Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale
2016, Vol. 70, No. 2, 154–164

© 2016 Canadian Psychological Association
1196-1961/16/$12.00 http://dx.doi.org/10.1037/cep0000081

A Computational Account of the Production Effect: Still Playing Twenty Questions With Nature

Randall K. Jamieson
University of Manitoba

D. J. K. Mewhort
Queen's University at Kingston

William E. Hockley
Wilfrid Laurier University

People remember words that they read aloud better than words that they read silently, a result known as the production effect. The standing explanation for the production effect is that producing a word renders it distinctive in memory and, thus, memorable at test. By 1 key account, distinctiveness is defined in terms of sensory feedback. We formalize the sensory-feedback account using MINERVA 2, a standard model of memory. The model accommodates the basic result in recognition as well as the fact that the mixed-list production effect is larger than its pure-list counterpart, that the production effect is robust to forgetting, and that the production and generation effects have additive influences on performance. A final simulation addresses the strength-based account and suggests that it will be more difficult to distinguish a strength-based versus distinctiveness-based explanation than is typically thought. We conclude that the production effect is consistent with existing theory and discuss our analysis in relation to Alan Newell's (1973) classic criticism of psychology and call for an analysis of psychological principles instead of laboratory phenomena.

Keywords: production effect, recognition memory, MINERVA 2, distinctiveness

Psychology often uses a divide-and-conquer strategy to understand how people learn and remember. For example, the standard multiple systems view assumes different systems for explicit and implicit memory with a further subdivision into episodic, semantic, procedural, priming, classical conditioning, and nonassociative learning (Squire, 1994, 2004). Although the strategy offers the methodological convenience of allowing researchers to address one function of memory at a time, the divide-and-conquer strategy risks leaving us with a fractured perspective: a view of the trees for lack of the forest.

An alternative view is that memory is a single system capable of producing complex and even perplexing patterns when faced with different test scenarios and materials. For example, Jamieson and Mewhort (2009a, 2009b, 2010, 2011; see also Jamieson, Holmes, & Mewhort, 2010), Higham, Vokey, and Pritchard (2000), Kinder and Shanks (2001, 2003), Nosofsky and Zaki (1998), and Benjamin (2010) have all argued that implicit and explicit learning can be understood using a single set of principles and mechanisms to handle phenomena traditionally distinguished as being implicit versus explicit. The number of memory systems needed is a fundamental issue. If we need to develop a different account for each phenomenon, scientific psychology is a futile discipline. Of course, we are not the first to note the problem.

Surprenant and Neath (2009) recently called for a critical reappraisal of theories of memory. To frame their argument, they asked "If the goal of science is to identify invariants and regularities within a particular domain (Russell, 1931; Simon, 1990), one might ask, what are the laws and principles of human memory?" (p. 2). They then pointed out that whereas 100 years of psychological research has produced an ample database of empirical effects and demonstrations, the field has failed to develop a unified explanation of those effects and demonstrations. Based on the failure, they argued that psychology should orient away from growing the already overwhelming database and focus instead on developing a coherent theoretical framework that identifies and articulates key principles and laws of human behavior. But Surprenant and Neath's criticism follows from a more classic example.

In his cri de cœur titled "You Cannot Play 20 Questions With Nature and Win," Newell (1973) pointed out that psychology had become seduced into playing an empirical game—one that he likened to playing the parlor game of 20 questions. Researchers pose a binary question such as, "Is the memorial benefit of production due to distinctiveness or strength" and then resolve the opposition by experimental analysis. Having solved that one bit's worth of uncertainty, the true state of nature becomes more certain. At first blush, this strategy is entirely rational. But, Newell argued that the strategy does not work and forecasted that in 30 years (i.e.,

Randall K. Jamieson, Department of Psychology, University of Manitoba; D. J. K. Mewhort, Department of Psychology, Queen's University at Kingston; William E. Hockley, Department of Psychology, Wilfrid Laurier University.

This work was supported by Natural Sciences and Engineering Research Council of Canada Discovery grants to all three authors.

Requests for computer code as well as correspondence concerning this article should be addressed to Randall K. Jamieson, Department of Psychology, University of Manitoba, Winnipeg, MB, R3T 2N2. E-mail: randy.jamieson@umanitoba.ca

2003), our discipline would amount to a database of experimental demonstrations without a coherent explanation. In place of the 20 questions approach, Newell proposed that psychology should shift from its goal of empirical demonstration to refocus its effort on the goal of developing a coherent and general explanation of behavior.

Newell's (1973) advice has been largely ignored: Psychology has continued to play 20 questions with nature. Unfortunately, as Newell warned, psychology in 2016 has grown its database of demonstrations but has not tried to develop a unified account of human behavior. How, then, should we go about developing the kind of unified account of memory and cognition that Newell envisaged?

Consider the topic of this special issue: the production effect. People remember words that they have read aloud better than words that they have not. Although a memorial advantage for produced over unproduced words has been known for some time (e.g., Conway & Gathercole, 1987; De Haan, Appels, Aleman, & Postma, 2000; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972), it has recently been renamed and reexamined (e.g., Bodner & Taikh, 2012; Jamieson & Spear, 2014; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010).

