Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale
2016, Vol. 70, No. 2, 93–98

1196-1961/16/$12.00 http://dx.doi.org/10.1037/cep0000082

The Production Effect in Recognition Memory: Weakening Strength Can Strengthen Distinctiveness

Glen E. Bodner
University of Calgary

Randall K. Jamieson and David T. Cormack
University of Manitoba

Dawn-Leah McDonald and Daniel M. Bernstein
Kwantlen Polytechnic University

Producing items (e.g., by saying them aloud or typing them) can improve recognition memory. To evaluate whether production increases item distinctiveness and/or memory strength we compared this effect as a function of the percentage of items that participants typed at encoding (i.e., 0%, 20%, 50%, 80%, and 100%). Experiment 1 revealed a strength-based pattern: The production effect was similar across pure-list (i.e., 0% vs. 100%) and mixed-list (i.e., 20%, 50%, 80%) designs, and there was no observed influence of statistical distinctiveness (i.e., 20% vs. 80%). In Experiment 2, we increased the study time for unproduced items to minimise the strength difference between produced and unproduced items. The manipulation attenuated the pure-list effect without eliminating the mixed-list effect, providing support for the inference that the mixed-list effect reflects distinctiveness. An influence of statistical distinctiveness also emerged: The mixed-list effect was larger when participants produced only 20%, rather than 80%, of the items. These findings suggest that both strength and distinctiveness contribute to the production effect in recognition.

Keywords: production effect, recognition, percent produced, distinctiveness, memory strength

If you want to learn a list of words, then reading them silently is a poor tactic. Much research, spurred by MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010), has suggested that producing words yields superior retention, whether productions are vocal, including silent mouthing and singing (e.g., MacLeod et al., 2010; Quinlan & Taylor, 2013), or nonvocal, including writing and typing (e.g., Forrin, MacLeod, & Ozubko, 2012; Jamieson & Spear, 2014). Researchers have mapped out conditions that can increase, decrease, attenuate, or even reverse the production effect. These maps are useful for constraining theories of why production works (when it works) and why it fails (when it fails). Largely, this evidence has been taken as support for what we term a *distinctiveness influence*, according to which producing an item at study adds something distinctive to its encoding that can later enhance memory for its occurrence. We next provide a few examples of how most current evidence is also consistent with what we term a *strength influence*, according to which producing an item at study improves memory by enhancing the overall strength of its representation in memory. We then report evidence that distinctiveness and strength can both influence recognition, and furthermore, that reliance on one influence might reduce reliance on the other.

MacLeod et al. (2010; see also Hopkins & Edwards, 1972) considered both distinctiveness and strength as sources for the production effect, but two of their findings favoured distinctiveness. First, production enhanced recognition in a *mixed-list design* (in which half the words were produced and half were not) but not in a *pure-list design* (in which one group produced 100% of the words and another group produced 0%). This pattern fit well with Hunt's (2006) definition of distinctiveness, which states that one class of items (e.g., produced) can become distinctive only when set against a backdrop of a second class of items at study (e.g., unproduced). However, subsequent experiments and meta-analyses have revealed pure-list effects (e.g., Bodner, Taikh, & Fawcett, 2014; Fawcett, 2013). Second, MacLeod et al. (2010) found null effects of production on an implicit speed-reading test in which distinctiveness was not expected to have an influence. However, Bodner and Taikh (2012) argued that memory strength would not be expected to have an influence on this implicit test because participants do not need to evaluate memory strength when reading test words aloud. Evidence for a distinctiveness influence obtained using a list-discrimination task (Ozubko & MacLeod, 2010) has also since been challenged (Bodner & Taikh, 2012).

