Memory, 2014
Vol. 22, No. 5, 470–480, http://dx.doi.org/10.1080/09658211.2013.798417

Routledge
Taylor & Francis Group

Production improves memory equivalently following elaborative vs non-elaborative processing

Noah D. Forrin, Tanya R. Jonker, and Colin M. MacLeod

Department of Psychology, University of Waterloo, Ontario, Canada

Words that are read aloud are better remembered than those read silently. Recent research has suggested that, rather than reflecting a benefit for produced items, this production effect may reflect a cost to reading silently in a list containing both aloud and silent items (Bodner, Taikh, & Fawcett, 2013). This cost is argued to occur because silent items are lazily read, receiving less attention than aloud items which require an overt response. We examined the possible role of lazy reading in the production effect by testing whether the effect would be reduced under elaborative encoding, which precludes lazy reading of silent items. Contrary to a lazy reading account, we found that production benefited generated words as much as read words (Experiment 1) and deeply imagined words as much as shallowly imagined words (Experiment 2). We conclude that production stands out as equally distinct—and consequently as equally memorable—regardless of whether it accompanies deep or shallow processing, evidence that is inconsistent with a lazy reading account.

Keywords: Production effect; Distinctiveness; Elaborative encoding.

Relative to silent reading, reading aloud enhances memory for words. This simple encoding technique, first reported by Hopkins and Edwards (1972), was only sporadically examined (see Conway & Gathercole, 1987; Gathercole & Conway, 1988; Gregg & Gardiner, 1991) until MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010) delineated the phenomenon and named it the *production effect*. Across numerous studies, production has now been shown to benefit both recognition (Forrin, Ozubko, & MacLeod, 2012; Hourihan & MacLeod, 2008; MacLeod, 2011; Ozubko, Hourihan, & MacLeod, 2012) and recall (Castel, Rhodes, & Friedman, 2013; Conway & Gathercole, 1987, Expt 3; Lin & MacLeod, 2012) and to persist across a wide variety of stimulus types, including nonwords (MacLeod et al., 2010, Expt 6), pictures (MacLeod, Ozubko, Forrin, & Hourihan, 2013), and even text (Ozubko, Hourihan, et al., 2012). In short, production is a simple but robust encoding technique, improving memory performance substantially.

The production effect has been studied primarily using a within-participant design, in which "aloud" and "silent" words are randomly intermixed in a study list (i.e., mixed list). MacLeod et al. (2010) posited that distinctiveness is the key factor driving the production effect, taking up an idea suggested earlier by Conway and Gathercole (1987).[1]

[1] Distinctiveness has been claimed to be the mechanism underlying several other mnemonics, including enactment (e.g., Engelkamp & Zimmer, 1997), generation (e.g., Begg & Snider, 1987), and bizarreness (e.g., McDaniel & Einstein, 1986). Notably, like the production effect, these other encoding techniques produce memory benefits mainly in a mixed-list design (McDaniel & Bugg, 2008), although sometimes also in a between-participants design.

Address correspondence to: Noah D. Forrin, Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1. E-mail: nforrin@uwaterloo.ca

This research was supported by Discovery Grant A7459 from the Natural Sciences and Engineering Research Council of Canada. We thank Sukhdip Grewal for her assistance in collecting the data.

A distinctiveness account emphasises the relative nature of stimuli or processes (Hunt, 2006, 2013). In the case of production in a mixed study list, speech constitutes a distinct processing dimension that causes aloud items to stand out relative to silent items. At test, participants can strategically retrieve this record of speech to confirm prior study.[2]

Several studies have refined the distinctiveness framework. Forrin et al. (2012) demonstrated that speech is not the only distinct feature that benefits memory: The production effect extends to non-verbal productions such as mouthing, writing, and typing, all of which involve distinct motor processing that is absent in silent reading. MacLeod (2011) showed that the benefit of production was larger for items produced by oneself than by another person, although there was a benefit relative to silent reading in both cases. Hourihan and MacLeod (2008) found that the distinctiveness brought about by production made words immune to intentional forgetting in a directed forgetting paradigm. And Ozubko, Gopie, and MacLeod (2012) demonstrated that production benefits both recollection and familiarity. Distinctiveness, they argued, conferred the recollective advantage on spoken words (see also Gregg & Gardiner, 1991).

