

# I said, you said: The production effect gets personal

Colin M. MacLeod

Published online: 22 September 2011  
© Psychonomic Society, Inc. 2011

**Abstract** Saying a word out loud makes it more memorable than simply reading it silently. This robust finding has been labeled the *production effect* and has been attributed to the enhanced distinctiveness of produced relative to unproduced items (MacLeod et al. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 671–685, 2010). Produced items have the additional information that they were spoken aloud encoded in their representations, and this information is useful during retrieval in certifying prior encoding. The present study explored whether production must be self-performed to be beneficial, or whether another person's production also makes an item more memorable. In two experiments, the production effect was shown to be reliable when production was done by someone other than the rememberer (i.e., by the experimenter or by another participant), but substantially smaller than the benefit from self-performed production. Intriguingly, the effect was intermediate when production was done by both the rememberer and another person. Distinctiveness—and hence the production effect—is greatest to the extent that it is personal.

**Keywords** Episodic memory · Human memory · Long-term episodic memory · Encoding effects

In accord with intuition, when an event or an element of an event stands out from its surrounding context, we remember it better. In the experimental literature, this is usually traced back to the von Restorff effect (Hunt, 1995; von Restorff, 1933)—although it in fact goes back farther (e.g.,

Calkins, 1894). Hunt (1995, 2006) has argued that this discriminability then leads to distinctive processing, and thereby results in enhanced memory. Critically, this link from discriminability to distinctive processing is not restricted to an isolated single item, but can arise whenever discrimination occurs within a set of items. Hunt (2006) makes a convincing argument for distinctiveness not as an empirical phenomenon but as a theoretical mechanism, in accord with, for example, Lockhart, Craik, and Jacoby (1976, p.86), who suggested that one of “the beneficial effects of depth of encoding is that deeper, richer encodings are also more distinctive and unique.” There is by now ample evidence that distinctiveness improves memory (Hunt & Worthen, 2006).

Recently, MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010) examined a phenomenon that they labeled the *production effect* and that they theorized—in accord with earlier proposals (Conway & Gathercole, 1987; Gathercole & Conway, 1988)—also relies on distinctive processing. Simply put, the production effect is the better memory observed for items read aloud than for items read silently. In a series of experiments, MacLeod et al. showed this to be a large and robust benefit (see also Hourihan & MacLeod, 2008; Ozubko, Gopie, & MacLeod, 2011; Ozubko & MacLeod, 2010). They argued that the mechanism underlying the effect is distinctiveness, with items produced during study having the distinctive additional element in their encoding of having been spoken aloud. People then use successful recovery of an item having been said aloud to verify that it was studied. Ozubko et al. showed, using the receiver operating characteristic and remember/know procedures, that the benefit of production during study is apparent in both recollection and familiarity.

Two key findings favor the idea that the production effect hinges on an additional dimension of encoding rather than simple strengthening of produced items. First, Hopkins

---

C. M. MacLeod (✉)  
Department of Psychology, University of Waterloo,  
Waterloo, Ontario N2L 3G1, Canada  
e-mail: cmacleod@uwaterloo.ca

and Edwards (1972), Dodson and Schacter (2001), and MacLeod et al. (2010) have all shown that the effect occurs in within-subject designs but not in between-subjects designs. This is inconsistent with a basic strength account, and suggests instead that the distinction between produced and unproduced items must be apparent during study so that it will be used during test. Second, Ozubko and MacLeod (2010) showed that if the distinctiveness of the items said aloud is interfered with in the critical mixed list by studying an additional list of items that are all read aloud, the benefit of production disappears. Again, this does not fit with a strength account, but suggests that when the additional dimension that a word was produced is rendered useless, the production effect is eliminated.