Other effects of production attributed to distinctiveness could instead be attributed to strength. For example, production can enhance both recollection and familiarity (Ozubko, Gopie, & MacLeod, 2012). Although the recollection effect may be due to

Glen E. Bodner, Department of Psychology, University of Calgary; Randall K. Jamieson and David T. Cormack, Department of Psychology, University of Manitoba; Dawn-Leah McDonald and Daniel M. Bernstein, Department of Psychology, Kwantlen Polytechnic University.

This research was supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada to G. E. Bodner (RGPIN 238599-2015) and R. K. Jamieson (RGPIN 355882-2013), and by Canada Research Chair 950-228407 to D. M. Bernstein.

Correspondence concerning this article should be addressed to Glen E. Bodner, Department of Psychology, University of Calgary, 2500 University Drive, Northwest, Calgary AB T2N 1N4, Canada. E-mail: bodner@ucalgary.ca

enhanced distinctiveness, the familiarity effect could be due to enhanced strength (Fawcett & Ozubko, 2016). That older adults show a reduced production effect has been taken as evidence that aging reduces the distinctiveness influence (Lin & MacLeod, 2012), but perhaps it also reduces the strength influence. Quinlan and Taylor (2013) found that singing improved recognition over saying words aloud, which in turn was better than silent reading. Singing may be more distinctive than aloud production, but it might also strengthen memory traces.

Two recent studies have provided more persuasive evidence for a distinctiveness influence. Icht, Mama, and Algom (2014) tested whether the production effect varied as a function of the *percent produced*. Groups said 20%, 50%, or 80% of the words aloud on the study list (the rest were read silently) prior to recall and recognition tests, in turn. In recognition, the production effect was greatest in the group that produced only 20% of the words (hence, when production was the rarer encoding event and was "statistically distinctive"), and it disappeared in the group that produced 80% of the words (hence, when not producing an item was the more statistically distinctive type of encoding event). Indeed, recall in the 80% group was actually greater for silent items than for aloud items. A strength influence, conversely, predicted equivalent effects for the three mixed-list groups on each test. Icht et al.'s (2014) findings highlighted the utility of separating encoding distinctiveness (e.g., the contents of one's processing) from statistical distinctiveness (e.g., the rarity, hence informational value of one's processing).

Ozubko, Major, and MacLeod (2014) also contrasted predictions based on distinctiveness versus strength. Their tack was to first strengthen some of the silent items at study by presenting them twice, such that overall recognition was equivalent for twice-silent items and aloud items. Participants then completed a study-mode test in which they had to classify items as aloud, silent, or new. Study-mode judgments were very accurate for aloud items. Critically, twice-silent items tended to be attributed to the silent-study mode, suggesting that participants evaluated the distinctiveness of their encoding on this task, and not just their (equated) strength. A strength influence incorrectly predicted that participants would be unable to distinguish the mode of the twice-studied silent items from the aloud items, given their equivalent strength.

The experiments presented here were conducted prior to the publication of Icht et al. (2014) and Ozubko et al. (2014), but share features with each study. Like Icht et al. (2014), we used a percent-produced manipulation. Important to note is, unlike Icht et al., we also included pure-list groups (i.e., 100% vs. 0% produced), and we tested recognition memory uncontaminated by a prior recall test. There is ample evidence that a pure-list production effect is absent in recall (Forrin & MacLeod, 2016; Jones & Pyc, 2014; Jonker, Levene, & MacLeod, 2014; Lambert, Bodner, & Taikh, 2016), suggesting that the influences of production do not influence recall. Moreover, the mixed-list production effect in these studies has always included a cost to recall for the unproduced items, which neither distinctiveness nor strength influences predict. Instead, the production effect in these studies has typically been ascribed to an item-order account (e.g., Jonker et al., 2014; see McDaniel & Bugg, 2008, for a review). Thus, the initial recall test in Icht et al. (2014) may have biased participants away from focusing on the strength of their recognition experiences. By testing only recognition, and by including pure-list groups, we predicted that we could specify the contributions of distinctiveness and strength to recognition. Finally, as described below, like Ozubko et al. (2014), our second experiment aimed to attenuate the strength difference between produced and unproduced items (i.e., weakening strength), to test whether a distinctiveness influence might then emerge (i.e., strengthening distinctiveness).