Despite this converging support for a distinctiveness account, it is of course important to explore explanations of the production effect other than distinctiveness. Bodner et al. (2013) directly compared the results of within-participant vs between-participants experiments and reported data suggesting that the larger within-participant effect might reflect a cost: Reading silently in a mixed list may be worse than in a pure list. This finding is inconsistent with the distinctiveness account, which predicts a benefit of reading aloud in a within-participant design because speech stands out as a distinct processing dimension in a mixed list, but not in a pure aloud list (see Forrin, Groot, & MacLeod, 2013, for evidence of this benefit). Instead Bodner and colleagues suggest that this cost to silently reading items in a mixed list might reflect lazy reading, a hypothesis that has not yet been directly tested.

Furthermore, Bodner et al. (2013) obtained a between-participants (i.e., pure list) production effect. This result also is not consistent with a distinctiveness account because in a study list where all words are read aloud, speech is not distinctive. This between-participants effect is, however, consistent with a strength account whereby speech ‘‘strengthens’’ the memory trace of the associated word, which should occur regardless of experimental design. Although other experiments have failed to yield a significant between-participants production effect (e.g., Hopkins & Edwards, 1972; MacLeod et al., 2010; cf. Gathercole & Conway, 1988), a reliable between-participants effect has been obtained in a recent meta-analysis (Fawcett, 2013), providing support for a strength account. It should be emphasised, however, that the size of the between-participants effect in Fawcett’s meta-analysis ($g = 0.37$) is relatively small compared to the typical within-participant production effect (e.g., $g = 1.07$ in Experiment 1 of MacLeod et al., 2010). Under a strength account, it is not clear why the production effect would be smaller between participants; it would be reasonable to expect speech to enhance memory strength equivalently regardless of the study design.

Ozubko, Major, and MacLeod (2013) recently investigated the strength account by having participants study words in a mixed list once aloud, once silently, or twice silently. In a subsequent recognition test, accuracy for items presented once aloud or twice silently was equivalent, both reliably better than for items presented once silently. Thus strength of the once aloud and twice silent items was empirically equated. Participants also identified the study mode of each word (‘‘aloud’’, ‘‘silent’’, or ‘‘new’’). Ozubko et al. posited that if reading silently twice had in fact increased memory strength to a level similar to that of reading aloud once, then participants should tend to incorrectly recognise twice silent items as ‘‘aloud’’ at test. Contrary to the strength account, however, the results showed that the accuracy of study mode judgements was independent of memory strength.

Taken as a whole, then, research on the production effect appears to be more compatible with a distinctiveness account than a strength account, although it is quite possible that both factors play roles in enhancing memory for spoken words. Speech may provide a boost to signal strength—as evidenced by the significant between-participants production effect in Fawcett’s

[2] Dodson and Schacter (2001) found that participants used this ‘‘distinctiveness heuristic’’ for speech to correctly reject silent lures on a recognition test, resulting in lower false alarm rates. Here, and in our previous research, we show that this same heuristic can increase hit rates for aloud items.

(2013) meta-analysis—while distinctiveness may provide an additional benefit to aloud words in a mixed-list, which would account for the larger within-participant production effect. There is, however, another possibility that should be considered. Perhaps participants in within-participant experiments correctly assume that the researchers are studying the mnemonic benefit of reading aloud. This may lead to a demand characteristic whereby participants are motivated to pay more attention to aloud words than to silent words, or to spend more time rehearsing the aloud words. Indeed, Underwood (1983) argued that such a demand characteristic might explain why several memory effects only occur in mixed lists and not in pure lists—including the pronunciation effect of Hopkins and Edwards (1972), the point of origin of the production effect. This demand characteristic would likely not be present in a between-participants experiment, in which participants would be unlikely to guess the purpose of the experiment. In short, the production effect in a mixed-list may stem from the ‘‘active’’ processing of seemingly important aloud items (Fawcett, 2013), and the ‘‘lazy’’ processing of seemingly unimportant silent items (for more on this idea, see Bodner et al., 2013; for evidence of lazy reading in the context of the generation effect, see Begg & Roe, 1988; Begg & Snider, 1987).

How might we minimise the likelihood of lazy reading to ascertain what, if any, role it plays in the within-participant production effect? We reasoned that one way would be to require elaborative processing of all target words prior to the signal as to whether to say the word aloud or to read it silently. In this way lazy reading of the silent items would be impossible, as they would already have been elaboratively processed. The goal of the present article, then, was to determine whether the within-participant production effect is as robust when both aloud and silent items have already been elaboratively encoded—either via generation (Experiment 1) or via imagery (Experiment 2). Derived from the distinctiveness account, we had two main predictions: first, that production would improve memory for words even when those words had already been well encoded; and second, that the production effect for these well-encoded words would be comparable in size to the production effect for less well-encoded control words.