MacLeod et al. (2010) explained the production effect as a corollary of distinctiveness, as opposed to strength (see also Richler, Palmeri, & Gauthier, 2013). According to the distinctiveness view, producing a word renders it distinctive in memory; distinctiveness, in turn, aids remembering at test. Forrin, MacLeod, and Ozubko (2012) further defined the memorial distinctiveness gained by production in terms of the additional motoric and perceptual features that arise from the productive act:

> Relative to silent reading, reading a word aloud involves the encoding of an additional dimension that stands out as distinct—speech. According to the proceduralist framework (Kolers, 1973; Kolers & Roediger, 1984), the process of vocalizing at study will be retained in a record of that processing. In an explicit memory test, participants can then retrieve this distinctive speech information to determine whether a word was studied. (p. 1046)

In contrast, the strength position argues that producing a word strengthens it in memory and that strength, in turn, aids remembering at test (see both Bodner & Taikh, 2012, and Bodner, Taikh, & Fawcett, 2014, for a discussion of strength). In addition to the strength/distinctiveness opposition, the production effect has also been examined in opposition to the enactment effect (in which people remember a written instruction better if they enact it or imagine enacting it; e.g., Engelkamp, 1995; Engelkamp & Dehn, 2000; Engelkamp, Zimmer, Mohr, & Sellen, 1994; Peterson & Mulligan, 2010) and the generation effect (in which people remember a word better if they generate it than if they read it; e.g., Begg, Vinski, Frankovich, & Holgate, 1991; Johns & Swanson, 1988; Slamecka & Graf, 1978).

In the work that follows, we take up Newell's (1973) challenge by integrating an explanation of the production effect within a framework for memory, MINERVA 2, that is already known to explain a diverse range of phenomena including confidence–accuracy inversions in recognition (Clark, 1997), false recognition (Arndt & Hirshman, 1998), speech normalization (Goldinger, 1998), decision-making (Dougherty, Gettys, & Ogden, 1999), implicit learning (Jamieson & Mewhort, 2009a, 2009b), and levels of processing (Hintzman, 1986), among others. Specifically, we will provide a formal description of Forrin et al.'s (2012) sensory-feedback account and, then, evaluate the theory's ability to accommodate data in recognition. Finally, we will use the model to examine the relation between distinctiveness and strength in the production effect. Our immediate aim is to develop a formal account of the production effect; our broader goal is to show that the production effect fits with established theory and, therefore, does not require a unique explanation.

MINERVA 2

MINERVA 2 is a multitrace model of memory (see also Kelly, Mewhort, & West, 2014). According to the model, episodic traces are stored in memory; repetition produces multiple traces of an item. Retrieval is cue-driven and parallel such that each trace is activated in proportion to its similarity to the retrieval cue, and the information retrieved from memory is a weighted sum of all activated traces.

Computationally, the model treats memory as an $m \times n$ matrix, **M**. Each row in memory stores a trace that includes n features. Each feature is assigned one of three possible values: $+1$, -1, or 0. Values of $+1$ and -1 represent the presence of information on the corresponding feature; a value of 0 indicates that information on the corresponding feature is missing or irrelevant. Each nonzero feature takes a value of $+1$ or -1 with equal probability. Traces include information about different features of a stimulus. For example, a studied word might be represented in memory by a vector of $n = 30$ features.

Encoding an item involves copying each feature in a stimulus representation to a corresponding feature in memory. Each feature is copied correctly with probability L; irrelevant or incorrectly copied features take a value of zero in memory. Thus, when $L = 0$, no elements are encoded and when $L = 1$, all elements are encoded. As L increases, encoding quality improves.

Presenting a probe, **p**, to memory, **M**, activates each trace in memory in parallel and in proportion to its similarity to the probe,

$$a_i = \left(\frac{\sum_{j=1}^{n} p_j \times M_{ij}}{n_R} \right)^3$$

where, a_i is the activation of trace i, p_j is the value of feature j in the probe, M_{ij} is the value of feature j of trace i in memory, and n_R is the number of features relevant to the comparison (i.e., the number of features for which either p_j or M_{ij} is nonzero). Activation ranges between -1 and $+1$. When the trace and probe are identical $a = 1$, when the trace and probe are orthogonal $a = 0$, and when the trace and probe are opposite $a = -1$.

The activated traces are, then, retrieved into an *echo*, which has two key properties: content and intensity. Echo content is a vector, **c**, that represents the sum of information retrieved from memory. The echo content is obtained by weighting each trace in memory by its activation and then summing the weighted traces into a composite,

$$c_j = \sum_{j=1}^{m} a_i \times M_{ij} \left\{ for\ j = 1 \ldots n \right\}$$

where c_j is the *jth* feature in the echo content, m is the number of traces in memory, a_i is the activation of trace i, and M_{ij} is the feature in the *ith* row and *jth* column in memory. Echo intensity is a scalar, f, that indexes the sum of activation elicited by the probe,

$$f = \sum_{i=1}^{m} a_i$$

where, m is the number of traces in memory and a_i is the activation of trace i in memory. Echo intensity is interpreted as an index of familiarity and is the feature used to model recognition behavior. If f is greater than a decision criterion, the probe is recognized as OLD; else, it is rejected as NEW.

As explained earlier, MINERVA 2 was designed to explain performance in a number of tasks, one of which was levels of processing. In a levels-of-processing study, participants study some words deeply (e.g., judge synonymy) and others shallowly (e.g., count vowels). A levels-of-processing effect is observed when participants remember the deeply processed words better than the shallowly processed words (Craik & Lockhart, 1972). To explain the result, researchers have argued that deep processing causes a person to encode higher-order semantic features, in addition to the lower-order features associated with shallow encoding. It is the additional information encoded that renders a word distinctive in memory and, thus, easier to remember at test.

Hintzman (1988) implemented the explanation in MINERVA 2 by encoding deeply processed words with more features than shallowly processed words. The model captured the levels-of-processing effect, and he concluded that the levels-of-processing effect is compatible with a multitrace model of memory and recognition.