In all of these previous studies of the production effect (Conway & Gathercole, 1987; Gathercole & Conway, 1988; Hopkins & Edwards, 1972; MacDonald & MacLeod, 1998), the rememberer has done his or her own production. The central question in the present study is whether a production benefit also occurs when one hears someone else do the production, and if so, how that benefit compares to the benefit experienced for one's own productions. There is an extensive literature on the benefit of self-reference for memory (e.g., Rogers, Kuiper, & Kirker, 1977), which is consistent with the utility of the production effect for the individual actually doing the production. Quite a large literature has also shown that auditory versus visual presentation has little impact on long-term retention, whether measured by recognition (Bray & Batchelder, 1972; Kirsner, 1974; Lehman, 1982) or recall (Crowder, 1970; Hintzman, Block, & Inskip, 1972). Together, these findings could be combined to imply a first-hand production effect but no second-hand effect. Note, however, that having a word presented from an audio speaker may well be quite different from having a person sitting next to the participant read the word aloud. A soundtrack is certainly less personal; moreover, in a modality study, words are presented either auditorily or visually, whereas in a production study all words are presented visually, and some of them are also produced. There is, therefore, an additional encoding feature in a production study, but not in a modality study.

Yet the idea that saying some items aloud provides an additional dimension of encoding that makes those items distinctive does not require retrieval of details that are unique to the rememberer. Rather, what is required is simply that some discriminative information be accessed that can certify that a given item was actually studied. Thus, remembering that someone else spoke some of the items should be beneficial, too, as long as the rememberer chooses to probe his or her memory for that information. Perhaps, though, probing memory for one's own productions is the default, whereas it is not the default for another

person's productions. One may pay more attention to one's own productions (cf. MacDonald & MacLeod, 1998) but not to those of someone else. This leads to the interesting question of what would happen when the rememberer and another person both produce some of the same items. Would this provide two potential sources of distinctiveness, thereby improving memory further, or would the other person's productions interfere with the rememberer's own productions, undermining the use of this distinctive information?

To address the questions and predictions just outlined, a second person was introduced into the paradigm in the experiments reported here. In [Experiment 1](#), this was an experimenter; in [Experiment 2](#), it was another participant. In both experiments, four study conditions were randomly intermixed: only the rememberer saying a word aloud, both the rememberer and the other person saying a word aloud, only the other person saying a word aloud, or neither person saying a word aloud. This "production schedule" was implemented using screen location to indicate what the rememberer and the other person were to do on each study trial. The first and last of these conditions constituted the standard production effect, but what would happen when the other person produced a word, either alone or in addition to the rememberer?

## Method

### Participants

All participants were undergraduates from the University of Waterloo who took part for credit in their courses. In [Experiment 1](#), there were 36 participants; in [Experiment 2](#), there were 48 participants, 24 sitting to the left and 24 sitting to the right of the display.

### Materials

The stimuli were 120 nouns, from 5 to 10 letters long, with frequencies greater than 30 per million (Thorndike & Lorge, 1944); this was the same set used by MacLeod et al. (2010). For each individual in [Experiment 1](#), or each yoked pair in [Experiment 2](#), 40 words were randomly selected to form the study list, and another 40 to serve as the recognition distractors. All stimuli were presented in Courier New 18 bold font against a black background.

### Apparatus

The experiments were controlled by PC computers with 15-in. color monitors. The controlling programs were written in E-Prime (Schneider, Eschman, & Zuccolotto, 2002).

## Procedure

The instructions informed the participant(s) that the task was to study a list of individual words for a later memory test. Study words were presented in white. When a word was presented on the left of the screen, only the person on the left was to read it aloud. When a word was presented on the right, only the person on the right was to read it aloud. When a word was presented at the top, both people were to read it aloud simultaneously (they routinely succeeded in doing so at their usual volume). When a word was presented at the bottom, both people were to read it silently. Participants were reminded that, when they were not to read a word aloud, they should still always read it silently, without moving their lips. A sample screen was shown to indicate where the words could appear and what to do when a word appeared at a given location, and participants appeared to have no difficulty remembering how to respond. Each word appeared for 2,000 ms during study, with location randomized. A 1,000-ms fixation point appeared between successive words.