The distinctiveness account of the production effect (Conway & Gathercole, 1987; MacLeod et al., 2010) holds that production benefits memory because speech stands out as a distinct processing dimension relative to silently read items. There is no reason under this explanation that the additional record of speech should be more or less influential depending on the type of encoding with which it is combined. Therefore we expected speech to improve memory for elaboratively encoded words as effectively as it does for words that are simply read. For example, in a mixed list featuring words that were generated aloud or silently, speech would still constitute distinct, non-overlapping information relative to the process of silently generating a word from a cue. Thus we expected that a memory trace of ‘‘I said that’’ should be equally distinct and memory-enhancing regardless of whether it accompanies generating or reading.

The findings of previous research have been consistent with these predictions. MacLeod et al. (2010) showed that production provided an additional memory boost to items that were already elaboratively encoded, either by generation (Experiment 7) or by an animacy judgement (Experiment 8). They argued that production further enhanced memory for these well-encoded items in the same manner as it did for silently read words—by adding a distinct feature to the record of processing. Importantly, the elaborative encoding also encouraged participants to pay roughly equivalent attention to both aloud and silent items, preventing lazy reading of silent words. However, as other researchers have commented (Bodner et al., 2013; Fawcett, 2013), the size of the production effect in the two MacLeod et al. elaborative encoding experiments was numerically smaller than that for words that were simply read, leaving open the possibility that differential attention might have played some role in the production effect.

However, the MacLeod et al. (2010) experiments did not contrast elaborative with non-elaborative encoding: In their Experiments 7 and 8, because their goal was to explore the basic production effect, all items were elaboratively encoded. Thus, in the present article, we seek to extend the MacLeod et al. research by directly testing whether the size of the production effect is related to the degree of elaboration of encoding during study. In Experiment 1 we examined whether production improves memory for words that are generated from a definition. Extending Experiment 7 of MacLeod et al., we incorporated a read condition into their experimental design. This enabled us to compare the production effect for generated words to that for read words. We expected production to enhance memory for generated words to a similar extent as for read words. Whether study is deep (generation) or shallow (reading), production consistently contributes a distinct processing dimension that is diagnostic of prior study.

In Experiment 2, to generalise our findings, we employed another very well-established mnemonic—imagery (Paivio, 1971). We predicted that production would incrementally benefit words that were first imagined and, critically, that this benefit would be of consistent size regardless of whether the imagery was deep (imagining words as their corresponding objects) or shallow (imagining words in uppercase print). The distinctiveness account again informed this prediction: Production adds a unique processing dimension that stands out as distinct regardless of imagery type. Thus a recollection of "I said that" should be equally informative for both deeply imagined and shallowly imagined words.

EXPERIMENT 1

In Experiment 1 our goal was to replicate and extend MacLeod et al.'s (2010; Expt 7) finding that there is a production effect for generated words. We presented participants with a mixed list of words, each of which was studied in one of four different ways: generated aloud, generated silently, read aloud, or read silently. Of course, generation should lead to better memory than reading (see Bertsch, Pesta, Wiscott, & McDaniel, 2007, for a meta-analytic review). We hypothesised that production would benefit the read and generated words to a similar extent due to the addition of production distinctiveness to the record of processing in both cases. To guard against ceiling effects, particularly for the generated items, we further modified the MacLeod et al. procedure by incorporating a 3-day delay between study and test in Experiment 1A, the recognition experiment. Regardless of the elaboration of the encoding, distinct speech information should be equally available and memory-enhancing at test. To replicate and generalise our findings, in separate experiments we measured remembering with a recognition test (Experiment 1A) and with a recall test (Experiment 1B).

Experiment 1A: Method

Participants. A total of 32 University of Waterloo students participated individually for bonus course credit. The data of two participants were rejected due to difficulty generating the words (our rejection criterion for both Experiments 1A and 1B was four or more incorrect generations), resulting in final data for 30 participants. Excluding these two participants did not change our results.

Apparatus. A PC-compatible computer with a 17-inch colour monitor was used for testing. The controlling program was written in E-Prime 1.2. The apparatus was identical in all subsequent experiments.

Stimuli. The item pool comprised 80 words, each with a corresponding generation cue (e.g., for the word "wine", the cue was "the alcohol produced when grapes ferment – w?"). These generation cues were a subset of the 90 cues used by Masson and MacLeod (2002). A total of 40 word-cue pairs were randomly selected for each participant, 10 of which were assigned to each of the four study conditions: generate aloud, generate silently, read aloud, and read silently.