Adapting MINERVA 2

We model the production effect in much the same way that Hintzman (1988) modelled the levels-of-processing result. However, we have made an additional change to explain how memory of the production features stored at study are retrieved and used at test.

First, to acknowledge the influence of production at study, we will encode produced targets with 25 features (i.e., 20 base features in Dimensions 1 through 20 of a vector representation plus 5 sensory feedback features in Dimensions 21 through 25). In contrast, unproduced targets will be encoded with 20 (i.e., 20 base features in Dimensions 1 through 20 but no sensory feedback features in Dimensions 21 through 25). In both cases, Features 26 through 30 will be set to zero. This first change acknowledges the influence of production on memory for studied words.[1]

Second, we modelled the retrieval and use of production features at test by iterative retrieval. Forrin et al. (2012; Fawcett, Quinlan & Taylor, 2012; MacLeod et al., 2010) argued that when an item has been vocalized at study, the production is retained in the record of studying that item. Consequently, "in an explicit memory test, participants can then retrieve this distinctive speech information [and use it] to determine whether a word was studied" (Forrin et al., 2012, p. 1046). The idea implies two steps in retrieval: retrieval of the fact that a word had been produced and use of that information to recover specific information about the target in question.

To implement the iterative retrieval idea, we used MINERVA 2's iterative retrieval process called *deblurring* (see Hintzman, 1986, pp. 416–417). In particular, we presented a test word as a retrieval probe to retrieve an echo. On this initial retrieval, the test probe contained only the base features (i.e., Dimensions 1 through 20) without production features (i.e., Dimensions 21 through 25). But, if the word was produced at study, the retrieved echo content, **c**, will include some information about the corresponding production features that were stored at study. In the next retrieval (and on each subsequent retrieval) we used the normalized version of the echo content, $\mathbf{c}'$, to compute another echo, where $\mathbf{c}' = \mathbf{c}/\max(\mathrm{abs}(\mathbf{c}))$.[2] Thus, echo intensity in the iterative model is computed as,

$$f = \sum_{i=1}^{m} \left(\frac{\sum_{j=1}^{n} c_j^{'} \times M_{ij}}{n_R} \right)^3$$

where f is echo intensity, m is the number of traces in memory, n is the number of features in a trace, n_R is the number of features relevant to the comparison of the probe and trace (i.e., the number of features for which either c_j' or M_{ij} is nonzero), M_{ij} is the feature in the *ith* row and *jth* column in memory, and c_j' is the *jth* feature in the normalized echo. In all simulations that follow, we report echo intensity following a third retrieval because, "Three or four echo-probe conversions are usually sufficient to produce a virtually perfect copy of one of the category names that were originally stored" (Hintzman, 1986, p. 416).

In summary, our model offers a formal representation of the distinctiveness account presented by Forrin et al. (2012). The formalization adds two mechanisms to the standard model for recognition: one to acknowledge that production affects memory of a word and another to accommodate how the production information is retrieved and used in recognition. Otherwise, our model is consistent with the principles developed in MINERVA 2. We now turn to a test of the model across a number of standard results in the production effect database.

Simulations

The Mixed-List Production Effect

The production effect is defined as a memorial advantage for produced over unproduced words. In the standard test, participants study words, half of which they produce and half of which they do not. Following study, they are tested for recognition of the studied targets that they did and did not produce, relative to unstudied foils.

We used the iterative retrieval model to simulate the standard mixed-list production paradigm (e.g., MacLeod, 2010). For each simulated subject, we generated 160 random vectors (i.e., 80 targets and 80 foils). Each vector was constructed by assigning one of two values + 1 or −1 with equal probability to each of the n = 30 dimensions. As explained earlier, Features 1 through 20 correspond to the word's base semantic features, Features 21 through 25 correspond to production features, and Features 26 through 30 correspond to higher order associative features (to be used later).

[1] Our decision to use 20 base features and 5 production features was an arbitrary one. However, we wish to emphasize that our results are not peculiar to that decision. As long as the number of production features is smaller than the number of base features, the same results obtain.

[2] This is the normalization function suggested and used by Hintzman (1986, p. 416).

Next, we stored the 80 targets to memory. Of the 80 targets stored to memory, half were stored with Features 1 through 25 filled in (i.e., produced targets) and half were stored with only Features 1 through 20 filled in (i.e., unproduced targets); in both cases, all other features were assigned a value of zero (i.e., Dimensions 26 through 30). Once all items had been stored to memory, the learning parameter was applied so that every element in memory was rewritten as a zero with probability $1 - L$ after which we computed the echo intensity for each of the 160 test items (i.e., the 80 targets and 80 foils) using the iterative retrieval process. Finally, we set a decision criterion equal to the mean echo intensity and converted all 160 echo intensities to OLD/NEW decisions accordingly. The decision to use the mean echo intensity as the criterion solves three problems. First, it makes the criterion a fixed instead of free parameter of the model. Second, the strategy can be applied at all levels of encoding quality and is, therefore, a principled way to fit the decision criterion over simulations. Third, using the mean ensures that the criterion reflects a consideration of the range of familiarities over all test items. To offer a clear picture of the model's performance, we conducted 250 independent simulations (i.e., simulated subjects) for each of five different levels of encoding quality: L = .2, .4, .6, .8, and 1.[3]

Figure 1 shows the simulation results. The top panel shows the percentage of OLD responses for the produced targets, the unproduced targets, and the foils. The bottom panel shows the size of the production effect. Whiskers show one standard deviation above and below each mean.

