Productions need not match study items to confer a production advantage, but it helps

Megan O. Kelly[1], Xinyi Lu[1], Tyler M. Ensor[2], Colin M. MacLeod[1], and Evan F. Risko[1]

[1]*University of Waterloo*

[2]*California State University, Bakersfield*

Author Note: This work was supported by Alexander Graham Bell Canada Graduate Scholarships to MOK and XL from the Natural Sciences and Engineering Research Council of Canada (NSERC), and by a Discovery Grant (#04091) from NSERC, by an Early Researcher Award from the Province of Ontario (#ER14-10-258), by funding from the Canada Foundation for Innovation and Ontario Research Fund (#37872), and by funding from the Canada Research Chairs (#950-232147) program to EFR.

Abstract

The production effect is the finding that, relative to silent reading, producing information at study (e.g., reading aloud) leads to a benefit in memory. In most studies of this effect, individuals are presented with a set of unique items, and they produce a subset of these items (e.g., they are presented with the to-be-remembered target item TABLE and produce “table”) such that the production is both unique and representative of the target. Across two preregistered experiments, we examined the influence of a production that is unique but that does *not* match the target (e.g., producing “fence” to the target TABLE, producing “car” to the target TREE, and so on). This kind of production also yielded a significant effect—the *mismatching production* effect—although it was smaller than the standard production effect (i.e., when productions are both unique and representative of their targets) and was detectable only when targets with "standard" productions were included in the same study phase (i.e., when the type of production was manipulated within participant). We suggest that target-production matching is an important precursor to the production effect, and that the kind of production that brings about a benefit depends on the other productions that are present.

Memory can be enhanced using a variety of active encoding techniques such as generating (Slamecka & Graf, 1978; see Bertsch et al., 2007 for a review), enacting (Engelkamp, 1998; see Roberts et al., 2022 for a review), and drawing (Fernandes et al., 2018). One particularly easy form of active encoding is production, whether it is accomplished through spelling, typing, writing, mouthing, whispering, speaking, or singing (Forrin et al., 2012; Jamieson & Spear, 2014; Quinlan & Taylor, 2019). Indeed, relative to silently reading, producing items at study benefits memory performance, and is known generally as the *production effect* (e.g., MacLeod et al., 2010; see MacLeod & Bodner, 2017 for a review). Across two preregistered experiments, we used the typing production effect to examine the influence of target-production matching on the resulting production effect. Here, the target refers to a to-be-remembered study item and production refers to the result of the production act. We present results that challenge extant accounts of the production effect and that provide further insight into how active engagement with material can influence memory.

Accounts of the production effect

Since its naming by MacLeod et al. (2010), several accounts have been proposed to explain the mechanisms underlying the production effect. MacLeod et al. (2010) proposed that producing items increases the *relative distinctiveness* of those items when compared with items not produced (i.e., silently read). This is consistent with the idea that the translation of items from one modality (e.g., visual) to another (e.g., auditory) increases distinctiveness (Conway & Gathercole, 1987). Forrin et al. (2012) elaborated upon this relative-distinctiveness account by suggesting that the act of producing an item at study encourages creation of a distinctive record of that particular encoding instance due to distinct, item-specific dimensions or features (e.g., motoric, auditory) imparted by the production act at study, features that are not imparted by silent reading. Thus, the production-associated features arising from reading aloud differentiate items produced at study from items not produced (i.e., silently read), as well as from one another. Further, Forrin et al. (2012) suggested that this distinctive record imparted by the act of production allows for replaying of the record to aid retrieval. Specifically, they wrote that:

> *Any unique production provides a distinctive cue that participants can use at test to help remember studied words. In line with the proceduralist account, this distinct encoding activity is preserved in the original processing record (Kolers, 1973; Kolers & Roediger, 1984) and can subsequently be replayed to aid retrieval. (p. 1054)*

A different explanation of the production effect is the strength account (e.g., Bodner & Taikh, 2012; Bodner, et al., 2014; Bodner et al., 2016; Fawcett, 2013). Under this explanation, producing words strengthens the representations of those words in memory more than does silent reading. This greater strength makes words read aloud easier to recognize and recall compared to silently read words, and hence results in a production advantage. The strength account can be and has been modelled in computational instantiations of the production effect by varying encoding or learning quality parameters (e.g., Jamieson et al., 2016; Kelly et al., 2022). The exact underlying mechanisms through which production may strengthen item representations remains unspecified with this account, however. It is possible it is through some inherent property of the production act such as increased attention—that said, the production effect does not appear to be driven by “lazy reading” of silent items in recognition memory (e.g., MacLeod et al., 2010, although see Forrin et al., 2016).

Recently, computational models of the production effect based in MINERVA2, REM, and the Feature Model have been constructed (e.g., Jamieson et al., 2016, Kelly et al., 2022, Saint-Aubin et al., 2021). Each of these models implements the notion that producing items at study results in the encoding of additional production-associated features that are unique to each item studied aloud and are not present for items read silently, providing a computational implementation of the Forrin et al. (2012) distinctiveness account. To illustrate, Jamieson et al. (2016) were able to account for a variety of findings including that (i) the production effect and the generation effect contribute unique benefits to memory compared to each other, (ii) the magnitude of the production effect is moderated by production mode (e.g., silently read < whisper < spoken), and (iii) the effect is larger in a mixed-list (within-participants) design than in a pure-list (between-participants) design.

Target-production matching

In most studies of the production effect, individuals are presented with a set of unique targets (i.e., to-be-remembered items) and they produce a subset of them. For example, if presented with the to-be-remembered target item TABLE, the participant produces “table” aloud. That is, the production made at study in response to the target matches the target and is unique from other productions (provided the study list is composed of all unique targets). In contrast to this standard procedure wherein the target and production match, MacLeod et al. (2010) had participants make the *same* production—that is, participants said “YES”—for *every* item in the production condition, thereby preventing item-specific information otherwise afforded by the act of production (an analogous key press condition yielded the same qualitative results). In this case, the auditory and motoric features of the productions were identical across items in the production condition.

