Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

The production effect is consistent over material variations: support for the distinctiveness account

Colin M. MacLeod, Jason D. Ozubko, Kathleen L. Hourihan & Jennifer C. Major

MEMORY
2022, VOL. 30, NO. 8, 1000–1007
https://doi.org/10.1080/09658211.2022.2069270

The production effect is consistent over material variations: support for the distinctiveness account

Colin M. MacLeod [a], Jason D. Ozubko [b], Kathleen L. Hourihan[c] and Jennifer C. Major[d]

[a]Department of Psychology, University of Waterloo, Waterloo, Canada; [b]SUNY, Geneseo, NY, USA; [c]Memorial University of Newfoundland, St. John's, Canada; [d]Wilfrid Laurier University, Waterloo, Canada

ABSTRACT
The production effect is the superior memory for items read aloud as opposed to silently at the time of study. The distinctiveness account holds that produced items benefit from the encoding of additional elements associated with the act of production. If so, then that benefit should be consistent regardless of item type. Three experiments, using three different sets of materials and three different methods, tested this hypothesis. Experiment 1, using recognition testing, showed consistent production benefits for high and low frequency words. Experiment 2, using free recall, showed consistent production increments for pictures and words. Experiment 3, using incidental learning, showed consistent production benefits for recognition of nonwords and words. Taken together, these results fit with the distinctiveness account: Production at encoding dependably adds information to the memory record, regardless of item type or method of testing, producing a consistently reliable memory benefit.

ARTICLE HISTORY
Received 17 June 2021
Accepted 18 April 2022

KEYWORDS
Memory; production effect; distinctiveness

Over the past 50 years, studies have repeatedly shown the memory benefit that accrues for words that people read aloud compared to words that they read silently. Hopkins and Edwards (1972) were the first to show this enhancement (although Barlow, 1928, had reported a closely related finding). Fifteen years later, Conway and Gathercole (1987; Gathercole & Conway, 1988) replicated and extended their finding. Ten years later, MacDonald and MacLeod (1998) also replicated and extended the finding (and see Dodson & Schacter, 2001). And then, after another 12 years, MacLeod et al. (2010) again replicated and extended the phenomenon and finally gave it a name – *the production effect*. In all that time, the production effect has yet to achieve the status that it warrants: to stand beside other well-known encoding tasks as a simple manipulation that has a robust effect on memory.

We now know a considerable amount about the production effect (for a brief review, see MacLeod & Bodner, 2017). It is evident both in recognition (as shown in all of the studies just cited) and in recall (as shown, e.g., by Conway & Gathercole, 1987; Lin & MacLeod, 2012; Putnam et al., 2014). It is expressed whether learning is incidental (Conway & Gathercole, 1987; Gathercole & Conway, 1988; MacDonald & MacLeod, 1998) or intentional (most of the other studies to date). MacLeod et al. (2010) further delineated the phenomenon by showing that it is observed whether encoding is relatively shallow – reading the words – or deep – generating the words or making a semantic judgment (see also Forrin et al., 2014) – and even when the words were only mouthed, not actually spoken (see also Forrin et al., 2012). Forrin et al. (2012) showed that spoken production is best but that other kinds of productions – typing, writing, or spelling – are also beneficial. And Ozubko et al. (2012b) demonstrated enhanced long-term retention after production and extended the effect to more complex materials than words, including sentences and texts (see also Rommers et al., 2020). All of these studies have shown the production effect to be a robust phenomenon.

In the beginning, the effect did not seem to occur between subjects (i.e., for pure lists), as Hopkins and Edwards (1972) and MacLeod et al. (2010) had observed (see also Dodson & Schacter, 2001). Rather, it seemed to occur only within subject in mixed lists. This suggested to Conway and Gathercole (1987) and to MacLeod et al. (2010) that, for the benefit to be expressed, the produced items had to stand out from other items that were not produced. In turn, this gave rise to the explanation that produced items are well remembered because distinct aloud information is preserved in the record of encoding and can be retrieved at test to confirm prior study. Forrin and MacLeod (2018) suggested that two of these distinct elements likely were articulation and audition. Computational modelling efforts using MINERVA2 (Jamieson et al., 2016), the Revised Feature Model (Saint-Aubin et al.,

CONTACT Colin M. MacLeod cmacleod@uwaterloo.ca Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada

2021), and REM (Kelly et al., 2022) have all successfully implemented this additional feature(s) idea of distinctiveness to capture key findings in the production effect literature. Distinctiveness has a quite long history in explaining other phenomena (see, e.g., Hunt & Elliot, 1980; Schmidt, 1991; for a theoretical overview, see Hunt, 2006; 2013) and provides a good fit for the production effect.