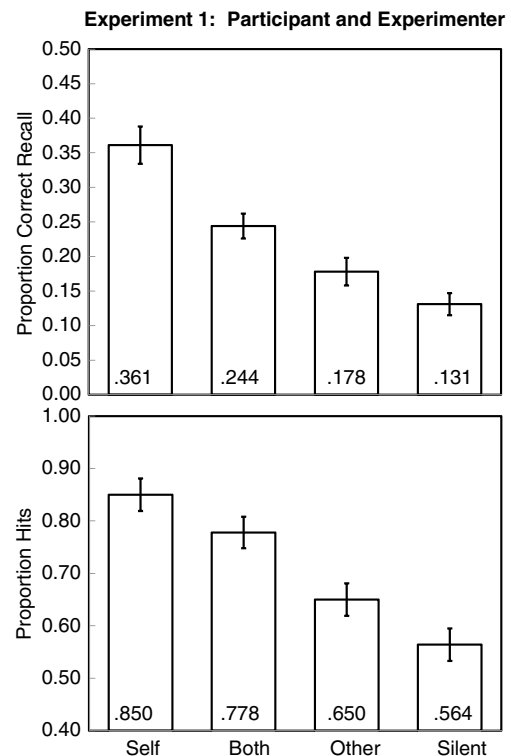
Immediately following study, participants performed a self-paced free recall test, writing any words that they recalled on a sheet of white paper. Immediately after the recall test, the participants performed a yes/no recognition test in which individual test words were presented in yellow at the center of the screen until the participant responded with a keypress indicating that the test word was previously studied (“m” key) or was previously not studied (“z” key). Each response was followed by a 500-ms white fixation point, and both the accuracy and the latency of each response were collected. For this test, the 40 studied words were randomly intermingled with 40 new distractors.

## Differences between the experiments

In **Experiment 1**, the participant sat to the right of the experimenter, so only the experimenter read aloud words on the left, and only the participant read aloud words on the right. Only the participant performed the recall and recognition tests. In **Experiment 2**, one of the participants sat on the left and read aloud words on the left, and a second participant sat on the right and read aloud words on the right. In **Experiment 2**, the participants studied the same list together but had separate recognition tests containing the same targets and distractors, using two different randomizations on two computers.

## Results

Figures 1 and 2 present the data for **Experiments 1** and **2**, respectively.



**Fig. 1** **Experiment 1** (participant and experimenter): Mean proportions of items correct in free recall (top panel) and mean proportions of hits in recognition (bottom panel), as functions of the study condition. The error bars are standard errors of their respective means

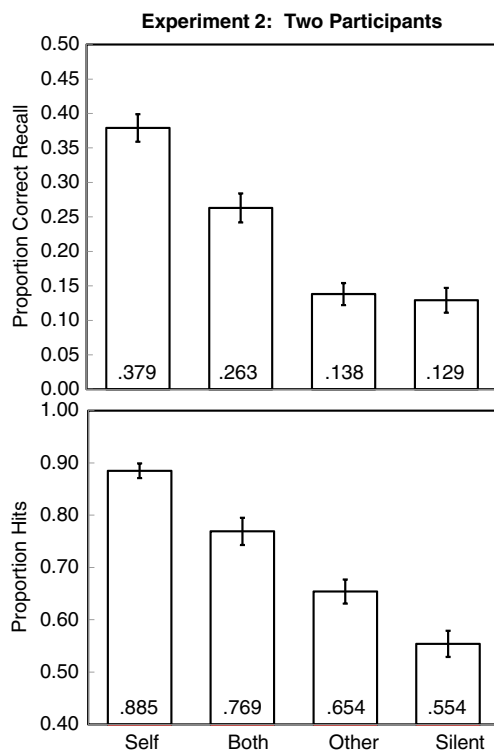
## Experiment 1

The free recall accuracy pattern was very clear. The conditions differed reliably overall,  $F(3, 105) = 21.45$ ,  $MSE = .017$ ,  $p < .001$ ,  $\eta^2 = 0.73$ . Most importantly, reliably more words were recalled in the *self* condition than in the *both* condition,  $t(35) = 3.46$ ,  $p < .001$ ; in the *both* condition than in the *other* condition,  $t(35) = 2.30$ ,  $p < .05$ ; and in the *other* condition than in the *silent* condition,  $t(35) = 1.99$ ,  $p = .05$ .