Experiment 1: Evidence for Strength

In Experiment 1, we varied the percentage of words that participants produced at study (i.e., typed) across five groups: 0%, 20%, 50%, 80%, and 100%, and then tested them in an "old"/"new" recognition test. This design tested two opposing predictions of distinctiveness and strength. First, a distinctiveness influence, based on Hunt's (2006) definition, does not predict a pure-list production effect (i.e., 0% vs. 100% group). In contrast, a strength influence predicts a pure-list effect, given that produced words should yield stronger memory traces than unproduced words. Second, following Icht et al. (2014), a distinctiveness influence should predict a larger mixed-list effect in the 20% group, in which produced words would be relatively more distinct, in a statistical sense, than in the 80% group, In which produced words would be relatively less distinct, in a statistical sense. A strength influence should predict similar mixed-list production effects in these two groups because memory traces should be stronger for produced words than for unproduced words, regardless of the percent produced.

Method

Participants. The participants were 130 undergraduates from the University of Manitoba participant pool who reported being capable typists. They were assigned randomly to the five groups, resulting in 26 per group.

Materials. Each participant was assigned a random sample of 100 of the 120 words listed in the appendix of MacDonald and MacLeod (1998). Of the 100 words, 50 were to be studied and 50 were foils on the recognition test. Of the 50 studied words, 0, 10, 25, 40, or 50 were produced, and the rest were unproduced. Words appeared in capital letters in 40-point Arial font. At study, produced words appeared in green and unproduced words appeared in red. At test, all words appeared in black.

Procedure. Participants were tested in small groups on computers. They were told that they would study words presented in red and/or green for a later memory test. They were asked to type only the green words and to keep their fingers on the keyboard during the study phase. The ratio of red to green words was not mentioned. Each word on the study list appeared for 3 s, with a 1-s blank interval between words. On production trials, keystrokes were recorded, but not echoed, to the screen. Following this study phase, the test instructions informed participants that 100 words would appear sequentially: 50 "old" (studied) and 50 "new" (not studied). Their task was to identify each word as studied or unstudied by clicking the tick box on the screen labelled "old" or "new" and then clicking the "OK" button, after which a 1-s blank screen came up.

Results and Discussion

Figure 1 shows the mean percentage of "old" responses for produced and unproduced words (i.e., hits) and foils (i.e., false alarms). The figure highlights our two main findings: The production effect was equivalent in the mixed-list and pure-list designs, and a similar mixed-list production effect occurred in the 20% and 80% groups. Both findings are more consistent with a strength influence than a distinctiveness influence (cf. Icht et al., 2014).

Analyses supported these impressions. There was a 9% pure-list production effect in hits advantaging the 100% group over the 0% group (73.9% vs. 64.9%), $t(50) = 2.91$, $p < .01$; the false-alarm rate was correspondingly lower for the 100% group than for the 0% group (8.3% vs. 16.7%), $t(50) = 3.00$, $p < .01$. There was a 14.2% mixed-list production advantage (76.1% vs. 61.9%) in hits for produced over unproduced words across the mixed-list groups overall, $F(1, 75) = 64.29$, $p < .001$. Contrary to a statistical distinctiveness influence, there was no convincing evidence that the mixed-list effect was larger in the 20% group than in the 80% group (15.2% vs. 11.1%), $F(1, 75) = 0.91$, $p = .34$.