Under the distinctiveness explanation, recollection plays a key role because at the time of test the fact that an item was produced earlier during study has been encoded and can be retrieved and used to certify prior study. Ozubko et al. (2012a; see also Fawcett & Ozubko, 2016), using the mixed list, within-subject procedure, have demonstrated that the production effect is evident for recollected items. Ozubko et al. found this to be the case using both remember/know judgments (Gardiner, 1988; Rajaram, 1993) and receiver operating characteristic procedures (Yonelinas, 1994, 1997) to examine the influences of recollection and familiarity. Under both approaches, and to a consistent extent as indexed by the dual-process signal-detection model, there was a recollection advantage as well as a familiarity advantage for words produced aloud at study.

There are other findings that also dovetail with those just summarised in supporting the distinctiveness account of the production advantage. The effect is larger when a person produces the words themselves than when they listen to someone else produce them (MacLeod, 2011): One's own productions are more distinctive (see also Forrin & MacLeod, 2018). Produced items cannot be forgotten on cue in a directed forgetting paradigm (Hourihan & MacLeod, 2008): Distinctive items appear to be invulnerable to intentional forgetting, as MacLeod and Daniels (2000) also showed for items that had been generated at study. Other modes of production, such as typing or handwriting also produce benefits (Forrin et al., 2012); they too provide distinctive extra information that can be retrieved to assist remembering.

By 2013, it became clear that there is in fact a production effect in between-subjects pure list designs (Fawcett, 2013), although that effect is notably smaller than the within-subject mixed list effect (but see Zhou & MacLeod, 2021). The between-subjects pure list effect could be evidence of a considerably reduced use of the distinctiveness heuristic or it could be that it reflects a different process: an increment in strength for aloud items relative to silent items (see MacLeod & Bodner, 2017, for discussion). Consistent with the strength explanation, Fawcett and Ozubko (2016) showed that the between-subjects pure list effect was apparent only for familiarity and not for recollection, unlike the within-subject mixed list effect, which was apparent for both familiarity and recollection (confirming the findings of Ozubko et al., 2012a). Thus, as MacLeod and Bodner (2017) summarise, the production effect may rely solely only on strength when production is manipulated between subjects, but a larger distinctiveness contribution is added when production is manipulated within subject.

In this article, our primary goal is to put the distinctiveness account to further test. If the addition of distinctive elements supports the production effect in mixed list experiments, then that contribution would be expected to be quite similar across situations: The increment due to production should be consistent for different types of processing and for different types of materials. Forrin et al. (2014) addressed the first situation – different forms of processing. They showed that production resulted in a consistent benefit whether processing was more elaborative (generation, imagining the pictorial referent) or less elaborative (silent reading, visualising the word), despite the considerably better overall performance in the more elaborative conditions.

Here, we address the second situation – different types of materials – at the same time addressing different kinds of experimental procedures. The procedural changes are meant to generalise the results and to connect to other studies already in the literature, although admittedly at the risk of somewhat complicating conclusions. In three experiments, we examined production in the context of material effects – the better recognition of low frequency words than high frequency words (Experiment 1), the better recall of pictures than words (Experiment 2), and the better recognition of words than nonwords (Experiment 3). Under the distinctiveness account, we predicted that the benefit of production would be consistent across materials because, independent of type of material, production provides an additional encoding dimension that is essentially the same regardless of the type of material and that can be retrieved and used to assist remembering.

Experiment 1

What would happen to the production effect if the stimuli were already distinctive in another way? We explored the relative benefit of production for different types of materials. To begin, it is well established that low frequency words are better recognised than high frequency words (e.g., MacLeod & Kampe, 1996) – typically by low frequency words exhibiting higher hit rates and lower false alarm rates than high frequency words. How would the benefit of production compare for low versus high frequency words? Under the distinctiveness account, the benefit should not be modulated by item type because the distinctive information that an item was said aloud does not hinge on what that item was. Thus, we predicted that word frequency would not interact with whether encoding mode was aloud or silent: The two types of items should both show consistently reliable memory advantages when produced, given reliance on the same additional encoded features.