Such productions (the repeated YES's) convey little to no information about the to-be-remembered targets themselves (i.e., "YES" does not match any to-be-remembered target). Moreover, each production (of YES) was presumably rather indistinct assuming that saying the same thing multiple times leads to the storage of virtually the same features (e.g., auditory, and motoric) each time. Consequently, this form of production prevents clear distinctions of one target from another in the production condition. MacLeod et al. found that this constant production condition did not yield a production effect, supporting the idea that unique productions (i.e., distinct, non-overlapping productions) are key to the production effect, and therefore supporting a relative-distinctiveness account.

The work by MacLeod et al. (2010) provides evidence that productions must be unique from one another to yield a production effect. It remains unclear, however, whether unique productions must match the target (target: CHAIR; production: "chair"), or whether the mere uniqueness of productions is enough to confer a production effect even when productions do not match the target (e.g., remember CHAIR, produce "label"). That is, provided that productions are unique from each other (and from items read silently), is there a production effect in the context of a target-production mismatch (e.g., target: CHAIR; production: "label")?

The *distinctive record* account of Forrin et al. (2012) maintains that the stored record of the *production* episode can be used at test to facilitate memory for produced items. Under typical procedures examining the production effect (i.e., wherein productions both match their associated targets and are unique compared to each other), it is straightforward to imagine how this would be useful: The stored record of the production episode represents the to-be-remembered target—they must remember APPLE, they produce “apple,” and therefore they have auditory and motoric memory traces available at test representing the production of “apple” as evidence that APPLE was in fact studied. Critically, this would not be the case for items that they did not produce at study. Now consider the case of a target-production mismatch: Here, they must remember APPLE, but they produce “*chair.*” At test, participants would be presented with the target APPLE, not CHAIR. Provided that it is the target presented at test, which was not produced at study and therefore is not the item with the retrievable distinctive record of its production, it is unclear how having produced “chair” could facilitate memory for APPLE.

From the perspective of a strength-based account, the importance of target-production match/mismatch is unclear. On the one hand, if producing the target itself is what strengthens the target’s representation in memory and leads to a production advantage over silently read items, then a production effect would not be expected. On the other hand, if the act of a *unique* production itself facilitates attention to the target (despite the production being mismatched to the target), then a production effect *would* be expected.

Computational accounts of the production effect seemingly vary in whether they predict an influence of target-production matching/mismatching on the presence of a production effect. For example, the Jamieson et al. (2016) model implemented in MINERVA2 would appear to predict a production effect even if production does not match the target. In this model, the act of production causes more features to be stored into memory for produced items than for silently read items. At test, the presentation of a test item (i.e., a memory probe) initially contains no production-associated information. The contribution of the production-associated features emerges from an iterative retrieval process within MINERVA2 originally used to model levels of processing effects (Hintzman, 1988; Jamieson et al., 2016). Here, a series of three retrievals is made such that the retrieved content (the "echo") from each retrieval acts as the test probe for the succeeding retrieval, before arriving at a final recognition decision. Critically, for items that were produced at study, the retrieved content after the initial retrieval will contain some information about the production-associated features stored at study. Thus, for items in the production condition, when memory retrieval is prompted a second time, it is done with a test probe that contains some production-associated features based on those originally stored into memory at study. This iterative process seems unaffected by a target-production mismatch provided that the production-associated features incorporated into the test item upon "retrieval" derive from the original encoding trace stored in memory.

In contrast to the Jamieson et al. (2016) model, the account proposed by Kelly et al. (2022) using REM.1 (Shiffrin & Steyvers, 1997) seems to predict no production advantage in the case of a target-production mismatch. Like the Jamieson et al. account, this model uses additional production-associated features encoded at study to differentiate items produced from items not produced at study. However, in this model, the production-associated features are included in the probe. This was meant to capture the idea that individuals can use the production as a memory cue, as suggested by Forrin et al. (2012), and akin to individuals asking themselves “did I produce this item?” (see the idea of a distinctiveness heuristic in Dodson & Schacter, 2001). Although the utility of such an approach is clear when an individual produces the item itself, if the production was *not* of the item itself, then the utility of such an approach seems limited. Instead, for items at study that were accompanied by productions that did *not* match study items (remember CHAIR, produce “apple”), the production-associated information at test (imagining whether they produced “chair”) would not align with the production-associated information stored originally at study (having produced “apple”). This mismatch would seemingly predict no production effect when producing an item mismatching the to-be-remembered target.

Given the above analysis, examining the influence of target-production match on the production effect would advance our theoretical understanding of the production effect. This is the purpose of the present study.

The Current Investigation

We report three preregistered experiments using the typing production effect to examine the influence of target-production match/mismatch on the resulting production effect. The method was like typical studies of the production effect such that participants studied a list of targets (to-be-remembered items) and then completed a recognition test. Some of these targets were assigned to the production condition wherein participants typed, and the others were assigned to be read silently (i.e., no typing). Critically, targets in the production condition were paired either with productions that represented the targets themselves (target: TABLE, production: "table"; the typical production manipulation) or they were paired with productions that were *unrepresentative* of the targets themselves (target: TABLE, production: "apple"). The latter condition is like the "YES" manipulation from MacLeod et al. (2010) in the sense that the production does not match the target at study, but critically differs from the "YES" manipulation in that mismatched productions here were *distinct* from each other. Would there be a production benefit in the mismatched condition?

Experiment 1

Experiment 1 was preregistered at osf.io/9q2ty. Data and analysis code are available at osf.io/hnxm3/.