For the read trials a word was presented in either blue or white print and participants were asked to read blue words aloud and to read white words silently. For the generate trials a generation cue was presented in either blue or white print and participants were asked to generate the target word for the blue cues aloud and to generate the target word for the white cues silently. Both words and cues were presented at the centre of the screen in 12-point Times New Roman font against a black background.

Procedure. The study phase consisted of 20 words (10 blue, 10 white) and 20 generation cues (10 blue, 10 white) that were presented individually in random order. Each word or cue remained on the screen for 5000 ms, with a 500-ms blank screen separating successive trials. The experimenter was present and recorded incorrect generations in the generate-aloud condition.

Following a 3-day retention interval, included to ensure that performance was not at ceiling, participants completed a recognition test. The 40 words that were either read or generated (aloud or silently) at study were randomly intermixed with 40 unstudied words. Words were individually presented in the centre of the screen,

TABLE 1
Experiment 1: Mean proportions of "yes" responses in recognition (Experiment 1A), and mean proportions of correct recall (Experiment 1B)

	Generate aloud	*Generate silently*	*Read aloud*	*Read silently*	*New*
Exp. 1A: Recognition	.780 (.039)	.620 (.039)	.527 (.047)	.347 (.039)	.193 (.026)
Exp. 1B: Recall	.441 (.033)	.366 (.038)	.214 (.031)	.135 (.027)	

Standard errors are shown in parentheses below each corresponding mean.

and were printed in yellow font to avoid colour overlap between study and test. Participants were asked to press the "m" key if they recognised the word from the study phase or to press the "c" key if they thought that the word was new.

Results

The proportion of failed generations in the generate aloud condition was a modest .032.[3] The top row of Table 1 displays the hit rates for each of the four study conditions (i.e., the proportions of studied words that were recognised), along with false alarm rates for new words in the far right column.

A 2 × 2 within-participant ANOVA assessed the effects of generation (generate vs read) and of production (aloud vs silent) on recognition. As expected, there were significant main effects of both. Participants correctly recognised more generated words (.700) than read words (.437), $F(1,\ 29) = 54.61$, $MSE = .038$, $p < .001$, $\eta^2 = .65$, and more aloud words (.654) than silent words (.484), $F(1,\ 29) = 22.29$, $MSE = .039$, $p < .001$, $\eta^2 = .44$.

Most important for present purposes, there was no significant interaction between generation and production, $F < 1$, Power = .069. Consistent with our hypothesis the production effect for generated words (mean difference of 16%), $t(29) = 3.97$, $p < .001$, $\eta^2 = .35$, was closely comparable in size to the production effect for read words (mean difference of 18%), $t(29) = 3.91$, $p = .001$, $\eta^2 = .35$. We also analysed the interaction using the Bayesian approximation procedure proposed by Wagenmakers (2007); the posterior odds were estimated from the sums of squares from ANOVA using a calculator provided by Masson (2011). The posterior odds compare the likelihood of two different models. In this case one model assumes that Production and Generation interact, whereas the other assumes that they do not interact (the null). The posterior odds can be converted into p_{BIC}, which will quantify the degree of support favouring the null relative to the alternative hypothesis on a scale of 0 to 1, with 1 indicating full support for the null. According to Raftery's (1995) system for labelling the strength of the evidence, this analysis yielded "positive" evidence in favour of the null hypothesis, $p_{\text{BIC}} = .83$, suggesting that a model assuming no interaction between Production and Generation is preferable.

[3] We assume that participants had a similar failure rate in the generate silent condition, although it was not possible to unobtrusively keep track of generation failures in this condition. Removing failed generations from the data did not change the pattern of results.

Experiment 1B: Method

Experiment 1A demonstrated that, when measured using recognition, production provides a comparable boost to memory for both read and generated words. Thus it appears that words that are generated aloud reap the benefits of both generation and production—benefits that are independent and essentially additive. In Experiment 1B we sought to replicate and extend our results using a recall test. Although the great majority of the research on the production effect has involved recognition tests, recent research has shown a strong production effect for recall tests as well (Castel et al., 2013; Lin & MacLeod, 2012; MacLeod, Ozubko, Forrin, & Hourihan, 2013; see also Conway & Gathercole, 1987, Expt 3), leading us to examine whether production and generation would also make independent contributions to recall.

Participants. A total of 32 University of Waterloo students participated individually for bonus course credit. The data of one participant were rejected for failing to recall a single study item, and the data of two participants were rejected due to difficulty generating the words, resulting in final data for 29 participants. The exclusion of these participants did not affect our results.