Method

Participants. Twenty undergraduate students from the University of Waterloo took part. The sample sizes in the three experiments reported here were chosen because they had revealed robust production effects in prior studies (see, e.g., the series of experiments in MacLeod et al., 2010).[1] In all three experiments, only participants who reported English as their first language were eligible. All received either bonus credit in a course or payment. Note that, in each experiment reported in this article, the sample of participants had not taken part in any other production effect experiment. Each experiment received ethics approval from the Office of Research Ethics, University of Waterloo.

Materials. Two word pools – one of 145 high frequency words (e.g., army, mother, window) and one of 475 low frequency words (e.g., acid, frost, nephew) – were used. Care was taken to select low frequency words that would be familiar to undergraduate students whose first language was English. All words were 4–9 letters in length (mean of 5.69). According to the Kučera and Francis (1967) norms, the high frequency words had frequencies of 100–250 ($M = 154.92$) and the low frequency words had frequencies of 3–15 ($M = 7.04$). Each set was first randomised for each participant. From each frequency set, 60 words were then selected, 20 to be studied aloud, 20 to be studied silently, and 20 to serve as distractors on the recognition test. Thus, the study list contained 80 words and the test list contained 120 words, as in most of our previous production effect experiments (see, e.g., MacLeod et al., 2010). During study, words in blue were to be read aloud and words in white were to be read silently without moving lips.[2] At test, all words were presented in yellow, to prevent overlap with the colour at study. All word presentations at study and at test were in Courier New bold, point size 16, and were in lower case. The words were always presented at the centre of the screen, and the background colour was always black.

Apparatus. All testing was carried out on PC-compatible computers with 17-inch CRT colour monitors. Experimental control was accomplished via an E-Prime programme (www.pstnet.com).

Procedure. The blue and white words were intermingled in a different random order for each participant. Each word was presented for 2 seconds, with a 500-ms blank screen between successive words. An experimenter was present to ensure that participants actually spoke only the blue words aloud. During test, a new randomisation unique to each participant was created, with the 80 old words and the 40 new words intermingled. Each test word stayed on the screen until the participant pressed either the "m" key to indicate "studied" or the "c" key to indicate "not studied". There was a 500-ms blank between successive test words.

Table 1. Experiment 1 (high-low frequency): Mean proportion of "yes" responses in each condition.

	Hit Rate		False Alarm Rate
Word Frequency	Aloud	Silent	New
Low	.825 (.024)	.560 (.044)	.143 (.021)
High	.770 (.026)	.538 (.040)	.255 (.040)

Note: Standard errors are shown in parentheses below their respective means.

Results & discussion

The proportions of "yes" responses are shown in Table 1. As expected, the false alarm rate was greater for high frequency distractors than for low frequency distractors, a significant difference, $t(19) = 2.787$, $p = .012$, Cohen's $d = 0.623$.

A 2×2 repeated measures ANOVA was carried out on corrected recognition scores (hits minus false alarms), the two within-subject factors being Word Frequency (high vs. low) and Encoding Mode (aloud vs. silent). The main effect of Word Frequency was significant, $F(1, 19) = 10.909$, $MSE = 0.042$, $p = .004$, partial $\eta^2 = .365$, with accuracy higher for low frequency words ($M = .550$, $SE = .032$) than for high frequency words ($M = .399$, $SE = .032$), in keeping with typical frequency findings for recognition (see, e.g., MacLeod & Kampe, 1996). There was a significant main effect of Encoding Mode, $F(1, 19) = 75.400$, $MSE = 0.016$, $p < .001$, partial $\eta^2 = .799$, with words read aloud ($M = .599$, $SE = .027$) better recognised than words read silently ($M = .350$, $SE = .027$). The nonsignificant interaction, $F(1, 19) = 0.612$, $MSE = 0.009$, $p = .444$, partial $\eta^2 = .031$, indicated that the production effect (Aloud – Silent) was present both for high frequency words ($M = .233$; 95% CI [.158, .307]) and for low frequency words (M = .265; 95% CI [.191, .339]), and to a similar extent.

As predicted, despite recognition being better for low frequency words than for high frequency words, as is typically observed, the production advantage was consistent for the two item types. The additional information that a word was said aloud during study was available and useful to a similar extent for the two types of items, as predicted by the distinctiveness explanation.