There are four key features to note in Figure 1. First, the model predicts a consistent recognition advantage for produced over unproduced targets: the production effect. Second, the model predicts that the size of the mixed-list production effect should increase as a function of encoding quality (see bottom panel). Third, the model assumes that the false alarm rate for unstudied foils should decrease as encoding quality increases—a result that follows in part from our decision rule that specifies rejection of test items that have an echo intensity less than the mean of all echo intensities computed in the same simulation.[4] Fourth, the difference in standard deviations for produced and unproduced words changes as a function of L (i.e., as a consequence of an emerging ceiling effect for the produced words as L approaches 1). We conclude that our model based on MINERVA 2 anticipates the standard mixed-list production effect. But does the theory also predict a smaller corresponding pure-list production effect?

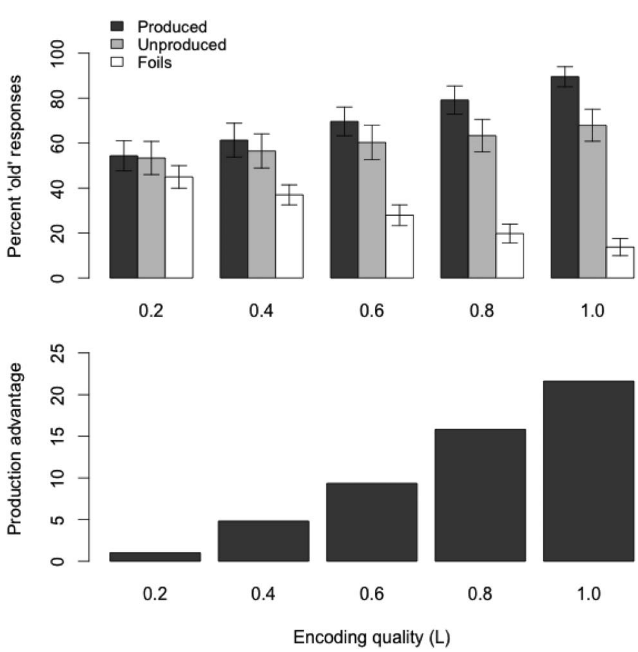

Figure 1. Simulation of the mixed-list production effect as a function of encoding quality. The top panel shows percent old performance as a function of L. The bottom panel shows the production advantage as a function of L.

The Pure-List Production Effect

In a pure-list production effect experiment, participants are tested for recognition of targets, all of which they either did or did not produce at study. A pure-list production effect is observed if participants who produced all items at study outperform participants who produced no items at study. Although the pure-list production effect is now known to be reliable, it is also known to be considerably smaller than its mixed-list counterpart (Fawcett, 2013).

To simulate the pure-list protocol, we ran two sets of simulations. In one series of simulations, we encoded all 80 targets with 25 features (i.e., Dimensions 1 through 25 filled in). In a different series of simulations, we encoded all 80 targets with 20 features (i.e., Dimensions 1 through 20 filled in). As before, we conducted 250 independent simulations at each of five levels of encoding quality: L = .2, .4, .6, .8, and 1.

Figure 2 shows the percentage of OLD responses for targets and foils as a function of encoding quality and production condition. The top panel shows results for the all-produced simulations and the middle panel shows results of the all-unproduced study simulations. Whiskers indicate one standard deviation above and below each mean. The bottom panel shows the size of the pure-list production effect.

There are three results to notice in Figure 2. First, there is a recognition advantage for produced over unproduced items: the pure-list production effect. Second, the size of the pure-list production effect grows with L. Third, the false-alarm rate decreases as encoding quality increases. In summary, the model correctly predicts that a pure-list production effect will be observed (Fawcett, 2013). More importantly, it also correctly predicts that the pure-list production effect will be much smaller than its mixed-list counterpart (compare bottom panels in Figures 1 and 2). We conclude that MINERVA 2 predicts the pure-list as well as the mixed-list production effects.

[3] The simulations were written and conducted using R 3.1.3 (R Development Core Team, 2010). Contact the first author for a copy of the code.

[4] One could misinterpret our simulated results as displaying a mirror effect, but, that would be a false impression. A mirror effect is defined as a negatively correlated relation in the hits and false alarms as a function of the dissociative factor (i.e., production). However, it is impossible to observe a mirror effect in the production effect because it is impossible to measure a false alarm rate for a produced foil: If a participant produced a "foil" at study, the participant actually studied that item.

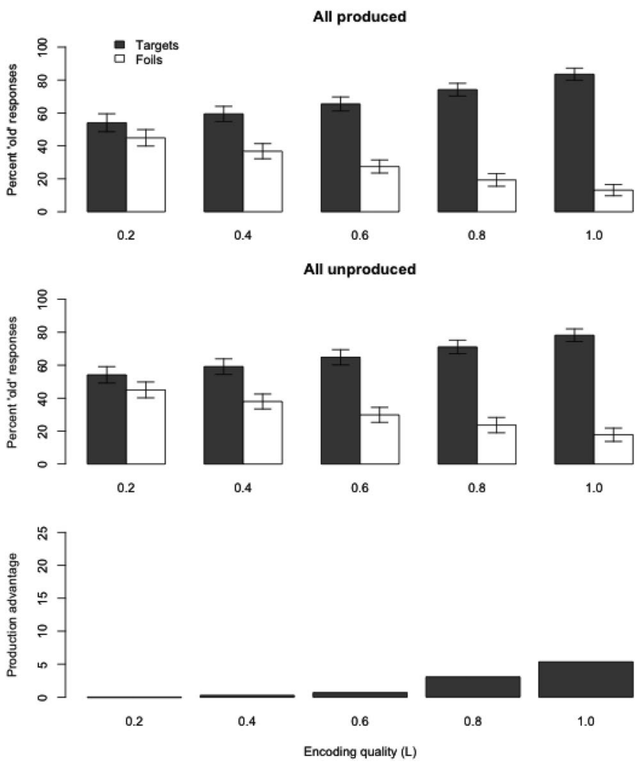

Figure 2. Simulation of the pure-list production effect as a function of encoding quality. The top panel shows percent old performance in the all produced condition and the middle panel shows percent old performance in the all unproduced condition. The bottom panel shows the production advantage as a function of *L*. There is a small pure-list production effect, but only at high values of *L*.