Stimuli. The item pool consisted of the same 80 word-cue pairs from Experiment 1A.

Procedure. The procedure was identical to Experiment 1A with one exception: Rather than performing a recognition test after a 3-day delay, participants were given a free recall test immediately following study. Prior to the recall test participants were instructed to write down, in any order, as many words as they could recall having previously studied. They were provided with a blank sheet of paper and were given as much time as they needed.

Results

The generation failure rate for the generate aloud condition was .066. The bottom row of Table 1 shows the proportions of words correctly recalled for each of the four study conditions. As was the case in Experiment 1A, a 2 × 2 ANOVA revealed significant main effects of generation, $F(1, 28) = 63.96$, $MSE = .024$, $p < .001$, $\eta^2 = .70$, and of production, $F(1, 28) = 14.46$, $MSE = .012$, $p = .001$, $\eta^2 = .34$. Generated items (.404) were recalled better than read items (.174), and aloud items (.328) were recalled better than silent items (.250). Crucially, the interaction between generation and production was again non-significant, $F < 1$, Power = .051. The production effect for generated words (mean difference of 7%), $t(28) = 2.31$, $p < .05$, $\eta^2 = .16$, was the same magnitude as it was for read words (mean difference of 7%), $t(28) = 2.58$, $p < .05$, $\eta^2 = .19$. Again, a Bayesian approximation procedure yielded "positive" evidence in favour of the null hypothesis, $p_{BIC} = .84$, suggesting that a model assuming no interaction between Production and Generation is preferable.

Discussion

In sum, the pattern of results for recall (Experiment 1B) replicated that for recognition (Experiment 1A). Participants' memory for both read and generated words was enhanced by production to a comparable extent regardless of type of test. Despite the elaborative encoding promoted by generation, it appears that additional distinctiveness was conferred on generated words by producing aloud the word that fit the cue. Moreover, the fact that production improved memory by a similar extent for both generated and read words suggests that aloud information stands out as equally distinct—and consequently equally memorable—regardless of whether it accompanies deep or shallow processing.

The lack of significant interaction between the size of the production effect and the encoding type is also inconsistent with lazy reading being responsible for the reported cost to silent items in a within-participant production effect (e.g., Bodner et al., 2013). In the present experiments silently generated words should be less susceptible to lazy processing than should silently read words, yet the production effect was just as robust for generation as it was for reading.

EXPERIMENT 2

In Experiment 2 we combined production with another well-established memory technique: visual imagery (see Paivio, 1971). To accomplish this we took advantage of a technique for imaging used by Hourihan (2009) in her dissertation and subsequently used by Danckert, MacLeod, and Fernandes (2011). Participants imagined each lowercase word in a study list either deeply (as the object to which the word referred) or shallowly (by imagining the word in uppercase). Immediately following imagery each word was read either aloud or silently. We expected that imagery and production would independently benefit memory, just as generation and production were found to have independent effects in Experiment 1.

In Experiment 2 we manipulated imagery between participants, unlike the within-participant manipulation of generation in Experiment 1. This design change was deliberate. We were concerned that in a within-participant design participants would realise that deep imagery is a more effective memory technique, and that they might therefore use deep imagery even when instructed to image shallowly. Thus we opted for a between-participants design, thereby further generalising our pattern with respect to experimental design. We had two main predictions: (1) production would further enhance memory even for deeply imagined words; and (2) the benefit of production would be of consistent magnitude, regardless of whether the imagery was deep or shallow.

Method

Participants. A total of 48 University of Waterloo students participated individually for bonus course credit; 24 participated in the deep imagery condition, and 24 in the shallow imagery condition.

Stimuli. The item pool consisted of 120 words taken from the MRC Psycholinguistic Database (http://www.psych.rl.ac.uk/MRC_Psych_Db.html). All of the words were nouns, four to eight letters long, that had frequencies of greater than 30 per million (Thorndike & Lorge, 1944). Imagery ratings were available for 99 of the words ($M = 583.30$, $SD = 38.24$), these ratings suggesting that the words were quite easily imaged. Each word was initially presented in white lowercase 16-point Courier New font against a black background.

Procedure. The study phase consisted of 60 words presented individually in random order. Participants in the deep imagery condition imagined each word as the object that it represented, whereas those in the shallow imagery condition imagined each word in all uppercase letters. Participants in both conditions were instructed to press the spacebar as soon as they had imagined the object or uppercase word. The program recorded how long it took participants to press the spacebar on each trial.