Experiment 2

Our view of the distinctiveness imparted by production is that the act of producing a word adds a further distinctive element to memory – that the encoding of the item includes extra features as a result of having been produced during study. Under this view, the production effect would be expected to be quite consistent across stimulus types, as Experiment 1 showed for high and low frequency words. In Experiment 2, the materials were pictures (line drawings) and words. The production effect has previously been demonstrated for recognition of pictures (Fawcett et al., 2012; Hourihan & Churchill, 2020), and the picture superiority effect – over words – is well documented in the memory literature (Nelson et al., 1976; Weldon et al., 1989). For generalisation, we switched from recognition to free recall in this experiment (for other instances of production effects in recall, see Conway & Gathercole, 1987;

Lin & MacLeod, 2012; Putnam et al., 2014) and also manipulated material type between subjects rather than within subject.

Method

Participants. Thirty different students from the same pool took part, with 15 assigned to each of the picture and word conditions.

Materials. The picture stimuli were 32 object pictures selected from the pool of 244 pictures publicly available online as freeware from the set of 520 used by Székely et al. (2003); https://crl.ucsd.edu/experiments/ipnp/method/getpics/getpics.html. These were chosen empirically for having very high naming consistency in those norms and intuitively for their visual clarity, with care taken to avoid related pictures or concepts. Each picture was a line drawing of a common object in black on a white background. The one-word names of the selected pictures served as the word stimuli for the word condition. All of the pictures were re-coloured as blue on a white background to be used when presented on Aloud trials. Words were presented in black or blue 18-point Courier font. All displays used a white background.

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. In the study phase, depending on condition, participants were told that they would be asked to remember a list of words or of pictures for a later memory test, but they were not informed of the exact nature of the test. Participants in the picture condition were instructed to name blue pictures aloud and black pictures silently; those in the word condition were instructed to read blue words aloud and black words silently. An experimenter was present to ensure that participants actually spoke only the blue items aloud. Each trial began with a 500-ms blank screen, then the study item was presented at the centre of the screen for 3 s. All 32 study items were presented, in random order, against a white background, with item colour assigned randomly per participant (with the restriction that 16 items were presented in blue and 16 items were presented in black).

Immediately following the study phase, participants were given a piece of paper by the experimenter and were instructed to write down as many picture names or words as they could recall, regardless of their encoding condition. There was no time limit imposed.

Table 2. Experiment 2 (picture-word): Mean proportion of items correctly recalled in each condition.

Stimulus Type	Aloud	Silent
Picture	.500	.304
	(.041)	(.034)
Word	.321	.175
	(.034)	(.028)

Note: Standard errors are shown in parentheses below their respective means.

Results & discussion

Mean proportions of items correctly recalled are displayed in Table 2. These were analysed in an Encoding Mode (aloud vs. silent) x Item Type (words vs. pictures) mixed ANOVA, with Encoding Mode as the within-subject factor and Item Type as the between-subjects factor. The main effect of Item Type was significant, $F(1,28) = 19.457$, $MSE = .018$, $p < .001$, partial $\eta^2 = .410$, showing the typical picture superiority effect: Picture names ($M = .402$, $SE = .025$) were recalled considerably more often than were the corresponding words ($M = .248$, $SE = .025$). The main effect of Encoding Mode was significant, $F(1,28) = 24.427$, $MSE = .018$, $p < .001$, partial $\eta^2 = .466$, with aloud items ($M = .411$, $SE = .025$) recalled considerably more often than Silent items ($M = .240$, $SE = .025$). There was, however, no evidence of an interaction, $F(1,28) = 0.526$, $MSE = .018$, $p = .474$, partial $\eta^2 = .018$, indicating that the production benefit (Aloud – Silent) did not differ reliably for pictures ($M = .196$; 95% CI [.080, .312]) versus words ($M = .146$; 95% CI [.054, .238]).

Experiment 2 complements Experiment 1 in showing another item effect that does not modulate the production effect, this time when materials were manipulated between subjects and the test was free recall. Just as the production effect was consistent for high and low frequency words, it was consistent for words and pictures. Experiment 2 also shows this additive boost due to production to be stable across test format, given that Experiment 1 had used a recognition test whereas Experiment 2 used a recall test.

Experiment 3

In Experiments 1 and 2, we provided evidence demonstrating that the size of the production effect does not depend on the stimulus material or on the test format – or on whether material type is manipulated within subject or between subjects. Experiment 3 continued this approach by examining production for words and nonwords. We predicted from the distinctiveness account that the magnitude of the production effect would be similar for words and nonwords, just as was the case for high and low frequency words (Experiment 1), and for pictures and words (Experiment 2). MacLeod et al. (2010, Experiment 6) did previously show a production benefit for nonwords, but this was in the context of intentional study of an unmixed list made up of nonwords only.

