production effect in general. The feature model represents memory for a stimulus as a vector (i.e., a coordinate in a memory space). Each vector dimension (i.e., axis in the space or feature of the stimulus) represents an item's value on a particular stimulus dimension. Some vector dimensions represent modality-independent features (i.e., those aspects of a situation that remain the same regardless of presentation modality or production) and others represent modality-dependent features (i.e., those aspects of a situation that are determined by the modality of presentation or production).

If one assumes that a word is represented in memory by the same modality-independent features, whether it was produced or not, that producing a word enriches the representation on modality dependent dimensions, and that producing a word motorically gives it more extreme values on the modality dependent dimensions, the results of our experiments fall out. Critically, so do the results of both Forrin et al. (2012), who showed that participants recognised words they spoke better than words they whispered better than words they read, and the results of Quinlan and Taylor (2013), who reported that participants recognised words they sang better than words they yelled better than words they spoke better than words they read.

Whereas our explanation is consistent with the spirit of Forrin et al.'s (2012) assertion that memory performance ought to improve the greater the number of sense modalities incorporated in production, it adds some constraint by providing a clear and principled translation from the sensory feedback gained in production to the mathematical and formal definition of distinctiveness implemented in current computational theories of memory (e.g., Brown et al., 2007; Nairne, 1988, 1990; Nosofsky & Zaki, 2003). In that sense, the explanation we provide preserves the distinctiveness account of the production effect, but identifies sensory feedback as the condition by which distinctiveness is achieved. An aim of ongoing work is to formulate the sketch we have provided in a working computational model and, if successful, use the model to examine the relationships between the production effect and its kissing cousins: the enactment and generation effects.

Nairne's (1988, 1990) appeal to a distinction between modality-dependent and modality-independent features presents a second perspective on our results. Nairne argues that an unusual method of word production—for example, silent mouthing or imagined typing—can lead to a *protracted encoding* of the stimulus. In fact, Nairne (1988, p. 350) has discussed the distinction between reading, mouthing, and whispering in exactly this manner to address both the modality and suffix effects in recall.

If one accepts the protracted encoding hypothesis, it is possible that the online and offline production effects occur for different reasons: the online production effect because of sensory distinctiveness, and the offline production because of protracted encoding. Unfortunately, the explanation points to complications of analysis that cannot be sorted by the data reported here or elsewhere.

An empirical shortcoming of our analysis is that we cannot verify that participants followed instructions to silently read and imagine typing words. Leveraging that weakness, a skeptic could argue that participants imagined typing some, but not all, words at study. If so, the intermediate hit rate for the words that people imagined typing might be due to the fact that they imagined typing sometimes but typed words always.

The criticism is valid and inescapable. It is also one that all examinations of the production effect must handle. All studies of the production effect depend on a comparison of memory for words that participants produced by an observable and thus verifiable act (e.g., speech, typing) and words they processed by an unobservable act (e.g., silent reading). Thus, the criticism that participants might not have imagined producing words when instructed is also a criticism for words they did not produce (e.g., by silent reading).

Of course, the problem has not gone unnoticed and has more recently been dealt with by running control conditions like the one we reported in Experiment 2. Unfortunately, the necessary control conditions are often missing in empirical reports, and so the problem and validity of the criticism is more severe than one might like to acknowledge (see Bodner & Taikh, 2012, for the original statement and empirical solution to the criticism).

The fact that a majority of studies on the production effect suffer the unverified nonproduction problem does not justify a dismissal of it. So what, then, does it mean for our results and conclusions? In short, we stand firm on the point that offline production benefits remembering. Our reasoning is twofold. First, the comparison of results in Experiment 1 against the control condition in Experiment 2 confirms a benefit of both online and offline production (without a nonproduction cost). Second, if the criticism is valid and participants did not always imagine typing when instructed, our measurements would underestimate, rather than overestimate, the magnitude of the offline production effect.