Forgetting

Introducing a delay between study and test weakens but does not eliminate the mixed-list production effect (e.g., Ozubko, Hourihan, & MacLeod, 2012, Experiment 1).

We simulated the effect of forgetting by simulating and testing recognition in a mixed-list design with 80 targets, half of which were stored with 20 base plus 5 sensory production features (i.e., produced targets) and half of which were stored with 20 base features alone (i.e., unproduced targets). We set *L* to .9 to bring the overall level of recognition performance in the simulation roughly in line with performance in Ozubko et al.'s (2012, Experiment1) immediate recognition condition. To simulate the delayed test, we retested recognition after deleting an additional 50% of the information in memory between the immediate and delayed simulations (i.e., a second encoding cycle in which each element reverted to zero with probability .5). The method is a standard manipulation to simulate forgetting as a function of time (see Hintzman, 1986).

We conducted 250 independent simulations of each test. Simulation results are presented in Figure 3. Whiskers show one standard deviation above and below each mean.

As shown, deleting information from memory weakened but did not eliminate the mixed-list production effect. The simulation matches the result from Ozubko, Hourihan, & MacLeod's (2012) experiment. We conclude that MINERVA 2 accommodates the influence of delay on the production effect.

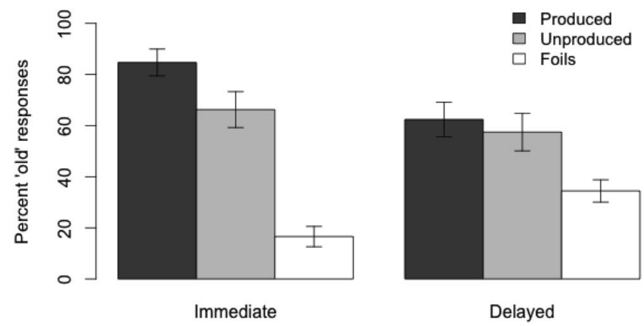

Figure 3. The mixed-list production effect as a function of study-test delay. Results are consistent with data from Ozubko, Hourihan, and MacLeod (2012, Experiment 1; see their Table 1, p. 320).

Method of Production

People recognize words that they produce more intensely better than words that they produce less intensely. For example, people recognize words that they read aloud better than words that they whisper. The top panel in Figure 4 shows an example of the result from Forrin et al. (2012, Experiment 2C). Whiskers show one standard deviation above and below each mean. See Quinlan and Taylor (2013), Gathercole and Conway (1988), and Fawcett et al. (2012) for other examples.

Forrin et al. (2012) explained the benefit of production intensity in terms of differential sensory feedback:

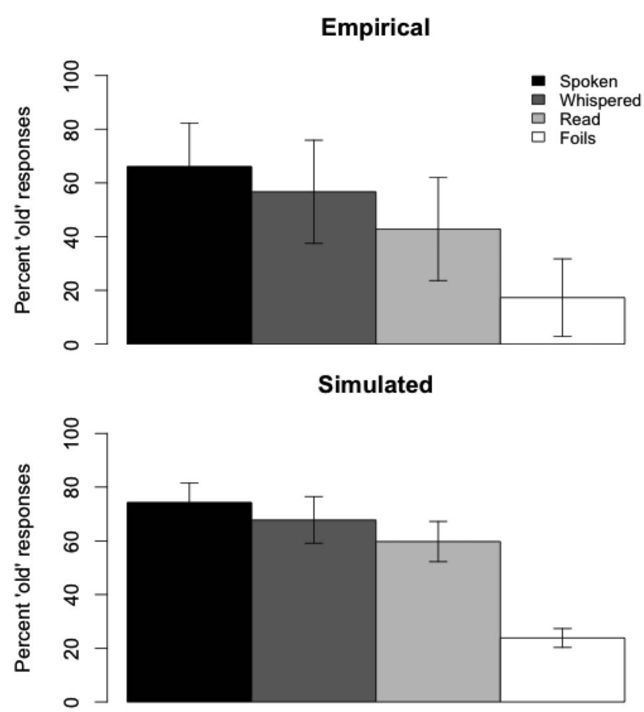

Figure 4. The mixed-list production effect as a function of study condition. Empirical data from Forrin et al. (2012, Experiment 2C) are shown in the top panel. Simulated results are shown in the bottom panel.

> Mouthing, writing, and whispering all involve additional processing that stands out as distinct—and therefore memorable—relative to silent reading. Reading aloud, however, benefits from additional processing on two distinct dimensions relative to silent reading: both articulatory and auditory processing. Moreover, the auditory processing is also distinct relative to the other modes of production, with the exception of whispering. Whispering also entails auditory processing, although the acoustic signal is weaker than the signal when reading aloud, due to the lower volume of the response, which may lead to a weaker record of processing. (p. 1053)

To capture Forrin et al.'s (2012) procedure and reasoning, we simulated recognition for spoken items with 25 features, whispered items with 23 features, and unproduced items with 20 features. The decision to represent whispered items with 23 features is arbitrary outside of the fact that it captures the idea that whispering provides more sensory feedback features than silent reading but less feedback than reading aloud.