In Experiment 3, we switched to incidental learning instructions to further test the generalisability of the production effect (note that MacDonald & MacLeod, 1998, had used incidental learning instructions). The study phase was disguised as a lexical decision experiment with two types of studied material – words and nonwords – read either aloud or silently. There was no mention of a memory test. That an item has been studied aloud stands out as distinct in a mixed list, and we expected this distinctive information to be routinely encoded, similarly to how frequency information is encoded (cf. Hasher & Zacks, 1979).

Method

Participants. Twenty-one different students from the same pool took part.

Materials and Apparatus. The pool of 120 words had been used in the experiments reported by MacDonald and MacLeod (1998) and by MacLeod et al. (2010). The pool of 120 pronounceable nonwords (e.g., binch, hest, prench) had been used in Experiment 6 of MacLeod et al. (2010). These nonwords were composed by altering a single letter – almost always a consonant – in a word such that the resulting nonword remained readily pronounceable. Each set was first randomised for each subject. From each set, 90 items were then selected, 30 to be studied aloud, 30 to be studied silently, and 30 to be distractors on the recognition test. Thus, the study list consisted of 120 items and the test list of 180 items. All other presentation details were the same as in Experiment 1, as was the apparatus.

Procedure. The study phase took the form of a lexical decision task with no mention made of the upcoming memory test. Thus, this experiment was the only one in this series to use incidental learning. The 120 study words were intermingled in a different random order for each participant. Each item remained on the screen until the participant had read it (aloud or silently) and made a lexical decision response, pressing the" m" key for words or the "c" key for nonwords. There was a 500-ms blank screen between successive items. An experimenter was present to ensure that participants spoke only the blue items aloud.

During test, a unique randomisation was created for each participant, with the 120 old items and the 60 new items intermingled for a yes/no recognition test. Each test item stayed on the screen until the participant pressed either the "m" key to indicate that an item (word or nonword) had been studied or the "c" key to indicate that an item (word or nonword) had not been studied. There was a 500-ms blank between successive test items.

Results & discussion

The proportions of "yes" responses on the recognition test are shown in Table 3. The false alarm rates did not differ reliably for word distractors and nonword distractors, $t(20) = 1.032$, $p = .314$, Cohen's d = .225.

Table 3. Experiment 3 (nonword-word): Mean proportion of "yes" responses in each condition.

	Hit Rate		False Alarm Rate
Lexical Status	Aloud	Silent	New
Word	.791	.694	.321
	(.035)	(.035)	(.042)
Nonword	.622	.553	.279
	(.035)	(.035)	(.037)

Note: Standard errors are shown in parentheses below their respective means.

A 2 × 2 repeated measures ANOVA was carried out on corrected recognition (hits minus false alarms), the two within-subject factors being Item Type (word vs. nonword) and Encoding Mode (aloud vs. silent). The main effect of Item Type was significant, $F(1, 20) = 13.855$, $MSE = 0.019$, $p = .001$, partial $\eta^2 = .409$, with accuracy considerably higher for words ($M = .421$, $SE = .034$) than for nonwords ($M = .309$, $SE = .034$), as would be expected. There was also a significant main effect of Encoding Mode, $F(1, 20) = 13.103$, $MSE = 0.011$, $p = .002$, partial $\eta^2 = .396$, with items read aloud ($M = .407$, $SE = .033$) better recognised than those read silently ($M = .324$, $SE = .033$). The nonsignificant interaction, $F(1, 20) = 1.031$, $MSE = 0.004$, $p = .332$, partial $\eta^2 = .049$, indicated that the production effect (Aloud – Silent) was present to a similar extent both for words ($M = .097$; 95% CI [.040, .154]) and for nonwords ($M = .069$; 95% CI [.015, .123]).

As in Experiments 1 and 2, both types of items – in this case, words and nonwords – showed reliable production effects, and those effects did not differ in magnitude. These results fit well with the idea that produced items benefit from enhanced distinctiveness, with recollection of having been said aloud providing confirmation that an item was studied. The production effect also clearly endured in incidental learning conditions; it was not eliminated when participants were less actively encoding the study context. This finding suggests that production information, by virtue of its distinctiveness, may be routinely encoded: Participants clearly do not need to intentionally choose to encode aloud information for it to be retained in their record of encoding.