The issue has important theoretical implications. If the offline production effect were in fact as strong as its online counterpart, we would need to question if the production effect is meaningful at all. What would be the added value of production if the online production confers the same magnitude advantage as mental rehearsal?[2]

Several researchers have reported experiments using a similar experimental design to ours. For example, Forrin et al. (2012) reported that participants remembered words they spoke better than words they whispered better than words they mouthed better than words they read. In a similar design, Quinlan and Taylor (2013) reported that participants remembered words they sang better than words they yelled better than words they spoke better than words they read. Those designs are, in turn, consistent with the designs of other experiments such as Murray, Birch, Chase, Eyolfson, and Simms's (1991) examination of recognition for items encoded by silent reading, silent mouthing, whispering, and vocal reading (see also Nairne, 1988).

Despite similarities between our comparisons and ones reported elsewhere, our experiments are different and ask unique and meaningful questions. First, our experiments are the only ones that compare people's memory for words they produced motorically and words they produced independent of motoric output. The difference is clear if one considers that participants in all previous experiments engaged in motoric production over all study conditions: singing, yelling, speaking, whispering, silent mouthing, and even silent reading (i.e., due to subvocal articulation). Thus, our experiments are the first to compare performance in motoric versus nonmotoric production. Second, whereas muted production (e.g., silent mouthing, whispering) could be construed as imagined production, muted production is still confounded with motoric production. Thus, our experiments are the first to compare motoric versus imagined production independent of that confound. Third, because reading compels inner speech, and imagined vocal production amounts to inner speech, it is difficult and perhaps impossible to examine the distinction between imagined production and nonproduction using a production-by-vocalization procedure. In summary, previous examinations of the production effect have confirmed that the benefit of online production varies with the intensity of production, but they have not examined or provided evidence for our two key conclusions: Offline production benefits remembering and online production benefits remembering more than offline production.

Finally, the pragmatic value of our contribution is made clearer if one asks how a production-by-vocalization experiment can be designed to address the offline production effect. For example, one might test recognition for words that people spoke, imagined speaking, and silently read at study. However, and despite a seemingly valid set of comparisons, the design cannot answer the questions that we have answered. Because all three conditions involve motor production, there is no way to test if an offline production effect is observed independent of motoric production and, consequently, no way to compare the magnitudes of the online and offline cases. We do not mean to argue that Forrin et al.'s (2012) production-by-typing protocol is the only, or even the best, experimental method to examine the distinction between online and offline production. Rather, we mean to argue that production-by-vocalization is a problematic method for doing so and that using the production-by-typing task goes some initial distance to solving those problems.

The production-by-typing procedure supports a distinction between online and offline production that the production-by-vocalization procedure cannot. We have used the procedure to reevaluate assumptions and conclusions about the production effect and to make three general contributions to a delineation and understanding of it. First, the production effect is observed for both motoric and imagined production. Thus, motoric production is a sufficient but not a necessary condition of the production effect. Second, motoric production confers a larger memorial benefit than imagined production. Thus, motoric production is a facilitative but not a necessary condition for the production effect. Third, the memorial benefit of motoric over imagined production is consistent over tests of both recognition and source memory. Thus, online production appears to facilitate remembering more than offline production, independent of how memory was tested.

[2] In fact, our initial intuition when we began conducting this work is that imagined typing would be a sufficiently bizarre method of production that participants would remember the words that they imagined typing better than the ones they actually typed—a prediction in line with Nairne's (1988) notion of protracted encoding.