To match the details of Forrin et al.'s (2012) procedure, we increased the length of the study list from 80 to 90 items so that memory included 30 spoken targets, 30 whispered targets, and 30 unproduced targets. We conducted 250 independent simulations to stabilize predictions and set L to .7 to bring the model's performance into the same range as participants' performance in the experiment.

Simulated results are shown in the bottom panel in Figure 4 as a function of the production manipulation; whiskers show one standard deviation above and below each mean. As shown, the model correctly anticipates the recognition advantage for spoken over whispered targets. It also predicts better recognition of produced over unproduced targets (i.e., spoken and whispered over read targets). Assuming Forrin et al.'s (2012) theoretical premise that speaking produces a stronger record of sensory feedback than whispering, and that the record of sensory feedback can be retrieved and used at test, the model anticipates that intensity of production will influence recognition performance.

Production and Generation

The production effect has a strong similarity to other well-known encoding effects, such as the generation effect, in which people remember a target better if they generate it (Begg et al., 1991; Johns & Swanson, 1988; Slamecka & Graf, 1978). It also has obvious similarities to the enactment effect in which people remember a written instruction better if they enact it (Engelkamp, 1995; Engelkamp & Dehn, 2000; Engelkamp et al., 1994; Peterson & Mulligan, 2010).

Despite the similarities between production and generation, evidence suggests that the two are not equivalent because people recognize words that they have generated and produced better than words that they have generated but not produced. The data in the top panel of Figure 5 present an example of the result (MacLeod et al., 2010, Experiment 7). Whiskers show one standard deviation above and below each mean.

Despite the empirical difference, MacLeod (2010) speculated that production and generation might benefit recognition for the same reason: "The production effect and the generation effect seem more than superficially similar. In each case, something must be retrieved from memory—either the item itself or something related to the item" (p. 236). Others have made the same point on the grounds that researchers have explained the production, generation, and enactment effects by the same mechanism: distinctiveness.

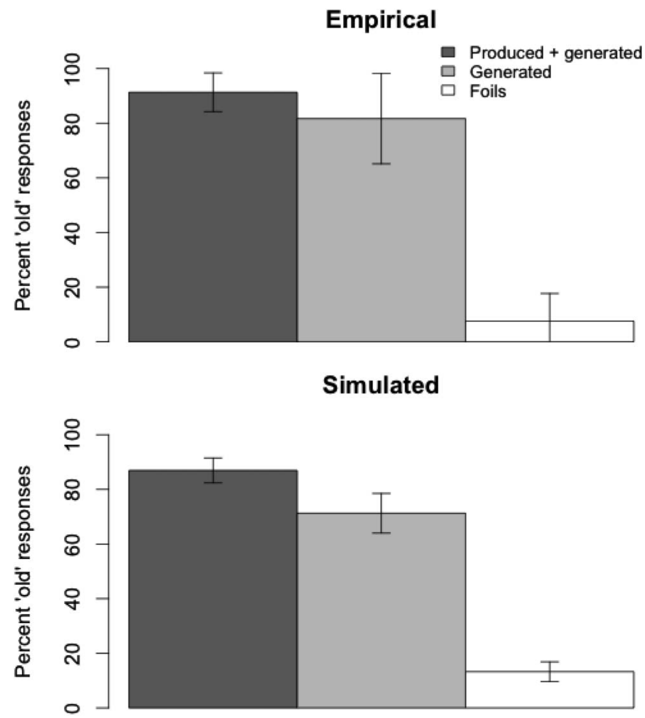

Figure 5. Recognition for words produced at study versus words produced-and-generated at study. The top panel shows empirical data from MacLeod et al. (2010, Experiment 7); the bottom panel shows our corresponding simulation.

Following MacLeod's (2010) speculation, we assumed that production and generation both enhance encoding by adding features to a studied target's representation in memory. However, production adds motoric and perceptual features whereas generation adds higher-order associative features. To implement the idea, we simulated recognition of targets encoded with 30 features (i.e., 20 base features in Dimensions 1 through 20, 5 production features in Dimensions 21 through 25, and 5 generation features in Dimensions 26 through 30) versus recognition of targets encoded with 25 features (i.e., 20 base features in Dimensions 1 through 20, no features in Dimensions 21 through 25, and 5 generation features in Dimensions 26 through 30).

The bottom panel in Figure 5 shows the percentage of OLD responses as a function of the encoding conditions. Whiskers show one standard deviation above and below each mean. We used L = .9 to bring simulated performance into the same range as in the empirical data. Means represent performance averaged over 250 independent simulations of the procedure.

As shown, the model anticipates better recognition in the production plus generation condition than in the generation alone condition: an additive benefit of production over generation. Our results are consistent with MacLeod's (2010) speculation that the two effects differ in what information they add to memory but that they both benefit recognition because of peoples' ability to retrieve and use memory of additional features encoded at study. We conclude that the distinction between production and generation can be modelled by assuming a common mechanism that operates during retrieval of both sensory and associative features.

What Does a Strength Account Actually Predict?

In the introduction to this article, we pointed out that the production effect has been developed against a binary distinction between distinctiveness and strength. According to the distinctiveness account, production renders a target distinctive and thus more memorable. According to the strength account, production renders a target more strongly encoded and thus more memorable. Received wisdom is that the distinctiveness account predicts a much stronger mixed-list than pure-list production effect whereas a strength account predicts equally probable and equally sized mixed-list and a pure-list production effects (see MacLeod et al., 2010). Because empirical data contradict predictions by the strength-based account, it has been rejected in favor of the distinctiveness account. But what does a strength-based account really predict?