Power analyses for the three experiments

As mentioned in the Participants section of Experiment 1, our sample sizes were selected to closely match those of prior experiments that have explored the production effect. However, those experiments primarily investigated main effects, not interactions. Unfortunately, power analyses are not well-developed to estimate sample sizes for interactions – and even less so for interactions in within-subject designs. Although we did not calculate power estimates before carrying out these experiments, here we provide estimates of power to detect interactions in each experiment. We used PANGEA (https://jakewestfall.shinyapps.io/pangea/) for this purpose.

For Experiment 1, our observed partial η^2 was .031 so our power level, given our sample size, was .23. With our design, we had an 80% chance of detecting effects that had partial $\eta^2 \geq .29$. PANGEA indicates that to have sufficient power (> .80) to detect an interaction with our observed partial η^2 would have required 105 participants. For Experiment 2, our observed partial η^2 was .018, so our power level was .14. With our design, we had an 80% chance of detecting effects that had partial $\eta^2 \geq .30$.

PANGEA indicates that to have sufficient power (> .80) to detect an interaction with our observed partial η^2 would have required 145 participants per between-subjects condition. For Experiment 3, our observed partial η^2 was .049, so our power level was .33. With our design, we had an 80% chance of detecting effects that had partial $\eta^2 \geq .29$. PANGEA indicates that to have sufficient power (> .80) to detect an interaction with our observed partial η^2 would have required 66 participants.

Although the chance of detecting a significant interaction was, therefore, limited in each individual experiment, given that we tried three times, the chance of detecting a significant interaction (if one exists) in at least one of our three experiments was .56. What we can say overall, though, is that even if there was an interaction, it would have to be quite small compared to the main effects. As noted, our sample sizes were selected based on past studies, which have routinely relied on samples of no more than 20–30 subjects, and have all easily demonstrated within-subject production effects.

General discussion

This set of experiments complements and extends the existing published work on the production effect (see MacLeod & Bodner, 2017). The effect is a robust one, as prior studies and those reported here demonstrate, so it is important to understand its boundary conditions and to refine its explanation. These were the goals of the experiments reported in this article.

Forrin et al. (2014) demonstrated that items produced aloud resulted in the same benefit whether the encoding task was a more elaborative one, like generation or imagery, or a less elaborative one, like simply reading or picturing the letters of a word. Here, we have reported a series of three converging studies focusing on material type. In these experiments, we have demonstrated that the boost from producing some items aloud does not appear to depend on other factors that certainly do influence memory in their own right. The production effect was reliable and comparable for high versus low frequency words, for words versus pictures, and for words versus nonwords, three material manipulations that powerfully affect memory. The present findings support the argument that the production effect adds to memorability through the encoding at study of a set of features that can then be used at test to support a judgment that the item was in fact studied. That participants are making such judgments has received direct support in the literature (e.g., Ozubko et al., 2020). This is the essence of the distinctiveness account of the effect (see MacLeod & Bodner, 2017, for a brief review).

Experiment 2 switched test procedure from recognition to recall, demonstrating again the robustness of the production effect. These experiments also confirm that the production effect is present whether learning is intentional (Experiments 1 and 2; as in MacLeod et al., 2010, for example) or incidental (Experiment 3; as in MacDonald & MacLeod, 1998). Clearly, it is not necessary to try to encode that an item was produced during study for produced items to be recognised better than silently read items. Instead, the results of Experiment 3 are consistent with the idea that distinctive "I said it aloud" information is routinely encoded, perhaps even automatically (cf. frequency encoding in Hasher & Zacks, 1979). This distinctive information then becomes useful at the time of test. The Ozubko et al. (2012a) finding that produced words have a familiarity advantage (in addition to a recollection advantage) over unproduced words in mixed-list experiments fits with this view of distinctiveness as being encoded very customarily.

The results of these experiments are consistent with the explanation of the production effect that we have been advocating – that production confers distinctiveness on words at study. Aloud information retains its distinctiveness at test, providing a basis for confirming previous study. This distinctiveness account was presented in detail in MacLeod et al. (2010), and is consistent with Conway and Gathercole's (1987) earlier proposal. The idea of distinctiveness is well connected to the empirical memory literature (see, e.g., Dodson & Schacter, 2001; Gallo et al., 2004; Hunt, 2003; Hunt & Elliot, 1980; Hunt & Mitchell, 1982; Waddill & McDaniel, 1998) and to theorising about memory (e.g., Hunt, 2006; Schmidt, 1991).