Application of the Erlebacher (1977) analysis enabled statistical comparison of the mixed-list and pure-list effects. These analyses were completed with R (Version 3.1.3, 2015) using the code implemented by Merritt, Cook, and Wang (2014). Three 2 (production: unproduced vs. produced) × 2 (design: mixed vs. pure) ANOVAs showed that the pure-list production effect was similar in size to the mixed-list production effect in each of the 20%, 50%, and 80% groups, $F(1, 50) = 1.62$, $p > .20$, $F(1, 50) = 2.77$, $p > .10$, and $F(1, 50) = 0.21$, $p > .60$. In summary, the pure- and mixed-list production effects were similar and no evidence for a distinctiveness influence was obtained.

Experiment 2: Evidence for Distinctiveness

Experiment 1 supported a strength account of the production effect in recognition. Experiment 2 was an examination of whether a distinctiveness influence might emerge if we minimised the difference in memory strength for produced and unproduced words. The mixed-list groups in Experiment 1 may have used differences in strength for produced and unproduced items to guide their recognition judgments, rather than trying to recollect whether they had typed a word (cf. Ozubko et al., 2014). In Experiment 2, we worked to attenuate this hypothesised strength difference by increasing the presentation duration for unproduced words, in an effort to bring their recognition up to the level for produced words. Given other evidence that the pure-list effect reflects strength (Fawcett & Ozubko, 2016), we reasoned that attenuating the pure-list effect would signal that the strength difference had been minimised, leaving only the distinctiveness influence to yield a mixed-list effect. Therefore, our key questions were: (a) Does the mixed-list effect persist when strength differences are minimised and if so, (b) is the mixed-list effect larger when production is more distinctive (i.e., in the 20% vs. 80% group)?

Method

The Experiment-1 method was followed, except that unproduced words were now presented for 9 s, whereas produced words were again presented for 3 s, thus intentionally confounding production with study duration. An additional 120 participants from the same pool were assigned randomly to the five groups, resulting in 24 per group.

Results and Discussion

As shown in Figure 2, tripling the study duration for unproduced words attenuated the pure-list production effect from Experiment 1. The hit rates for produced words in the 100% group and for unproduced words in the 0% group were nearly identical (71.9% vs. 72.4%), $t(46) = 0.11$, $p = .91$, and their respective false-alarm rates (12.0% vs. 14.6%) did not differ, $t(46) = 0.93$, $p = .36$. A 2 (group: 0% vs. 100%) × 2 (measure: hits vs. false alarms) × 2 (Experiment: 1 vs. 2) ANOVA revealed a significant three-way

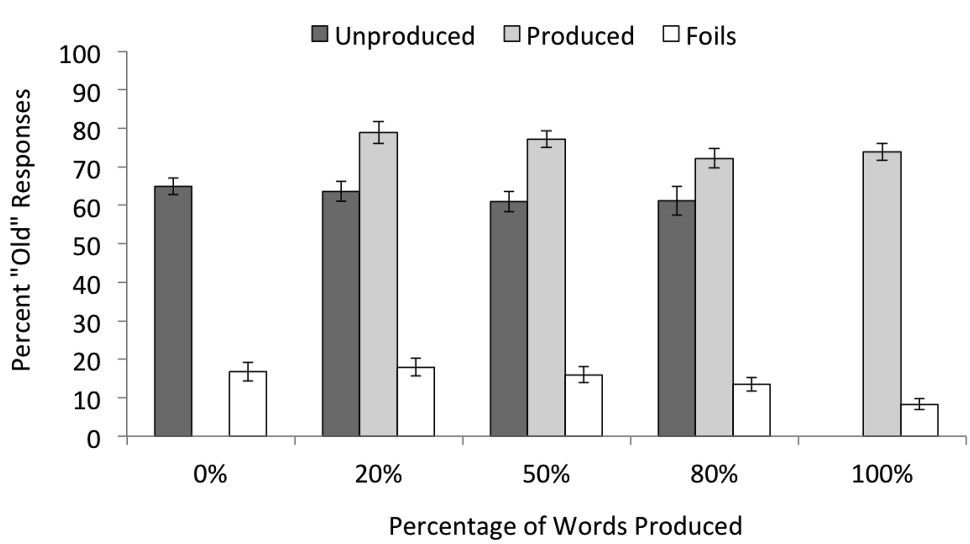

Figure 1. Experiment 1 (3 s per unproduced word): mean (*SE*) percent "old" responses for produced, unproduced, and foil items as a function of the percentage of words produced.