After the spacebar was pressed, the word immediately changed colour to either blue or pink. Half of the participants in each imagery condition were told to read the word aloud if it became blue and to read the word silently if it became pink. To counterbalance, the other half of the participants were told to read pink words aloud and blue words silently. Each word remained on the screen for 2000 ms after the colour change, and was followed by a 500-ms blank that preceded the next word in the study list. A free recall test immediately followed the study phase. Participants were instructed to type in all of the words that they could remember, one at a time, regardless of the colour that the words had changed into.

It is worth mentioning that participants spent an equivalent amount of time imagining words from the aloud ($M = 3.22$ seconds, $SE = .023$ seconds) and silent ($M = 3.14$ seconds, $SE = .024$ seconds) conditions, $F(1, 46) = 1.19$, *ns.* Given that participants did not know which condition (i.e., aloud or silent) each word belonged to until after they had imagined it, this result is not surprising; however, it does suggest that participants initially paid roughly equal attention to aloud and silent words. This is inconsistent with a lazy reading explanation (see Begg & Snider, 1987) for the observed production effect: Participants could not have lazily read the silent words because they first had to imagine them.

Results

Table 2 shows the proportions of aloud and silent words correctly recalled for the deep imagery (picturing words as objects) and shallow imagery (picturing words in all uppercase) conditions. The counterbalancing of print colour and condition did not affect the results.

A 2 × 2 mixed ANOVA was conducted, with imagery (deep vs shallow) as the between-participants variable and production (aloud vs silent) as the within-participant variable. We found a significant main effect for imagery, with deep imagery (.297) leading to better recall than shallow imagery (.185), $F(1, 46) = 16.45$, $MSE = .018$, $p < .001$, $\eta^2 = .26$. It is noteworthy that this difference did not result from participants spending significantly more time engaging in deep imagery ($M = 3.38$ seconds, $SE = .034$ seconds) than in shallow imagery ($M = 2.98$ seconds, $SE = .033$ seconds), $F < 1$. The main effect of production was also reliable, $F(1, 46) = 13.80$, $MSE = .007$, $p = .001$, $\eta^2 = .23$, with words said aloud (.272) recalled better than words read silently (.210). Critically, the interaction between imagery and production was non-significant, $F < 1$, $\eta^2 = .004$, Power = .069, which demonstrates that the production effect was of comparable size

TABLE 2
Experiment 2: Mean proportions of correct recall

	Aloud	*Silent*
Deep imagery	.332	.262
	(.024)	(.026)
Shallow imagery	.213	.157
	(.020)	(.022)

Standard errors are shown in parentheses below each corresponding mean.

regardless of whether the words were deeply imagined (mean difference of 7%), $t(23) = 3.23$, $p < .01$, $\eta^2 = .31$, or shallowly imagined (mean difference of 6%), $t(23) = 2.14$, $p < .05$, $\eta^2 = .17$. Indeed, a Bayesian approximation procedure yielded ‘‘positive’’ evidence in favour of the null hypothesis, $p_{\text{BIC}} = .86$, suggesting that a model assuming no interaction between Production and Generation is preferable.

GENERAL DISCUSSION

Across two experiments we have demonstrated that production improves memory for words even when those words have already been well encoded by generation (Experiment 1) or by deep imagery (Experiment 2). Moreover, the production effect for these elaboratively encoded items was just as large as the production effect for weakly encoded comparison items. In Experiment 1 production enhanced memory for both generated and read words by a comparable margin, and this occurred for both recognition and recall. In Experiment 2 production further enhanced the recall of words that had already benefited from imagery. Again, this benefit was equally large regardless of whether the words had been deeply imagined (as objects) or shallowly imagined (in all uppercase letters), in this case in a between-participants design.

This research provides evidence that production improves memory for deeply processed words to an extent similar to that for shallowly processed words. This finding is consistent with the distinctiveness account: Speech constitutes an additional, distinctive process regardless of the encoding that it accompanies, whether that encoding is shallow or deep. In either case, speech serves as a distinctive cue that can subsequently aid retrieval. We assume that the distinctiveness of speech was not diminished by either generation or imagery because the process of vocalising does not overlap with the processes of covertly generating a word from a cue or imagining the physical referent of a word; speech simply adds a new element to the record.