There are, of course, other accounts of the production effect – two in particular have been set out. The strength account (see, e.g., Bodner & Taikh, 2012) maintains, simply, that production makes the produced items stronger. As mentioned in the introduction, both strength and distinctiveness might be operative in a mixed-list, within-subject design whereas only strength might be operative in a pure-list, between-subjects design (see MacLeod & Bodner, 2017, for discussion of this idea). But when Ozubko et al. (2014) equated the strength of aloud and silent items by repeating silent items during study – which led to equivalent item recognition – participants were still adept at recalling whether an item had been studied aloud or silently, consistent with "aloudness" being recorded in memory, as proposed by the distinctiveness account.

The other explanation is the order account (see, e.g., Jonker et al., 2014) – that presenting both aloud and silent items in the mixed-list procedure disrupts the storage of order information that would ordinarily occur for the silent items in a pure-list procedure, thereby reducing memory for the silent items and resulting in an advantage for the aloud items. This idea is closely connected to distinctiveness: By virtue of being distinctive against the backdrop of the silent items in a mixed list, the aloud items prevent the encoding of the order information for the silent items, amounting to a two-process explanation.

What would these two accounts predict regarding the present manipulations? Regarding the strength account, it is difficult to derive precise predictions because it is a very general account. However, one challenge that we see for a strength account would be to explain how both highly memorable items and less memorable items can benefit to the same extent from production. It would seem as though items that are already high in memory strength would benefit less from further attempts to increase strength than would items low in memory strength. Of course, a strength model may exist to reconcile this pattern but in our view this seems to challenge the basic strength account at this time.

As for the item order account, although it would appear to be able to handle the results reported here, it does so by invoking two processes: the routine encoding of order information and the disruption of that routine encoding by the distinctiveness of the produced items. Consequently, we see the distinctiveness account as simpler and more parsimonious.

In the end, having produced a piece of information makes it memorable. By our view, a produced item becomes distinctive against the background of items that were not produced, and that distinctiveness can later be used as diagnostic regarding whether a piece of information has been experienced. The boost imparted by production is quite consistent across different types of encoding, across different study and test procedures, and now across different types of material. In this sense, production makes a reliable and consistent contribution, entirely in keeping with the modelling of distinctiveness as involving the encoding of additional features stemming from speech (Jamieson et al., 2016; Kelly et al., 2022; Saint-Aubin et al., 2021). The production effect operates just as many other mnemonic techniques do to benefit memory: It adds information that enriches encoding, and it consequently enhances retrieval.

Notes

1. After reporting all three experiments, and just prior to the General Discussion, we will consider the issue of the power in these experiments in more detail.
2. In many previous production experiments (e.g., MacLeod et al., 2010; Hourihan & MacLeod, 2008; Ozubko et al., 2012a), 2/3 of the items on the recognition test have been studied and 1/3 have been new. This makes the number of aloud, silent, and new items equal at test, but it also produces an unequal number of old and new items. We do know, though, from observing the benefit in recall experiments (Conway & Gathercole, 1987; Lin & MacLeod, 2012; Experiment 2 here) and in other recognition experiments in which the numbers of old and new items were equal (e.g., Forrin et al., Experiment 2, 2012) that this choice does not matter. Nor does counterbalancing the colours that signal aloud versus silent reading matter (see Lin & MacLeod, 2012). The extension of the production benefit to a fill-in-the-blanks test in Ozubko et al. (2012b, Experiment 3) further shows that the production effect does not rely on certain specific procedural features either during study or during test.

Acknowledgements

We thank Jennifer Allman, Emily Bryntwick, Grace Hsiao, Molly Pottruff, and Jonathan Zimmermann for their assistance with the data collection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by Discovery Grant A7459 from the Natural Sciences and Engineering Research Council of Canada.

ORCID

Colin M. MacLeod http://orcid.org/0000-0002-8350-7362
Jason D. Ozubko http://orcid.org/0000-0003-0351-0957

References

Barlow, M. C. (1928). The rôle of articulation in memorizing. *Journal of Experimental Psychology*, *11*(4), 306–312. https://doi.org/10.1037/h0071449

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1711–1719. https://doi.org/10.1037/a0028466

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*(3), 341–361. https://doi.org/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*(1), 155–161. https://doi.org/10.3758/BF03196152

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

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology*, *70*(2), 99–115. https://doi.org/10.1037/cep0000089

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory (Hove, England)*, *20*(7), 655–666. https://doi.org/10.1080/09658211.2012.693510

Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2014). Production improves memory equivalently following elaborative vs. Non-elaborative processing. *Memory (Hove, England)*, *22*(5), 470–480. https://doi.org/10.1080/09658211.2013.798417