Method

Participants. We collected and analyzed data from N = 160 (*n* = 80 per group) participants on PROLIFIC who were paid GBP 3.34 (~USD 4.00) for their time (~20 minutes) and who were current residents of Canada, US, or UK. Sample size was based on an a priori power analysis using ANOVA_exact (Lakens & Caldwell, 2021) for a two-way mixed ANOVA with the following parameters based on extant research: a1_b1 = .65 (standard-nonproduced), a1_b2 = .75 (standard-produced), a2_b1 = .60 (nonstandard-nonproduced), a2_b2 = .65 (nonstandard-produced), sample size per cell = 80, SD = .20, and correlation among within-subject factors = .50. This yielded approximately .80 power to detect a Cohen's f of 0.22 (alpha = .05, two-tailed) and over .99 power to detect a Cohens f of 0.38 (alpha = .05, two-tailed). No participants from one experiment participated in another experiment.

Design. Production (typing vs. silent reading) was manipulated within-participant, and production type (target-production match vs. target-production mismatch) was examined between-participants. That is, in Experiment 1, we used a between-participants design to compare the effect from a standard production procedure where all the targets and productions matched (e.g., target: TABLE, production: “table”) with the effect when all the targets and productions mismatched (e.g., target: TABLE, production: “apple”).

Materials. The stimulus set consisted of 180 words ranging from 5 to 8 letters. Frequencies ranged from 7 to 26334, with a median frequency of 471.5 (FREQCount from the Open Lexicon Project; Brysbaert & New, 2009). These words were randomly assigned to 60 distinct lists of 3 items (i.e., 60 triplets); the three items within each triplet were unrelated to one another. Two of the three items would be presented at study, as they were assigned to be either the to-be-remembered target or the produced word. Half of these instances were production trials, and the other half were non-production trials.[1] The remaining item of the triplet was then withheld from study and presented at test as a foil. Each item within a triplet had an equal chance of being assigned as the study word, the produced word, or a foil.

Procedure. Participants were instructed that they were to remember a list of words. First, there was a study phase wherein words appeared one at a time. Participants were told that their job was to remember all the study words that were presented. They were also told that study words would be presented in black or blue font and that they

[1]Nonproduction trials were split into half target-production matching and half target-production mismatching to be consistent with production trials, but participants were not to produce items on these trials and rarely did so (i.e., 99% production accuracy in each experiment).

should remember the study words (regardless of font colour) for a future memory test. In addition to seeing a study word, they were told about a second word that would appear in the textbox below the study word. Specifically, they were told that the word in the textbox may or may not match the study word, but that either way, they should type the gray word in the textbox only if the study word was black, and type nothing if the study word was blue.[2] Participants were shown examples of what they might see during the study phase, and that the font colour of the study word dictated whether they should type the word in the text box.

Following these initial instructions, participants completed 18 practice study trials before moving on to the actual study trials (there were no practice test trials). In these practice study trials, participants were presented with each study word (and textbox word) for 8 seconds during which time they were either to type or not to type the textbox word, depending on the trial type. Trials were separated by 500 ms. Participants were given feedback about whether they did the correct production on the previous trial. If they were incorrect, they were told the appropriate instruction and were to try again; if they were correct, they were told so and only then could move on to the next practice trial; see Table 1. The items on the practice trials were the same for every participant and did not overlap those on the experimental trials.

Table 1

Feedback given to participants regarding accuracy during practice trials

[2]This was to make instructions as consistent as possible between conditions. This was also counterbalanced such that approximately another half of the participants were told to type the gray word in the textbox only if the study word was blue and to type nothing if the study word was black.

Trial type	Participant behaviour on practice trial	Feedback given
Production trial	*Typed textbox word correctly*	“Correct!”
	Typed study word or type textbox word incorrectly	“You typed the wrong word or took too long. Please try again.”
	Typed nothing	“You did not type the textbox word. Please try again”
Nonproduction trial	*Typed nothing*	“Correct!”
	Typed anything, including the study word or textbox word	“You should not type anything. Please try again.”

After successfully completing the 18 practice study trials, participants responded to three comprehension checks about what they were to do at study before starting the actual task. Each comprehension check covered a key part of the instructions and participants received feedback regarding the accuracy of their response. Regardless of their accuracy, they were shown the response that they had made along with the correct response.

Table 2

Comprehension checks after practice trials

Prompt	Response options	Feedback given
“In this task, I will be presented with a study word and a word that Is in a textbox. My task is to:”	(a) try my best to remember the study word [correct response] (b) try my best to remember the word that is in the textbox [incorrect response] (c) try my best to remember the study word and the word that is	“[Correct! /Incorrect.] Your task is to remember the study word. You answered: [their response]. The correct answer was [correct response]”

	in the textbox [incorrect response]	
"When the study word is presented in [no production colour], my task is to:"	(a) type the word that is in the textbox, even if the same word is presented multiple times [incorrect response] (b) type nothing [correct response]	"[Correct! /Incorrect.] Please do not type anything when the study word is presented in [non production colour]. You answered: [their response] The correct answer was: [correct response]"
"When the study word is in production colour], my task is to:"	(a) type the word that is in the textbox, even if the same word is presented multiple times [correct response] (b) type nothing [incorrect response]	"[Correct! /Incorrect.] Please type the word that is in the textbox when the study word is presented in [production colour], even if the same word is presented multiple times. You answered: [their response] The correct answer was: [correct response]"

After this, participants progressed to the actual task. Here, items were presented for 5000 ms each and were separated by 500 ms. Participants were given feedback about their production accuracy (incorrect vs. correct) on these trials but, unlike during the practice trials, they were not given the chance to correct themselves. Once the study trials were completed, they performed a self-paced arithmetic distractor task, responding to a series of simple true/false arithmetic statements (e.g., 5 x 1 = 6; 12 + 1 = 12, etc.) for 2 minutes. Finally, they completed the old/new recognition test wherein they were presented with 120 items one at a time--60 targets and 60 foils.