Recognition hit rates duplicated the clear free recall pattern. (The false alarm rate was .096 [ $SE = .014$ ].) The conditions differed reliably overall,  $F(3, 105) = 20.05$ ,  $MSE = .029$ ,  $p < .001$ ,  $\eta^2 = 0.36$ . Just as with recall, more words were recognized in the *self* condition than in the *both* condition,  $t(35) = 1.85$ ,  $p = .07$ , although for the only time in these experiments, this effect was marginal. As with recall, though, reliably more words were recognized in the *both* condition than in the *other* condition,  $t(35) = 2.96$ ,  $p < .01$ , and in the *other* condition than in the *silent* condition,  $t(35) = 2.34$ ,  $p < .05$ .

## Experiment 2

The free recall accuracy pattern closely followed that of Experiment 1. The conditions differed reliably overall,  $F(3,$



**Fig. 2** Experiment 2 (2 participants): Mean proportions of items correct in free recall (top panel) and mean proportions of hits in recognition (bottom panel), as functions of the study condition. The error bars are standard errors of their respective means

141) = 42.41,  $MSE = .016$ ,  $p < .001$ ,  $\eta^2 = 0.47$ . As in the previous experiment, reliably more words were recalled in the *self* condition than in the *both* condition,  $t(47) = 4.40$ ,  $p < .001$ , and in the *both* condition than in the *other* condition,  $t(47) = 4.83$ ,  $p < .001$ , although this time recall in the *other* condition did not differ reliably from that in the *silent* condition,  $t(47) = 0.32$ .

Recognition hit rates duplicated the clear pattern seen in Experiment 1. (The false alarm rate was .118 [ $SE = .014$ ].) The conditions differed reliably overall,  $F(3, 141) = 58.22$ ,  $MSE = .017$ ,  $p < .001$ ,  $\eta^2 = 0.54$ . Again, reliably more words were recognized in the *self* condition than in the *both* condition,  $t(47) = 5.22$ ,  $p < .001$ ; in the *both* condition than in the *other* condition,  $t(47) = 4.36$ ,  $p < .01$ ; and in the *other* condition than in the *silent* condition,  $t(47) = 3.34$ ,  $p < .01$ .

#### Comparing the experiments

Two 2 (Exp. 1 vs. Exp. 2)  $\times$  4 (encoding condition) analyses of variance were conducted to confirm the almost identical patterns for both dependent measures—free recall and recognition—in the two experiments. In both analyses, the main effect of encoding was highly reliable,  $F_s \geq 60$ ,  $p_s < .001$ , but there was no suggestion of a main effect of experiment (both  $F_s < 1$ ) nor of an interaction of

experiment with encoding condition (both  $F_s < 1$ ). The retention results of the two experiments were, therefore, completely in agreement. It is also noteworthy that an earlier full-scale pilot study produced a virtually identical pattern.<sup>1</sup>

#### Discussion

The production effect has been shown to result in a robust benefit for memory (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972; MacDonald & MacLeod, 1998; MacLeod et al. 2010). Saying aloud some of what is studied makes that portion more memorable. MacLeod et al. (2010), inspired by Conway and Gathercole (1987; Gathercole & Conway, 1988), have argued for distinctiveness as the best explanation of the benefit—that retrieval of the fact that an item was said aloud during study is used to confirm that the item was indeed studied. Ozubko and MacLeod (2010) provided evidence that distinctiveness is the “active ingredient” in the production effect, and Ozubko et al. (2011) provided evidence for the roles of both recollection and familiarity.