To answer the question, we conducted a strength-based simulation of the mixed- and pure-list production effects by encoding both produced and unproduced targets with 20 features in Dimensions 1 through 20 but encoding produced targets more strongly than unproduced targets: $L_P = L_U + .1$.[5] We used L to represent differences in strength because that is the standard method to manipulate the strength of encoding in MINERVA 2.

Figure 6 shows the percentage of OLD responses as a function of encoding quality for the unproduced targets, L_U, where $L_P = L_U + .1$. Whiskers show one standard deviation above and below each mean. Results of the mixed-list simulations are presented in the left column and results of the pure-list simulations are shown in the right column. We show results for $L_U = .1$ through .9; results for $L_U = 1$ are excluded because L_P would equal 1.1 (i.e., greater than unity).

As shown in Figure 6, stronger encoding of the produced than unproduced targets *does not* predict equal sized mixed-list and pure-list production effects. In fact, a strength-based simulation predicts precisely the same distinction in the mixed-list and pure-list production effects as distinctiveness: a larger mixed-list than pure-list production effect. But, predictions for the mixed-list result for distinctiveness and strength are not equivalent. Whereas the distinctiveness model predicts a small mixed-list production effect that grows as a function of L, the strength model predicts a big mixed-list production effect at all levels of L—a prediction that might be refined to offer a proper test in the lab. For example, a test of the mixed-versus pure-list production effects as a function of study time or stimulus quality could discern the two (although, a manipulation of either study time or stimulus quality is limited by the practical fact that participants would need to be able to read the word as well as have enough time to produce it).

Clearly, additional work is needed to differentiate the strength versus distinctiveness accounts. But, the ability to check assumptions and generate clear testable predictions confirms the wisdom of adopting a formal model (Estes, 1975, 2002; Farrell & Lewandowsky, 2010; Hintzman, 1991, 2011; Lewandowsky, 1993).

General Discussion

People remember words that they say aloud better than words that they do not say aloud (Conway & Gathercole, 1987; De Haan et al., 2000; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972; MacLeod et al., 2010). This production advantage has been explained as a consequence of distinctiveness defined in terms of sensory feedback (Forrin et al., 2012).

We implemented the sensory-feedback account of the production effect using a model of memory inspired by MINERVA 2 (Hintzman, 1988). Our model specifies four characteristics essential to any memory account. It specifies representation assumptions: Each word is represented by a vector of binary features. It specifies encoding: The vector for each studied item is appended to the memory matrix. It specifies how a probe is handled on a recognition trial: The probe activates each item in memory and retrieves an echo; the echo content is used to compute echo intensity. It specifies how a recognition decision is made: The echo intensity is compared with a criterion and an OLD or NEW decision is given. To simulate the production effect, we adopted Forrin et al.'s (2012) premise combined with Hintzman's (1988) modelling assumptions about encoding and distinctiveness to store produced targets with more features than unproduced targets. We used an iterative retrieval process to model how participants retrieve and use the added information when performing recognition at test. The theory offers a competent account of a number of results associated with the production effect in recognition memory.

Our model-based analysis offers a novel approach to examining the production effect that involves a presentation of premises as well as mechanisms in a manner that can be analytically and objectively scrutinized (Estes, 1975, 2002; Farrell & Lewandowsky, 2010; Hintzman, 1991, 2011; Lewandowsky, 1993). For example, consider the difference between *distinctiveness* and *strength*. To make an event distinctive, we added features to its corresponding trace in memory. To encode an event more strongly, we increased encoding quality. The definitions are intuitively satisfying. But a formal analysis of the distinction exposed a problem. If increasing the number of nonzero features in memory renders a trace distinctive, why does increasing L (which also increases the number of nonzero features in memory) render a trace strong? Because both of the ideas are defined concretely in the model, the difference is easy to identify. Additional features in produced targets are shared amongst produced targets only (i.e., the added sensory features). By contrast, increasing L encodes features that are shared among both produced and unproduced targets (i.e., shared base features). Thus, added sensory features are distinct by virtue of being shared among only half of the studied targets—an idea consistent with the definition of distinctiveness that MacLeod and colleagues have been arguing.

There is, however, another issue exposed by the analysis. If distinctiveness and strength both work by adding features to a trace in memory, they are correlated concepts and, consequently, dissociating the two by experimental design will prove a more difficult or at least more complicated problem than is typically acknowledged. Naturally, the verbal definitions of strength and distinctiveness that others have been using might differ from the definitions in our model (Hunt, 2006). If so, we welcome others to import

[5] We conducted additional simulations with $L_P = L_U + .2$, $L_P = L_U + .3$, and so on. As the difference between L_P and L_U increased, the size of the pure-list production effect increased.