Because data collection was done online, participants then were asked two yes/no questions about their data quality during the experiment (informed that their responses would *not* affect their remuneration): “Did you take any notes, write anything down, or use a search engine (e.g., Google) while completing the task?” and “Were you doing anything else while completing this task? (e.g., Netflix)”.

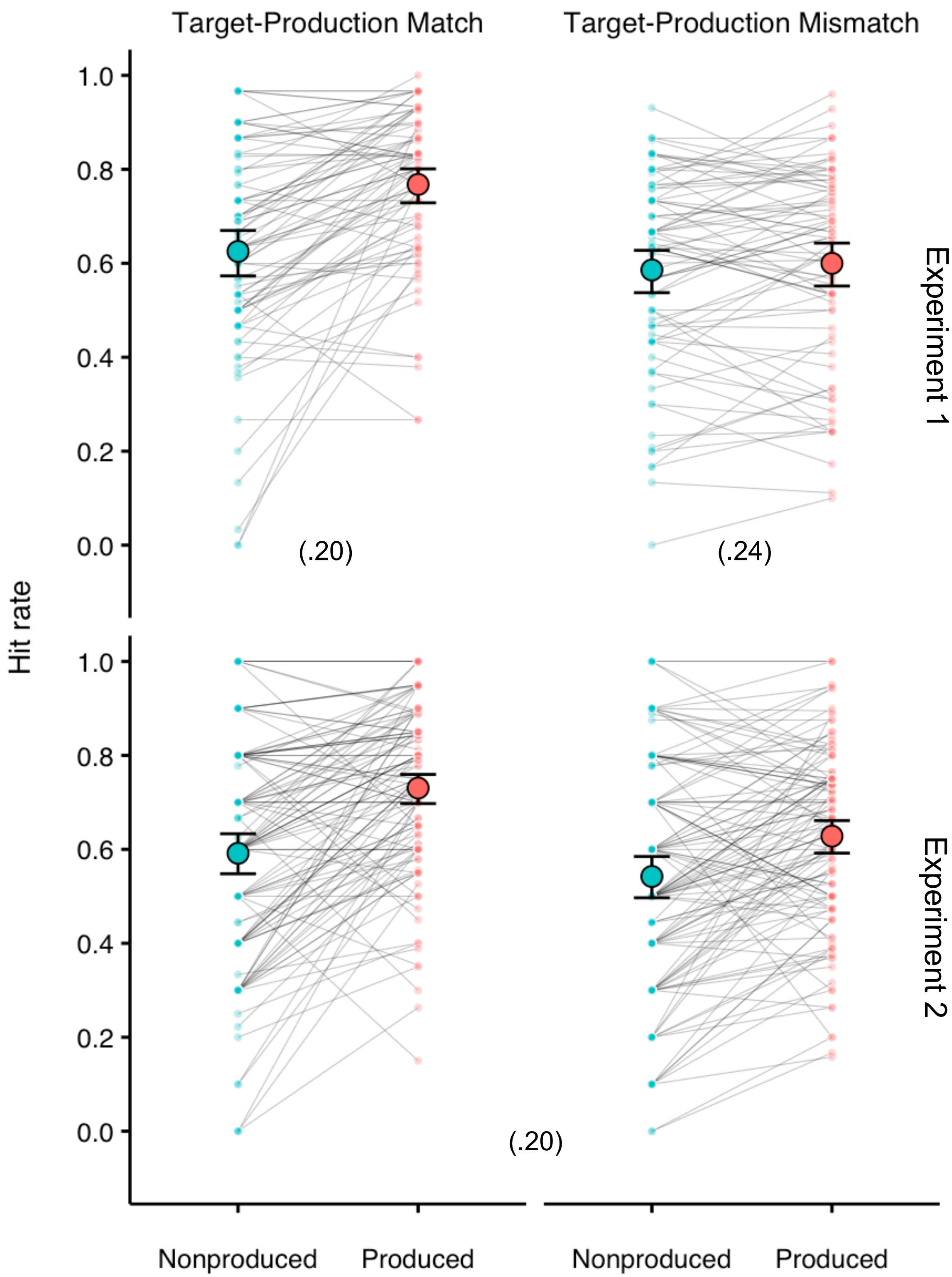
Results

We preregistered the exclusion (and replacement) of participants who indicated either that they were taking notes during the study or that they were doing anything else during task completion. Two participants who reported doing something else while completing the experiment were replaced for the final analyses, as preregistered. We also preregistered that participants would not be included in the study analyses if they could not obtain at least 70% on the arithmetic distractor task and/or if they did not respond to at least 14 arithmetic statements in the provided 2-minute duration. Three participants did not meet these criteria and were replaced in the final analyses. Finally, we had preregistered that participants must be at least 80% accurate within each production condition. Fourteen participants were replaced for not having this minimum production accuracy. Individual trials wherein participants did not appear to adhere to the production instructions (2.4% of data) were removed although results are qualitatively the same when including this data unless noted. All analyses reported were preregistered unless specified otherwise. We do not analyze d prime because the structure of the task meant that only the experimental conditions were manipulated across targets. Confidence intervals are 95% bias-corrected accelerated bootstrap confidence intervals using 10,000 samples and Effect sizes for ANOVAs are reported in terms of generalized eta squared, η_G^2 (ez package in R; Lawrence, 2016), and Cohen's *f*.