Résumé

Les personnes se souviennent mieux des mots qu'ils disent tout haut, un résultat appelé « effet production ». L'explication actuelle de cet effet est que la production d'un mot le rend distinct dans la mémoire et, de ce fait, le sujet peut s'en souvenir dans le cadre d'un test. Les avantages de la production motrice sont maintenant clairs, comparativement à une non-production, pour la mémorisation, et ces avantages sont plus grands lorsque s'accentue l'intensité de la production motrice. Toutefois, il n'y a eu aucune comparaison des avantages découlant de la production motrice comparativement à ceux de la production imaginée. Une raison pouvant expliquer l'écart est que la procédure standard de production verbale perturbe l'analyse. Pour effectuer la comparaison, nous avons eu recours à une procédure de production par dactylographie : nous avons vérifié la mémorisation des mots que les sujets ont tapé, imaginé avoir tapé et ceux qu'ils n'ont pas tapés. Les participants se sont mieux rappelés les mots qu'ils avaient tapés et ceux qu'ils avaient imaginé taper comparativement aux mots qu'ils n'avaient pas tapés ni imaginé taper. En outre, ils se sont mieux rappelés des mots qu'ils avaient tapés que ceux qu'ils avaient imaginés taper. Cet avantage s'est révélé cohérent dans les tests de reconnaissance mnémonique et de reconnaissance des sources. Nous concluons que la production motrice est une condition suffisante et facilitante (mais non nécessaire) pour l'observation de l'effet production. Nous expliquons nos résultats par une rétroaction sensorielle de l'effet production et esquissons un cadre computationnel pour mettre en application cette approche.

Mots-clés : effet production, caractère distinctif, reconnaissance, contrôle des sources, dactylographie

References

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19,* 487–497. doi:10.3758/BF03199571

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1711–1719. doi:10.1037/a0028466

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. doi:10.1037/0033-295X.114.3.539

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

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. (1995). Visual imagery and enactment of actions in memory. *British Journal of Psychology (London, England: 1953), 86,* 227–240. doi:10.1111/j.2044-8295.1995.tb02558.x

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., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22,* 34–39. doi:10.3758/BF03202759

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

Forrin, N. D., Groot, B., & MacLeod, C. (2013, June). *The benefits and costs of production*. Paper presented at the 23rd Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Alberta, Canada.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40,* 1046–1055. doi:10.3758/s13421-012-0210-8

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto word pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1080 words. *Behavior Research Methods & Instrumentation, 14,* 375–399. doi:10.3758/BF03203275

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Haber, L. R., & Haber, R. N. (1982). Does silent reading involve articulation? Evidence from tongue twisters. *The American Journal of Psychology, 95,* 409–419. doi:10.2307/1422133

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, 62,* 242–246. doi:10.1037/1196-1961.62.4.242

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 180–190. doi:10.1037/0278-7393.14.1.180

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

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

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. doi:10.1037/0033-2909.109.2.163

MacLeod, C. M. (2010). When learning met memory. *Canadian Journal of Experimental Psychology, 64,* 227–240. doi:10.1037/a0021699

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

McGuigan, F. J. (1970). Covert oral behavior during the silent performance of language tasks. *Psychological Bulletin, 74,* 309–326. doi:10.1037/h0030082

Murray, D. J., Birch, D. E., Chase, R. S., Eyolfson, B. H., & Simms, K. A. M. (1991). The effects of vocalization at presentation on short-term recognition of three-digit numbers. *Canadian Journal of Psychology, 45,* 462–473. doi:10.1037/h0084304

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16,* 343–352. doi:10.3758/BF03197045

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269. doi:10.3758/BF03213879

Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1194–1209. doi:10.1037/0278-7393.29.6.1194

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition, 38,* 1147–1160. doi:10.3758/MC.38.8.1147

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

Peterson, D. J., & Mulligan, N. W. (2010). Enactment and retrieval. *Memory & Cognition, 38,* 233–243. doi:10.3758/MC.38.2.233

Quinlan, C., & Taylor, T. (2013, June). *Enhancing the production effect in memory*. Paper presented at the 23rd Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Alberta, Canada.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning & Verbal Behavior, 10,* 645–657. doi:10.1016/S0022-5371(71)80071-3

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

Spear, J. A., & Jamieson, R. K. (2011). A reversed production effect in reality monitoring. *Canadian Journal of Experimental Psychology, 65,* 298.

Taikh, A., & Bodner, G. E. (2013, June). *Evaluating the basis of the between-subject production effect*. Paper presented at the 23rd Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Alberta, Canada.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8,* 513–520. doi:10.3758/BF03213770

Received June 13, 2013
Accepted September 10, 2013 ■