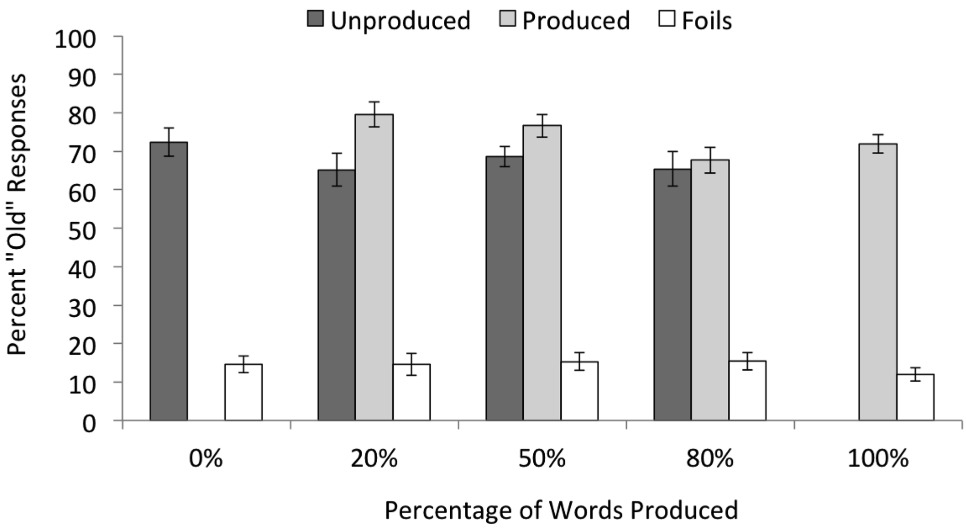

Figure 2. Experiment 2 (9 s per unproduced word): mean (*SE*) percent "old" responses for produced, unproduced, and foil items as a function of the percentage of words produced.

interaction, $F(1, 96) = 6.65$, $p < .001$, confirming that the pattern of increased hits and decreased false alarms in the 100% group relative to the 0% group was statistically smaller in Experiment 2 than in Experiment 1.

Assuming that strength differences drive the pure-list effect (see Fawcett & Ozubko, 2016), the strength basis for a mixed-list effect was thus attenuated. Yet, across the mixed-list groups, an 8.3% hit-rate advantage occurred for produced over unproduced words (74.7% vs. 66.4%), $F(1, 69) = 10.25$, $p < .01$. Attenuation of the strength influence also allowed an influence of statistical distinctiveness to emerge. Specifically, the mixed-list production effect was significant in the 20% group (79.6% vs. 65.2%), $t(23) = 2.72$, $p < .05$, but not in the 80% group (67.7% vs. 65.4%), $t(23) = 0.46$, $p > .65$. This pattern replicated Icht et al. (2014), although our interaction was not quite significant, $F(1, 69) = 3.69$, $p < .06$.

Erlebacher-method analyses confirmed that the 14.4% production effect on hits in the 20% group was larger than the 0% effect in the 0% group, $F(1, 50) = 4.94$, $p < .03$, whereas the production effects in the 50% and 80% groups did not differ from the 0% group, $F(1, 50) = 2.23$, $p > .10$, and $F(1, 50) = 0.16$, $p > .60$. Following Bodner et al. (2014), we followed up the interaction involving the 20% group to assess whether the mixed-list effect reflected a benefit for produced items (relative to the 100% group) and/or a cost to unproduced items (relative to the 0% group). The 7.7% benefit in produced hits in the 20% group relative to the 100% group was marginally significant, $t(46) = 1.92$, $p = .06$, whereas the 7.2% cost in unproduced hits in the 20% group relative to the 0% group was not significant, $t(46) = 1.28$, $p = .21$. A distinctiveness influence predicts a benefit for produced items in a mixed list, and does not predict a cost for unproduced items in a mixed list. The cost–benefit analyses were roughly consistent with this predicted pattern.