Hit Rate. A 2 (within-participant production: no-type vs. type) x 2 (between-participants target-production match: matching vs. mismatching) mixed ANOVA examined the influences of production and of target-production match on hit rate. This analysis revealed a significant main effect of production such that the hit rate was significantly higher for produced items (M = .68, CI_{95} [.65, .71]) than for non-produced items (M = .61, CI_{95} [.57, .64]), $F(1, 158) = 41.16$, $MSE = .01$, $p < .001$, $\eta_G^2 = .04$, $f = .51$. There was also a significant main effect of matching/mismatching such that the hit rate was higher when the target and production matched (M = .70, CI_{95} [.66, .73]) than when they mismatched (M = .59, CI_{95} [.56, .62]), $F(1, 158) = 12.39$, $MSE = .07$, $p < .001$, $\eta_G^2 = .06$, $f = .28$. Finally, the interaction was also significant such that the production effect was larger in the matching condition than in the mismatching condition ($M_{Matching}$ = .14, CI_{95} [.11, .19]; $M_{Mismatching}$ = .01, CI_{95} [-.01, .04]), $F(1, 158) = 27.66$, $MSE = .01$, $p < .001$, $\eta_G^2 = .03$, $f = .42$. Follow-up pairwise *t*-tests showed that whereas the production effect was significant in the matching condition [$t(79) = 6.86$, $p < .001$, $d = 0.77$] it was not significant in the mismatching condition [$t(79) = 1.10$, $p = .275$, $d = 0.12$]. Figure 1 depicts the mean hit rates (and false alarms) as a function of target type (produced vs. silent), target-production manipulation (matching vs. mismatching), and experiment (Experiment 1 vs. Experiment 2). Although not preregistered, we ran analogous Bayesian t-tests on the production effect within the matching and mismatching conditions, finding further evidence for a clear production effect in the matching condition (BF_{10} = 7759259) but not in the mismatching condition (BF_{01} = 4.54).

Figure 1

Experiment 1 (top) and Experiment 2 (bottom): Group mean and individual hit rates as a function of production and target-production match.



Note: False alarm rates are presented in parentheses. Error bars represent 95% bias-corrected accelerated bootstrap confidence intervals using 10,000 samples.

Production Accuracy. We also conducted a corresponding 2 (within-participants production: no-type vs. type) x 2 (between-participants target-production relation: matching vs. mismatching) mixed ANOVA to examine the influence of production and target-production relation on the accuracy of production (keeping in mind that participants had to have a minimum of 80% production accuracy). This analysis revealed a significant main effect of production such that production accuracy was higher when participants did not have to type anything (M = .99, CI_{95} [.99, .100]) than when they had to type something (M = .96, CI_{95} [.95, .97]), $F(1, 158) = 80.00$, $MSE = .001$, $p < .001$, $\eta_G^2 = .18$, $f = .71$. There was also a significant main effect of target-production relation such that production accuracy was higher in the matching condition (M = .98, CI_{95} [.98, .99]) than in the non-matching condition (M = .97, CI_{95} [.96, .98]), $F(1, 158) = 10.52$, $MSE = .001$, $p = .001$, $\eta_G^2 = .04$, $f = .26$. Production and target-production matching also interacted such that the influence of production on production accuracy was smaller in the matching condition than in the mismatching condition ($M_{Diff\_Matching}$ = -.03, CI_{95} [-.04, -.02]; $M_{Diff\_Mismatching}$ = -.04, CI_{95} [-.06, -.03]), $F(1, 158) = 5.95$, $MSE = .001$, $p = .016$, $\eta_G^2 = .02$, $f = .19$.

Relation between production accuracy and production effect. We preregistered a secondary analysis to examine the relation between production accuracy and the resulting production effect. Overall, there was no significant correlation between production accuracy and the magnitude of the production effect, $r(158) = -.09$, $p = .276$. For the matching condition, there was a marginal negative correlation, $r(78) = -.21$, p =

.058; for the mismatching condition, there was a small significant negative correlation, $r(78) = -.23$, $p = .043$.[3]

False alarm rate. We used a Welch two sample *t*-test (Hayes & Cai, 2007) to compare false alarm rate as a function of target-production matching and found no difference in false alarm rates between the matching condition ($M = .20$, CI_{95} [.17, .23]) and the non-matching condition ($M = .24$, CI_{95} [.20, .28]), $t(147.79) = 1.45$, $p = .149$, $d = 0.23$.

Discussion

Participants who had the (typical) target-production matching (e.g., target: TABLE; produce: “table”) condition demonstrated the usual production advantage documented in previous work (e.g., Forrin et al., 2012; MacLeod et al., 2010; Ozubko et al., 2012; Quinlan & Taylor, 2013). In contrast, participants who had mismatching productions (e.g., target: TABLE; produce: “apple”) showed no production effect (i.e., of typing vs. reading silently). Despite mismatching productions being distinctive, in that they were unique, they did not benefit memory.

We also observed lower production accuracy in the mismatching condition which could be attributable to the need to attend to typing the correct textbox word and not the to-be-remembered study word. That said, silent accuracy was high regardless because participants did not have to type anything, allowing them to attend fully to the to-be-remembered items on these trials. There was little relation between production accuracy and resulting production effect magnitude, although this was marginal for those in the

[3]Note that when the inaccurate production trials (2.4% of the data) were included, this small negative correlation became non-significant, $r(78) = 1.52$, $p = .132$.

matching production condition such that the greater the difference in production accuracy, the smaller the production effect.

Clearly, distinctive productions alone are not enough to yield a production effect. That there was no production effect when individuals made distinct productions that mismatched their targets is consistent with work by MacLeod et al. (2010) and Castel et al. (2012). Like Experiment 1, this previous work manipulating production distinctiveness (i.e., the distinctiveness of the productions from other productions) primarily examined the resulting production effect using between-participants manipulations of production distinctiveness. Previous work suggests that the production effect itself is sensitive to within- vs. between-participant contexts: The effect is far more robust when examined using a within-participants design (see MacLeod & Bodner, 2017 for a review). If the processes underlying the standard production effect are driven by a proceduralist mechanism wherein the distinctive production record can be replayed to aid retrieval, then moving to a within-participant design to examine the effect of production type should promote use of these potential strategies regardless of the type of production (target-matching vs. not target-matching). Thus, in Experiment 2, we replicated the method of Experiment 1 but moved to a completely within-participant design with respect to the encoding conditions such that every participant experienced all production conditions and production *type* conditions.