Until now, all studies of the production effect have compared self-performed production to no production; that is, each individual participant has read a word aloud or has read a word silently. The present study was designed to examine whether production by another person would also result in a memory benefit, and if so, what the relative benefits would be. These two experiments revealed that there is a “gradient” of benefit brought about by the production effect—from self-production, to joint production with another person, to production only by another person, to silent reading only—and that this gradient is consistent and reliable. Clearly, the production effect is largest when it is self-performed.

The *both* condition is particularly intriguing. Under the present explanation, its benefit being intermediate between

<sup>1</sup> Recognition latency data were also available for both experiments. Although less stable because of the relatively few observations per participant per encoding condition ( $\leq 10$  observations, depending on accuracy), the latencies fit with the consistent accuracy pattern. In Experiment 1, the mean response times were reliably shorter for the *self* condition (1,016 ms,  $SE = 69.60$ ) than for the other three conditions (*both*, 1,097 ms,  $SE = 63.18$ ; *other*, 1,095 ms,  $SE = 57.72$ ; *silent*, 1,129 ms,  $SE = 60.13$ ), which roughly mirrored the order of the hits analysis. (The correct rejection mean latency was 1,106 ms [ $SE = 61.57$ ].) In Experiment 2, each of the four conditions differed reliably from the adjacent conditions (*self*, 954 ms,  $SE = 35.86$ ; *both*, 1,066 ms,  $SE = 52.17$ ; *other*, 1,095 ms,  $SE = 57.72$ ; *silent*, 1,129 ms,  $SE = 60.13$ ), perfectly mirroring the accuracy pattern. (The mean latency for correct rejections was 1,159 ms [ $SE = 47.84$ ].) Put simply, recognition latencies were faster when responding was more accurate.

*other* and *self* production is not seen as social loafing occurring during encoding. Instead, it is interpreted as being due to disruption of the distinctiveness that arises from personal production: The test of distinctiveness at the time of retrieval is no longer conclusive when production at encoding was not unique to oneself. This idea has precedent in the work of Basden, Basden, Bryner, and Thomas (1997), who suggested that interfering with an individual's unique retrieval strategy impairs remembering. Note, however, that their account relied solely on activity at retrieval, whereas the production effect is seen as occurring due to the interaction of encoding and retrieval.

Very recently, Barber, Rajaram, and Aron (2010) presented evidence of a cost at encoding due to collaboration with another individual. Participants were asked to produce a sentence linking two words (e.g., *citizen–trail*), with one participant creating the first part of the sentence (e.g., *the citizen went*) and the other creating the second part (e.g., *along the trail*) versus a single participant creating the whole sentence. On a subsequent cued recall test, memory was superior for participants who had encoded individually than for those who had encoded collaboratively. Barber et al. argued that “collaborative encoding produces less effective cues for later retrieval” (p. 255). This disruption at encoding fits with what happened in the present experiments when both individuals produced responses. In place of cue effectiveness, one need only substitute distinctiveness of the additional dimension of encoding—that the word was produced—to align the accounts. Indeed, it is quite possible that differential distinctiveness provides a mechanism for differential cue effectiveness.

It is clear, then, that the production effect robustly extends to situations in which another person does the production, but that the benefit is largest when the production is done by oneself. Recollection of one's own production at the time of test is optimally distinctive, in part because individuals may focus on their own productions to verify that an item was previously studied. Any other production, therefore, is less distinctive: Indeed, the participants in these experiments did report routinely trying to remember whether they had produced an item themselves, but not routinely doing so for items produced by the other person. The likelihood of using the distinctiveness heuristic increases as production—the aspect of processing at study that can be retrieved—becomes increasingly personal and unique.

This analysis may resolve a discrepancy considered in the introduction: Modality does not typically influence long-term retention, whereas production does. Why would auditorily presented words not be remembered better than visually presented words when words spoken aloud are remembered better than those read silently? By the present

account, when modality is manipulated, words are presented either auditorily or visually, so additional modality information is present for both sets of items. In contrast, when production is manipulated, all items are presented visually, but some have the additional dimension of encoding that they were also produced aloud. This additional element provides the basis for a test of prior experience that goes beyond the item information. Add to this the more personal element of the items being produced by oneself or by another nearby person, and this may well explain why production causes a benefit, but auditory presentation does not.