General Discussion

Our findings highlight two nested points. First, recognition decisions can be influenced by both an item's memory strength and a memory for specific and distinctive aspects of its encoding; the two influences are not mutually exclusive. Second, characteristics of the encoding and retrieval situations likely dictate which of these influences are in play. We showed the flexibility of these influences on the production effect. In Experiment 1, typing words improved later recognition relative only to reading them, and this production effect was similar whether participants typed some (20%, 50%, 80%) or all (100%) of the words. Consistent with a strength influence, the mixed-list and pure-list production effects were similar, as was the mixed-list effect whether production was statistically more or less distinctive (20% vs. 80%).

In Experiment 2, we attempted to rob produced words of their apparently greater strength by presenting the unproduced words for a duration three times as long as in Experiment 1 (9 s vs. 3 s). We then obtained influences of distinctiveness on recognition. First, the pure-list production effect disappeared, but the mixed-list effect remained, consistent with the claim that production is distinctive only when other items are not produced (e.g., Hunt, 2006; MacLeod et al., 2010). Second, the mixed-list effect was significant when participants typed 20% of the words, and hence when production was more distinctive in a statistical sense (see Icht et al., 2014), but not when they typed 80% of them. Therefore, when we attenuated the strength difference between produced and unproduced items, evidence emerged for MacLeod et al.'s (2010) claim that production improves memory by increasing item distinctiveness in memory.

We propose that strengthening the unproduced items in Experiment 2 attenuated the use of strength as the basis for recognition decisions. Although this proposal meshes with our results, reliance on strength might have been expected to *increase* in Experiment 2, because both produced and unproduced items were now much stronger (and similarly stronger) than foil items. If reliance on strength had increased, then although the hit rate for produced and unproduced items would exceed the false-alarm rate, as we observed, the mixed-list production effect should have been absent, unlike what we observed.

It appears that participants in Experiment 2 did not capitalize on the overall strength difference between studied items and foil items. This finding provides an intriguing counterpoint to Ozubko et al. (2014), who required participants to make study-mode judgments at test under the assumption that doing so would force them to try to recollect study-mode information, thus yielding a distinctiveness influence. Our results suggest that participants may choose to evaluate distinctiveness even when study-mode judgments are not required—a possibility that warrants further testing. For example, by Ozubko et al.'s claim, the mixed-list effect should be larger when recognition judgments are followed by study-mode judgments than when they are not. Further supporting a distinctiveness influence, the mixed-list effect in the 20% group was greater than the pure-list effect, as shown using the Erlebacher (1977) analysis, and resulted more from a benefit for produced items than from a cost for unproduced items.

Our percent-produced manipulation did not fully replicate Icht et al. (2014). Their production effect was boosted when participants produced only 20% of the words; when participants produced 80% of the words, the production effect disappeared in recognition (and reversed in recall). Conversely, our percent-produced manipulation did not modulate the mixed-list production effect in Experiment 1. As reviewed in our introduction, the use of an initial recall test in Icht et al. may have biased participants away from using strength to guide their recognition judgments. This possibility also awaits further testing. Our Experiment-2 results were more consistent with Icht et al.'s recognition results: The production effect was significant in the 20% group, but not in the 80% group. However, these findings emerged only when we attenuated the strength difference between produced and unproduced items, unlike in Icht et al. (2014).