Experiment 2

Experiment 2 was a replication of Experiment 1 but with a fully within-participants design. Experiment 2 was preregistered at osf.io/c5nvz. Data and analysis code are available at osf.io/hnxm3.

Method

Data were collected and analyzed from N = 120 participants on PROLIFIC who were paid GBP 2.50 (~USD 3.17) for their time (~15 min) and who were current residents of Canada, US, or UK. Sample size was based on an a priori power analysis using ANOVA_exact (Lakens & Caldwell, 2021) for a two-way within-participants ANOVA with parameters inputted based on Experiment 1 while trying to account for the new fully within-participants design. For example, we anticipated lower nonstandard-silent items given that participants would now be seeing all four item types. Hence, the inputs were the following: a1_b1 = .64 (standard-nonproduced), a1_b2 = .76 (standard-produced), a2_b1 = .56 (nonstandard-nonproduced), a2_b2 = .60 (nonstandard-produced), sample size per cell = 120, SD = .20, and correlation among within-subject factors = .50. This yielded approximately .80 power to detect a Cohen's f of 0.28 (alpha = .05, two-tailed) and over .99 power to detect a Cohens f of 0.57 (alpha = .05, two-tailed). No participants from one experiment participated in another. The method of Experiment 2 was largely the same as that of Experiment 1 except for changes permitting the within-participant manipulation of target-production relation. That is, at study, each participant was presented 20 target-production matching trials, 20 target-production mismatching trials, and 20 nonproduction trials.[4]

Results

[4]As in Experiment 1, non-production trials were split into half target-production matching and half target-production mismatching to be consistent with the production trials, but participants were not to produce items on these trials and rarely did so (i.e., 99% production accuracy in each experiment).

We preregistered the same exclusion criteria as in Experiment 1 and all analyses were preregistered unless specified otherwise. Two participants were replaced due to their responses regarding the post-task data quality survey, two were replaced due to not meeting the minimum performance on the arithmetic distractor task, and 17 were replaced due to not meeting the minimum of 80% production accuracy within each of the four production conditions. Individual trials wherein participants did not appear to adhere to the production instructions were removed (3.5% of data; the results are qualitatively the same when including these trials). False alarm rate comparisons were not necessary because all conditions were presented on the same test, resulting in a single false alarm rate (see Figure 1).

Hit Rate. A 2 (production: no-type vs. type) x 2 (target-production match: matching vs. mismatching) repeated measures ANOVA examined the influence of production and of target-production matching on hit rate. There was a significant main effect of production such that hit rate was significantly higher for produced items (M = .68, CI_{95} [.65, .70]) than for non-produced items (M = .57, CI_{95} [.54, .60]), $F(1, 119) = 57.31$, MSE = .0, $p < .001$, $\eta_G^2 = .07$, $f = .69$. There was also a significant main effect of target-production matching such that hit rate was higher when target and production matched (M = .66, CI_{95} [.63, .69]) than when they mismatched (M = .58, CI_{95} [.56, .61]), $F(1, 119) = 48.61$, MSE = .01, $p <$ .001, $\eta_G^2 = .03$, $f = .64$. And there was a significant interaction, with the production effect being larger in the matching condition than in the mismatching condition ($M_{Matching}$ = .14, CI_{95} [.10, .18]; $M_{Mismatching}$ = .09, CI_{95} [.05, .12]), $F(1, 119) = 6.38$, MSE = .01, p = .012, η_G^2 < .01, $f = .23$. Pairwise t-tests showed that the production effect was significant both in the matching condition, $t(119) = 7.34$, $p < .001$, $d = 0.67$, and in the mismatching condition,

$t(119) = 4.94$, $p < .001$, $d = 0.45$. Like in Experiment 1, although not preregistered, we ran analogous Bayesian t-tests on the production effect within the matching and mismatching conditions and found further evidence for a clear production effect in the matching condition (BF_{10} = 292529974) and, unlike in Experiment 1, a clear effect in the mismatching condition (BF_{10} = 5222).

Comparing Experiments 1 and 2. We preregistered comparing the two experiments using two 2 (production: type vs. no type) x 2 (manipulation: within-participants vs. between-participants) mixed ANOVAs, one for the standard target-production matching condition, and one for the nonstandard target-production mismatching condition. In the matching condition, there was a significant main effect of production, $F(1, 198) = 96.02$, $MSE = .02$, $p < .001$, $\eta_G^2 = .10$, $f = .70$, no significant effect of experiment, $F(1,198) = 1.92$, $MSE = .06$, $p = .168$, $\eta_G^2 = .01$, $f = .10$, and no interaction between production and experiment, $F(1, 198) = 0.02$, $MSE = .02$, $p = .895$, $\eta_G^2 < .01$, $f = .01$. In the mismatching condition, there was a significant effect of production, $F(1, 198) = 17.71$, $MSE = .01$, $p < .001$, $\eta_G^2 = .01$, $f = .30$, no main effect of experiment, $F(1, 198) = 0.07$, $MSE = .08$, $p = .790$, $\eta_G^2 < .01$, $f = .02$, and a significant interaction between experiment and production, $F(1, 198) = 9.12$, $MSE = .01$, $p = .003$, $\eta_G^2 = .01$, $f = .21$. Taken together, these analyses indicate that the production effect in Experiment 2 was significantly larger than that in Experiment 1, a result due to the emergence of a production effect in the mismatching condition of Experiment 2 that was not evident in the mismatching condition of Experiment 1.