In recent years, the embodied cognition perspective has come to the fore (see, e.g., Robbins & Aydede, 2009). This approach emphasizes that cognition is a situated activity, replacing the idea of cognition as computation involving a set of formal operations applied to abstract symbols. As Anderson (2003, p. 91) maintained, “thinking beings ought therefore be considered first and foremost as acting beings.” This idea is certainly relevant to language, and reveals itself in many ways, such as in the added value of thinking aloud in intelligent problem solving (e.g., Fox & Charness, 2010). Under the embodied view, the production effect gains its value from the action of producing. The activity of another person producing can certainly be processed and used for remembering, but one's own actions are more direct, more distinctive—more embodied—and hence more memorable due to their uniqueness.

**Author Note** This research was supported by Discovery Grant A7459 from the Natural Sciences and Engineering Research Council of Canada. I thank Merrick Levene for collecting the data of both experiments, and Jingjing Ji and Erica Elderhorst for carrying out pilot studies. I also thank Ramona Bobocel and Myra Fernandes for initial helpful comments, and Reed Hunt, Henry Roediger, and an anonymous reviewer for additional helpful comments in the review process.

## References

- Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, *149*, 91–130.
- Barber, S. J., Rajaram, S., & Aron, A. (2010). When two is too many: Collaborative encoding impairs memory. *Memory & Cognition*, *38*, 255–264. doi:10.3758/MC.38.3.255
- Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L., III. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1176–1189. doi:10.1037/0278-7393.23.5.1176
- Bray, N. W., & Batchelder, W. H. (1972). Effects of instructions and retention interval on memory of presentation mode. *Journal of Verbal Learning and Verbal Behavior*, *11*, 367–374. doi:10.1016/S0022-5371(72)80100-2
- Calkins, M. W. (1894). Association. *Studies from the Harvard Psychology Laboratory Psychology Review*, *1*, 476–483.

- Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341–361. doi:10.1016/0749-596X(87)90118-5
- Crowder, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology*, *1*, 157–178. doi:10.1016/0010-0285(70)90011-3
- Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161. doi:10.3758/BF03196152
- Fox, M. C., & Charness, N. (2010). How to gain eleven IQ points in ten minutes: Thinking aloud improves Raven's Matrices performance in older adults. *Aging, Neuropsychology, and Cognition*, *17*, 191–204. doi:10.1080/13825580903042668
- Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*, 110–119. doi:10.3758/BF03213478
- Hintzman, D. L., Block, R. A., & Inskip, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, *11*, 741–749. doi:10.1016/S0022-5371(72)80008-2
- Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 534–537. doi:10.1016/S0022-5371(72)80036-7
- Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology*, *62*, 242–246. doi:10.1037/1196-1961.62.4.242
- Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*, 105–112. doi:10.3758/BF03214414
- 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: Oxford University Press.
- Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory*. New York: Oxford University Press.
- Kirsner, K. (1974). Modality differences in recognition memory for words and their attributes. *Journal of Experimental Psychology*, *102*, 579–584. doi:10.1037/h0036112
- Lehman, E. B. (1982). Memory for modality: Evidence for an automatic process. *Memory & Cognition*, *10*, 554–564. doi:10.3758/BF03202438
- Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. (1976). Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75–102). New York: Wiley.
- MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*, 291–310. doi:10.1016/S0001-6918(97)00047-4
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685. doi:10.1037/a0018785
- Ozubko, J. D., Gopie, N., MacLeod, C. M. (2011). *Production benefits both recollection and familiarity*. Manuscript submitted for publication.
- Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1543–1547. doi:10.1037/a0020604
- Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 3–10). New York: Cambridge University Press.
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677–688. doi:10.1037/0022-3514.35.9.677
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld: Analyse von Vorgängen im Spurenfeld. *Psychologische Forschung*, *18*, 299–342. doi:10.1007/BF02409636