Finally, Icht et al. (2014) did not include pure-list groups, which prevented them from being able to examine the cost–benefit basis of their mixed-list production effects. As it stands, it remains unclear whether their design would yield a pure-list effect, and if it does, whether that pure-list effect would most resemble the mixed-list effect in their 20%, 50%, or 80% group. Our study confirms that including these pure-list groups can help researchers gauge the influences of distinctiveness versus strength on the production effect in memory.

Résumé

La production d'éléments (en les prononçant à haute voix ou en les tapant) peut améliorer la mémoire de reconnaissance. Pour déterminer si la production augmente la distinctivité d'un élément et/ou la force de la mémoire, nous avons comparé cet effet en fonction du pourcentage d'éléments que les participants ont tapé à l'encodage (c.-à-d., 0 %, 20 %, 50 %, 80 % et 100 %). L'Expérience 1 a révélé une tendance basée sur la force : l'effet de la production était similaire pour l'ensemble des listes pures (0 % vs 100 %) et des listes mixtes (produites à 20 %, 50 %, 80 %). Aucune influence de spécificité statistique n'a été observée (c.-à-d., 20 % vs 80 %). Dans l'Expérience 2, le temps d'étude a été accru pour les éléments non produits pour réduire la différence de la force entre les éléments produits et les éléments non produits. La manipulation a atténué l'effet de la liste pure, sans éliminer l'effet de la liste mixte, ce qui appuie l'inférence selon laquelle l'effet de la liste mixte témoigne de la distinctivité. L'influence d'une distinctivité statistique a aussi été révélée : l'effet de la liste mixte était plus grand lorsque les participants ont produit seulement 20 % des éléments, au lieu de 80 %. Ces résultats suggèrent que tant la force que la distinctivité contribuent à l'effet de la production sur le plan de la reconnaissance.

Mots-clés : effet de la production, reconnaissance, pourcentage produit, distinctivité, force de la mémoire.

References

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. http://dx.doi.org/10.1037/a0028466

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review, 21,* 149–154. http://dx.doi.org/10.3758/s13423-013-0485-1

Erlebacher, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin, 84,* 212–219. http://dx.doi.org/10.1037/0033-2909.84.2.212

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

Fawcett, J. M., & Ozubko, J. D. (2016). *Those old familiar feelings: Familiarity, but not recollection, supports the between-subject production effect*. Manuscript submitted for publication.

Forrin, N. D., MacLeod, C. M. (2016). *Order information is used to guide recall of long lists: Further evidence for the item-order account*. Manuscript submitted for publication.

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

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

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (Ch. 1, no pagination). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195169669.003.0001

Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology: Cognition, 5,* Article No. 00886. http://dx.doi.org/10.3389/fpsyg.2014.00886

Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology, 68,* 20–28. http://dx.doi.org/10.1037/cep0000009

Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 300–305. http://dx.doi.org/10.1037/a0033337

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,* 441–448. http://dx.doi.org/10.1037/a0034977

Lambert, A. M., Bodner, G. E., & Taikh, A. (2016). *The production effect in long-list recall: In no particular order?* Manuscript submitted for publication.

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. http://dx.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,* 291–310. http://dx.doi.org/10.1016/S0001-6918(97)00047-4

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671–685. http://dx.doi.org/10.1037/a0018785

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15,* 237–255. http://dx.doi.org/10.3758/PBR.15.2.237

Merritt, P. S., Cook, G. I., & Wang, M. (2014). *Erlebacher's method for contrasting the within- and between-subjects manipulation of the independent variable using R and SPSS* (Unpublished manuscript). Retrieved from https://dl.dropboxusercontent.com/u/18192026/Merritt_Cook_Wang_CSDAv1.pdf

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

Ozubko, J. D., & Macleod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1543–1547. http://dx.doi.org/10.1037/a0020604

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory, 22,* 509–524. http://dx.doi.org/10.1080/09658211.2013.800554

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory, 21,* 904–915. http://dx.doi.org/10.1080/09658211.2013.766754

Received June 30, 2015
Accepted January 28, 2016 ■