Production Accuracy. A 2 (production: no-type vs. type) x 2 (target-production relation: matching vs. mismatching) repeated measures ANOVA tested the influence of production and target-production relation on the accuracy of production (again, keeping in mind that participants had a minimum accuracy of 80% production). This analysis revealed a significant main effect of production such that production accuracy was, again, higher when participants did not have to type anything (M = .99, CI_{95} [.98, .99]) that when they did have to type something (M = .95, CI_{95} [.94, .96]), F(1, 119) = 68.31, MSE = .002, p < .001, η_G^2 = .14, f = .76. There also was a significant main effect of target-production relation such that production accuracy was higher in the matching condition (M = .98, CI_{95} [.97, .98]) than in the mismatching condition (M = .96, CI_{95} [.96, .97]), F(1, 119) = 17.26, MSE = .002, p < .001, η_G^2 = .03, f = .38. The significant interaction indicated that the production effect on production accuracy was smaller in the matching condition than in the mismatching condition ($M_{Matching}$ = -.02, CI_{95} [-.03, .-.01]; $M_{Mismatching}$ = -.06, CI_{95} [.-.07, -.05]), F(1, 119) = 28.60, MSE = .002, p < .001, η_G^2 = .04, f = .49.

Relation between production accuracy and production effect. As in Experiment 1, we preregistered a secondary analysis to examine the relation between production accuracy and the resulting production effect. There was no correlation between production accuracy and magnitude of the resulting production effect overall, r(118) = .05, p = .595, and no correlation in either of the target-production conditions, matching: r(118) = .08, p = .408; mismatching: r(118) = -.06, p = .488.

Discussion

As in Experiment 1, the results of Experiment 2 demonstrated a clear production effect in the typical procedure wherein the target matches the production. However, unlike in Experiment 1, there now also was a clear production effect when the target and production mismatched—a *mismatching production effect*—although this effect was significantly smaller than the typical production effect. Indeed, further analysis comparing the results across Experiments 1 and 2 revealed that there was no significant difference between the target-production match conditions across the two experiments but that there was a production effect in the target-production mismatch condition in Experiment 2 unlike in Experiment 1. Production accuracy was again worse when the production did not match the target. As in Experiment 1, there was no clear relation between production accuracy and the resulting production effect magnitude in either condition or overall.

General Discussion

We have reported two preregistered experiments using the typing production effect to examine the potential influence of target-production matching versus mismatching on the resulting production effect. In the first experiment, there was no detectable production effect for participants making productions that did not match the to-be-remembered targets. Here, target-production match versus mismatch was manipulated between participants. In the second experiment, we moved to a within-participant design where participants experienced both target-production matching and mismatching trials. We now found a significant production effect for productions that did not match targets; this *mismatching production effect* was, however, clearly smaller than the standard production effect found when productions did match their targets. Thus, the mismatching production effect was evident only when participants experienced both target-production matching trials (the standard production manipulation) and target-production mismatching trials. It is also noteworthy that the mixing of target-production match conditions within-participant did not influence the magnitude of the standard production effect (i.e., when the target and production matched). Thus, the present investigation revealed two novel variables that modulate the magnitude of the production effect: (i) the match between the target and the production and (ii) the production context (i.e., the other types of production that a participant encounters).

The current results from the perspective of extant accounts

Taken together, the two novel findings just described are difficult for some extant accounts of the production effect to explain neatly. The lack of a mismatching production effect found in Experiment 1 is seemingly consistent with the *distinctive record* account by Forrin and colleagues (2012). Recall that according to this account, when presented with a particular test item, individuals can use the stored record of having produced the target for that test item during study as evidence for that item having been presented at study. This approach to remembering is useful in typical procedures, wherein productions match their targets. However, for a target that was studied but not itself produced at study—that is, when a different unique word had been produced—basing a memory decision on whether the target itself was produced at study would seem to be an unhelpful strategy. Consequently, the *distinctive record* account cannot easily account for the results of Experiment 2—i.e., that there was an evident production effect (albeit smaller) when target and production did not match and that this was dependent on the presence of standard target-production match trials during study.

We described two types of strength-based accounts in the introduction. First, there was a target-specific strength-based account where the production of the target itself is what enhances memory for target items. From this perspective, we would only expect a production effect when targets and productions match. Hence, this type of strength-based account is consistent with the findings of Experiment 1 but not with those of Experiment

2. Second, there was a target-nonspecific strength-based account under which the act of production in general—even of something other than the target—strengthens memory of the target (e.g., by facilitating attention) regardless of the relevance of the production to the to-be-remembered target. Such an account is inconsistent with both Experiments 1 and 2: There should have been a mismatching production effect in both Experiment 1 and Experiment 2, and it also is unclear why the production effect would be smaller for mismatching target-production cases than for standard target-production cases in Experiment 2. Perhaps even more perplexing from a strength-based perspective is why the mismatching production effect occurs only when mixed with standard matching production cases.

Turning to the computational accounts of the production effect, the REM.1 account by Kelly et al. (2022) predicts no clear production effect when targets and productions mismatch and is consistent with the findings of Experiment 1. That of course means that it has difficulty in explaining the findings of Experiment 2. Like the distinctive record account by Forrin et al. (2012), the REM.1 computational account relies on the idea that individuals can use the test items to probe memory for that item having been produced at study, which is difficult when the production at the time of study did not match the target.