Forrin, N. D., & MacLeod, C. M. (2018). This time it's personal: The memory benefit of hearing oneself. *Memory (Hove, England)*, *26*(4), 574–579. https://doi.org/10.1080/09658211.2017.1383434

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*(7), 1046–1055. https://doi.org/10.3758/s13421-012-0210-8

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language*, *51*(3), 473–493. https://doi.org/10.1016/j.jml.2004.06.002

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*(4), 309–313. https://doi.org/10.3758/BF03197041

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*(2), 110–119. https://doi.org/10.3758/BF03213478

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356–388. https://doi.org/10.1037/0096-3445.108.3.356

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, *11*(4), 534–537. https://doi.org/10.1016/S0022-5371(72)80036-7

Hourihan, K. L., & Churchill, L. A. (2020). Production of picture names improves picture recognition. *Canadian Journal of Experimental Psychology*, *74*(1), 35–43. https://doi.org/10.1037/cep0000185

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*(4), 242–246. https://doi.org/10.1037/1196-1961.62.4.242

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

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). Oxford University Press.

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, *22*(1), 10–15. https://doi.org/10.1177/0963721412463228

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, *109*(1), 49–74. https://doi.org/10.1037/0096-3445.109.1.49

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(1), 81–87. https://doi.org/10.1037/0278-7393.8.1.81

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology*, *70*(2), 154–164. https://doi.org/10.1037/cep0000081

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 441–448. https://doi.org/10.1037/a0034977

Kelly, M., Ensor, T. M., Lu, X., MacLeod, C. M., & Risko, E. F. (2022). Reducing retrieval time modulates the production effect: Empirical evidence and computational accounts. *Journal of Memory and Language*, *123*, 104299. https://doi.org/10.1016/j.jml.2021.104299

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

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*(3), 212–216. https://doi.org/10.1037/a0028309

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*(2-3), 291–310. https://doi.org/10.1016/S0001-6918(97)00047-4

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*(6), 1197–1202. https://doi.org/10.3758/s13423-011-0168-8

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*(4), 390–395. https://doi.org/10.1177/0963721417691356

MacLeod, C. M., & Daniels, K. A. (2000). Direct vs. indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review*, *7*(2), 354–359. https://doi.org/10.3758/BF03212993

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*(3), 671–685. https://doi.org/10.1037/a0018785

MacLeod, C. M., & Kampe, K. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 132–142. https://doi.org/10.1037/0278-7393.22.1.132

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 523–528. https://doi.org/10.1037/0278-7393.2.5.523

Ozubko, J. D., Bamburoski, L. D., Carlin, K., & Fawcett, J. M. (2020). Distinctive encodings and the production effect: Failure to retrieve distinctive encodings decreases recollection of silent items. *Memory (Hove, England)*, *28*(2), 237–260. https://doi.org/10.1080/09658211.2019.1711128

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012a). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012b). Production benefits learning: The production effect endures and improves memory for text. *Memory (Hove, England)*, *20*(7), 717–727. https://doi.org/10.1080/09658211.2012.699070

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory (Hove, England)*, *22*(5), 470–480. https://doi.org/10.1080/09658211.2013.800554

Putnam, A. L., Ozubko, J. D., MacLeod, C. M., & Roediger III, H. L. (2014). The production effect in paired-associate learning: Benefits for item and associative information. *Memory & Cognition*, *42*(3), 409–420. https://doi.org/10.3758/s13421-013-0374-x

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*(1), 89–102. https://doi.org/10.3758/BF03211168

Rommers, J., Dell, G. S., & Benjamin, A. S. (2020). Word predictability blurs the lines between production and comprehension: Evidence from the production effect in memory. *Cognition*, *198*, 104206. https://doi.org/10.1016/j.cognition.2020.104206

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language*, *118*, 104219. https://doi.org/10.1016/j.jml.2021.104219

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, *19*(6), 523–542. https://doi.org/10.3758/BF03197149

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621–633. https://doi.org/10.3758/BF03195542

Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition*, *26*(1), 108–120. https://doi.org/10.3758/BF03211374

Weldon, M. S., Roediger, H. L., & Challis, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition*, *17*(1), 95–105. https://doi.org/10.3758/BF03199561

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*(6), 747–763. https://doi.org/10.3758/BF03211318

Zhou, Y., & MacLeod, C. M. (2021). Production between and within: Distinctiveness and the relative magnitude of the production effect. *Memory (Hove, England)*, *29*(2), 168–179. https://doi.org/10.1080/09658211.2020.1868526