As mentioned in the introduction, it has recently been claimed that the production effect yields only costs to memory in a mixed-list design (Bodner et al., 2013; cf. Forrin et al., 2013). Bodner and colleagues argue that such a cost might result from the lazy reading of some of the silent items in a mixed list, just as lazy reading might contribute to the generation effect (Begg & Roe, 1988; Begg & Snider, 1987). The present research addressed this lazy reading idea by having participants elaboratively encode words prior to production, under the assumption that elaboration would largely prevent lazy processing of silent items. If lazy reading was indeed imposing a cost on silent items, then one would expect this elaborative encoding manipulation to attenuate the size of the production effect. This is not what we found in our experiments. In Experiment 1 the production effect, whether measured by recognition or by recall, was as robust for generating as it was for reading, even though silently generated words would be expected to be less susceptible to lazy processing than would silently read words. In Experiment 2 the production effect was as large for deeply imagined words as it was for shallowly imagined words, even though deep imagery would have ensured that silent items were elaboratively processed before a change in print colour indicated that they belonged to the silent condition. In fact, the Bayesian approximation procedure demonstrated that, in all cases, a model assuming no interaction between Production and Generation was preferable.

Thus, contrary to the lazy reading account, the present research suggests that the magnitude of the mixed-list production effect is not inflated by the cursory encoding of silently read words. But if lazy reading does not impose a cost to the silent items, then what accounts for the cost reported by Bodner et al. (2013)? One possibility is that participants may be biased towards classifying silent items as new because silent items lack distinct aloud information. If participants are using a distinctiveness heuristic to help identify studied words (Dodson & Schacter, 2001), then the *absence* of aloud information constitutes evidence that the word was not studied. In the context of a Deese-Roediger-McDermott (DRM) paradigm, Dodson and Schacter found that participants were less likely to incorrectly recognise lures if the study words had been read aloud than if they had been heard, leading them to conclude that participants were using the lack of distinct aloud information to infer that a word was not studied (cf. Bodner & Taikh, 2012, with regard to a list-discrimination task). This ‘‘negative’’ distinctiveness account is supported by several production experiments that have examined modality judgements and have found that participants are more likely to incorrectly recognise new items as ‘‘silent’’ than as ‘‘aloud’’ (Conway & Gathercole, 1987; Forrin et al., 2013; Ozubko et al., 2012). This is presumably because both silent items and new items lack a distinct aloud record in memory. In short, distinctiveness may be a double-edged sword when it comes to the mixed-list production effect: The use of a distinctiveness heuristic at test might help participants to differentiate between aloud and new items, but this same heuristic may increase the likelihood that participants conflate silent and new items.

Despite the apparent cost of reading silently in a mixed list, evidence remains that production also benefits aloud items. The recent meta-analysis by Fawcett (2013) shows a reliable between-participants production effect, and recent research in our laboratory has demonstrated that the production effect is larger using a mixed-list design than a pure-list design (Forrin et al., 2013), consistent with a distinctiveness account. Although the present paper suggests that silent items do not suffer from lazy reading, it is still possible that aloud items benefit from greater attention in that they require an overt response.

In short, the present research suggests that production significantly improves memory for generated words and for imagined words. Indeed, production even appears to provide an independent contribution to the memorability of both generated and imagined words; that is, production and generation may have additive effects, as might also be the case with production and imagery. The evidence that the production effect is as strong for generated words as for read words (Experiment 1) and for deeply imagined words as for shallowly imagined words (Experiment 2) is certainly consistent with the claim that production provides an independent boost to the memorability of words, irrespective of the type of processing that they have undergone.

Notably, the proceduralist framework (Kolers, 1973) supports the claim that the production effect should be independent of the effects of generation and imagery. When a word is produced at study, the process of vocalisation (“I said that”) is added to the processing record; likewise, the processes of generation and imagery are also preserved in this record (Kolers & Roediger, 1984). At test, participants can then replay these processing records to facilitate remembering. We therefore assume that the record of vocalising will be *independent* of the records of generating or imagining. Previous research using the Remember/Know paradigm has demonstrated that production (Gregg & Gardiner, 1991; Ozubko, Gopie, et al., 2012), generation (Gardiner, 1988; Gardiner, Java, & Richardson-Klavehn, 1996; Sheridan & Reingold, 2011), and imagery (Dewhurst & Conway, 1994, Expt 5; O’Neill, 2005), all enhance the conscious recollection of words. Thus, in the case of generating aloud, participants essentially have two chances to remember a word: The recollection of either “I said that” or “I generated that” will lead to successful retrieval. A parallel case can be made for imagery.