The results of Experiment 1 are also difficult for the computational account by Jamieson et al. (2016). As discussed earlier, this computational account implements a retrieval process whereby the originally encoded production-associated features (which are specific to produced items) are integrated into the final recognition decision (unlike the computational account by Kelly et al., 2022). From this perspective, a production effect would still be expected even when targets do not match productions because it is the original production-associated features that contribute to the recognition decision of a test item. As a result, the findings of Experiment 1 appear difficult for this account. This model fares better with the results of Experiment 2, wherein we do observe a mismatched production effect. However, it is unclear why the mismatched production effect would be smaller than the standard production effect and why it is sensitive to the other productions in the list. Taken together, when considering Experiments 1 and 2 together, no single account of the production effect appears able to explain the full set of current results.

Making sense of the present findings

One way to explain our findings would be a kind of distinctiveness account that differs subtly from the *distinctive record* account based on Forrin et al. (2012) wherein one “replays” the episode of producing the target. Instead, we begin with the idea that standard production-target matching trials are important in producing the mismatched production effect. Perhaps the presence of production-target match trials encourages reflection (or motivation) at test about whether a probe was accompanied by a production at study (matching or not matching the target). This general increase in reflection at test about the study episode would benefit memory for the to-be-remembered target even when production and target mismatched at study, albeit less effectively than when they matched. This is because presumably, certifying targets with matching productions as having been studied is more direct than is certifying targets with mismatching productions as having been studied.

Consistent with this idea of increased reflection is the higher hit rate in target-production matching trials (.73) compared to target-production mismatching trials (.62) in Experiment 2. Given this proposed role of standard target-production matching trials in driving the mismatching production effect observed in Experiment 2, there ought to be no mismatching production effect when the productions at study only mismatch. This is indeed what was observed in Experiment 1 when target-production matching was manipulated between participants. When exposed only to trials wherein targets and productions mismatch, participants may treat the consistent production of items mismatching the targets as a distracting secondary task, concentrating instead only on remembering the to-be-remembered target. This could also explain the main effect of matching in Experiment 1 such that performance was lower for the *mismatching* condition compared to the matching condition (mismatching: .59; matching: .70).

Interestingly, this explanation for the differences in findings between Experiments 1 and 2 also provides a potential explanation for why we found a mismatching production effect in Experiment 2 whereas MacLeod et al. (2010) did not find a production effect when every production was “yes.” As we alluded to in discussing Experiment 1, a potentially important difference between the MacLeod et al. (2010) study and Experiment 2 is that the target-production matching manipulation here was within-participants. If the act of production per se is what is critical, perhaps MacLeod et al. (2010) would have found a “yes” production effect had they used a within-participant design and included during study the usual subset of items wherein participants made productions which match the study items together with a subset of items with “yes” productions. Of course, it may be that the productions must be unique, in which case “yes” productions would not work under any circumstances. Thus, the present work suggests a valuable avenue for future research into how the effect of production on memory is influenced by other productions.

Conclusion

We examined the potential for unique productions to benefit memory performance both when they match and when they mismatch the to-be-remembered study items. Productions at study apparently need not match to-be-remembered study items to confer a memory benefit relative to no production. This nonmatching production effect is, however, smaller than when production at study matches study items and is only clear when a set of study items that match their productions is also present (in a within-participant design). Taken together, the present findings appear challenging for extant accounts of the production effect to neatly capture and support the idea that the kind of production that brings about a production effect likely depends on the other productions that are present.

Authorship Contributions

M. O. Kelly: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing
X. Lu: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – review & editing
T. M. Ensor: Conceptualization, Software, Writing – review & editing
C. M. MacLeod: Conceptualization, Supervision, Writing – review & editing
E. F. Risko: Conceptualization, Methodology, Supervision, Writing – review & editing

References

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201–210. https://doi.org/10.3758/BF03193441

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. https://doi.org/10.1037/a0028466

Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology, 70*(2), 93-98. https://doi.org/10.1037/cep0000082

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

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments & Computers, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2012). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition, 41*, 28-35. https://doi.org/10.3758/s13421-012-0249-6

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

Engelkamp, J. (1998). *Memory for actions*. Hove, UK: Psychology Press.

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science, 27*(5), 302-308. https://doi.org/10.1177/0963721418755385

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

Forrin, N. D., Groot, B., & MacLeod, C. M. (2016). The d-prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 42,* 1090-1111. https://doi.org/10.1037/xlm0000214

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

Hayes, A. F., & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. British Journal of Mathematical and Statistical Psychology, 60, 217-244. https://doi.org/10.1348/000711005X62576

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

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

Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology, 68*(1), 20-28. https://doi.org/10.1037/cep0000009

Lakens, D, Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science, 4*(1). https://doi:10.1177/2515245920951503

Kelly, M. O., 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*, Article 104299. https://doi.org/10.1016/j.jml.2021.104299

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition, 1*(3), 347–355. https://doi.org/10.3758/BF03198119

Kolers, P. A., & Roediger, H. L. III. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23*(4), 425–449. https://doi.org/10.1016/S0022-5371(84)90282-2

Lawrence, M. A. (2016). ez: Easy analysis and visualisation of factorial experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

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., 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

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

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory, 21*(8), 904–915. https://doi.org/10.1080/09658211.2013.766754

Quinlan, C. K., & Taylor, T. L. (2019). Mechanisms underlying the production effect for singing. *Canadian Journal of Experimental Psychology, 73*(4), 254–264. https://doi.org/10.1037/cep0000179

Roberts, B. R. T., MacLeod, C. M., & Fernandes, M. A. (2022). The enactment effect: An integrative review of behavioral, neuroimaging, and patient studies. *Psychological Bulletin*, *148*(5-6), 397-434. https://doi.org/10.1037/bul0000360

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*, Article 104219. https://doi.org/10.1016/j.jml.2021.104219

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. Psychonomic Bulletin & Review, 4, 145–166. https://doi.org/10.3758/BF03209391

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604. https://doi.org/10.1037/0278-7393.4.6.592