Perhaps incongruent with this independence claim, however, is the view that distinctiveness may underlie not only production (Conway & Gathercole, 1987; MacLeod et al., 2010), but also generation (Begg, Snider, Foley, & Goddard, 1989; Gardiner & Hampton, 1985, 1988; Kinoshita, 1989; cf. Schmidt, 1992) and imagery (Einstein, McDaniel, & Lackey, 1989; Marschark, Richman, Yuille, & Hunt, 1987; McDaniel & Einstein, 1986). Assuming that distinctiveness underlies the benefits of each of these encoding techniques, it might seem unlikely, at first blush, that their effects would be additive. However, if the distinctiveness conferred by production is independent of the distinctive processing involved in generation and imagery—if it is different in kind—then production should provide an additive benefit, as observed in the present experiments. Along these lines, Hunt and Mitchell (1982) have demonstrated that the effects of conceptual distinctiveness and orthographic distinctiveness are additive.

To sum up, the present research clearly demonstrates that production improves memory for elaboratively encoded words—both for words that are generated (Experiment 1) and for those that are deeply imagined (Experiment 2). Moreover, these production benefits appear equal in magnitude to the standard production effect for read words (Experiment 1; or for words that are shallowly imagined and then read, Experiment 2). Therein lie the simplicity and elegance of production as a mnemonic device: Vocalisation can easily supplement other memory techniques, and can do so without losing its potency.

Manuscript received 21 October 2012
Manuscript accepted 17 April 2013
First published online 23 May 2013

REFERENCES

Begg, I., & Roe, H. (1988). On the inhibition of reading by generation. *Canadian Journal of Psychology*, *42*, 325–336.

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 553–563.

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 977–989.

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

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.

Bodner, G. E., Taikh, A., & Fawcett, J.M. (2013). Assessing the costs and benefits of production in recognition. Manuscript submitted for publication.

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

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341–361.

Danckert, S. L., MacLeod, C. M., & Fernandes, M. A. (2011). Source-constrained retrieval influences the encoding of new information. *Memory & Cognition*, *39*, 1374–1386.

Dewhurst, S. A., & Conway, M. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1088–1098.

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.

Einstein, G. O., McDaniel, M. A., & Lackey, S. (1989). Bizarre imagery, interference, and distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *1*, 137–146.

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

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

Forrin, N. D., Groot, B., & MacLeod, C. M. (2013). *The production effect in memory: Costs and benefits.* Manuscript in preparation.

Forrin, N. D., Ozubko, J. D., & MacLeod, C. M. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*, 1046–1055.

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

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 732–741.

Gardiner, J. M., & Hampton, J. A. (1988). Item-specific processing and the generation effect: Support for a distinctiveness account. *American Journal of Psychology*, *101*, 495–504.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology*, *50*, 114–122.

Gregg, V. H., & Gardiner, J. M. (1991). Components of conscious awareness in a long-term modality effect. *British Journal of Psychology*, *82*, 153–162.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, *11*, 534–537.

Hourihan, K. L. (2009). The power of optimal encoding: Distinctiveness and differentiation defeat directed forgetting. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *69*, 7159.

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.

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.

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

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

Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, *17*, 563–571.

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

Kolers, P. A., & Roediger, H. L. III (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior*, *23*, 425–449.

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.

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*, 1197–1202.

MacLeod, C. M., & Daniels, K. A. (2000). Direct vs indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review*, *7*, 354–359.

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.

MacLeod, C. M., Ozubko, J. D., Forrin, N. D., & Hourihan, K. L. (2013). *Assessing the basis of the* *production effect: Evaluating the distinctiveness account*. Manuscript in preparation.

Marschark, M., Richman, C. L., Yuille, J. C., & Hunt, R. R. (1987). The role of imagery in memory: On shared and distinctive information. *Psychological Bulletin*, *102*, 28–41.

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679–690.

Masson, M. E. J., & MacLeod, C. M. (2002). Covert operations: Orthographic recoding as a basis for repetition priming in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 858–871.

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

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

O'Neill, W. (2005). Word-imagery effects on recollection and familiarity in recognition memory. *Perceptual and Motor Skills*, *100*, 716–722.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*, 326–338.

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

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

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

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–196). Cambridge, MA: Blackwell.

Schmidt, S. R. (1992). Evaluating the role of distinctiveness in the generation effect. *The Quarterly Journal of Experimental Psychology*, *44A*, 237–260.

Sheridan, H., & Reingold, E. M. (2011). Recognition memory performance as a function of reported subjective awareness. *Consciousness and Cognition*, *20*, 1363–1375.

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

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

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804.