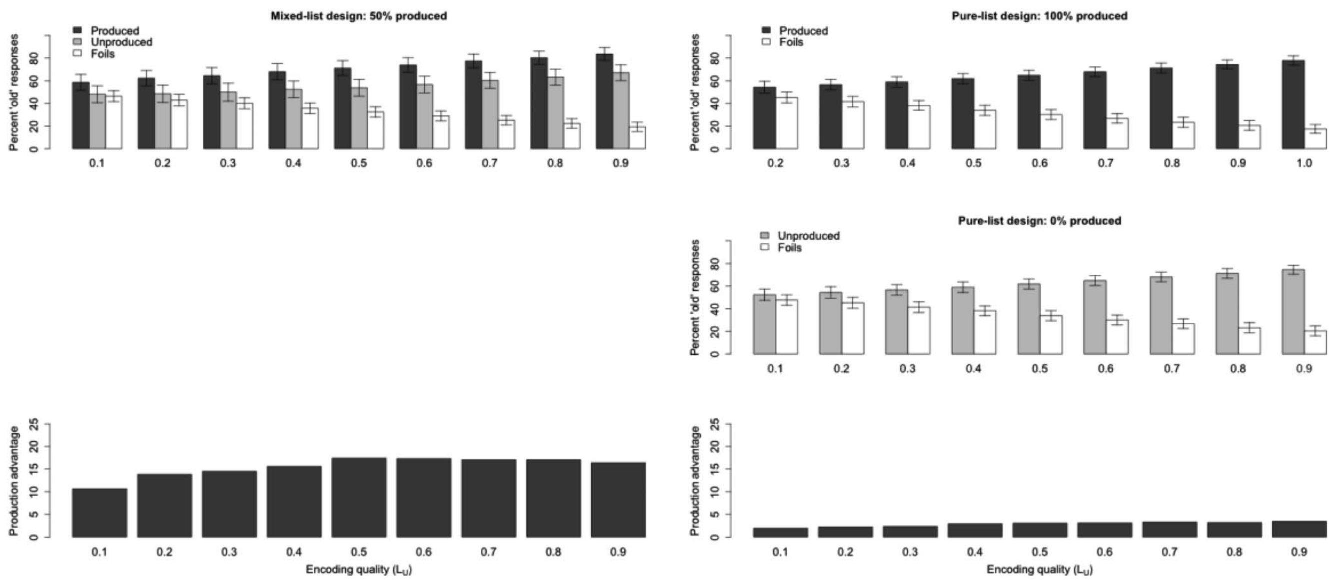

Figure 6. Strength-based simulations of the mixed-list and pure-list production effects. Contrary to expectations, a strength-based manipulation does not predict equal-sized mixed-list and pure-list production effects.

their own definitions into the model and test them. In our opinion, that exercise would be an excellent outcome of the work presented here because it would deepen the formal analysis of distinctiveness and how it figures into an analysis of memory (Brown, Neath, & Chater, 2007; Surprenant & Neath, 2009).

We adopted iterative retrieval to explain how the additional information encoded during production becomes available at test. Although simulations with MINERVA 2 have typically used noniterative retrieval, the mechanism follows directly from MINERVA 2's iterative retrieval mechanism called deblurring (Hintzman, 1986, pp. 416–417) that involves retrieving an echo, normalizing the echo, and presenting the normalized echo as a secondary probe.

In our simulations, we deblurred to three iterations. But in the original description, deblurring can be carried out indefinitely. Our decision to deblur to three iterations follows advice by Hintzman (1986) who suggested deblurring to a specified number of iterations (i.e., he suggested three or four iterations). However, additional simulations deblurring to two and four iterations confirmed that our model predictions are affected by but are not conditional on that decision. Still, additional work might focus on a more complete analysis of iterative retrieval and the role it plays in the production effect.

Our account of distinctiveness and our discussion of the role it plays in production is consistent with ideas developed in other models. For example, memory theories focused on distinctiveness typically conceive of memory as a high-dimensionality geometric space, where each studied item is plotted as a coordinate in the space. Distinctive items (i.e., items that differ from other items in memory) are located in sparse and poorly populated regions. Indistinctive items (i.e., items that do not differ much from other items in memory) are located in dense and well-populated regions. When a probe is presented to memory, it retrieves items near it. Thus, a distinctive probe (i.e., one located in a sparse region of memory) can retrieve itself without retrieving very many near neighbors (i.e., because there are few). An indistinctive probe, on the other hand, cannot help but retrieve many near neighbors (i.e., because it has many). By a difference of interference at retrieval, distinctive items are retrieved more clearly than indistinctive items—a factor that benefits recognition. This scheme is well described in Brown et al.'s (2007) SIMPLE model of memory and in the models it was developed from (e.g., Nairne, 1988, 1990). The scheme is also largely consistent with our own: Adding information on additional dimensions (i.e., production features) moves produced words to a different region in memorial space compared with unproduced words. But, there are some important differences.

The SIMPLE and Feature Models are developed for recall whereas the MINERVA 2 model was developed—and in this analysis tested—for recognition. Thus, a comparative analysis of distinctiveness in SIMPLE and MINERVA 2 would require not only measurements but also a degree of redesign. We leave this comparison and cross-examination to future effort.

The simulations that we have presented offer postdictions. But, of course, the purpose and true value of a model is prediction. For example, the theory could be used to generate predictions for the interaction of semantic and sensory distinctiveness. This can be accomplished in an experiment by presenting words that are semantically congruous versus incongruous, some of which are produced and some of which are not. The model might also be used to examine the influence of various experimental factors such as correlated sensory features, the number of study opportunities, the length of the study list, and depth of processing. As Lewandowsky (1993) has noted, computational theory is valuable because it can be used to generate predictions about performance under a large number of conditions and manipulations. It is also valuable because it ensures that predictions over multiple complex designs agree with one another and are not led astray by intuitive assumptions or oversights. We hope that the model will serve that purpose and help researchers to develop new and sophisticated experimental designs for exploring the role of distinctiveness in recognition memory.

In his now famous critique, Newell (1973) argued that psychology had overinvested in discovering phenomena when it should have been working to develop a coherent theory of behavior. His grim assessment likened psychology's phenomena-driven strategy to Victorian biology, prompting the image of a pith-helmeted avuncular gentleman pursuing specimens with butterfly net in hand. Based on that analysis, he forecasted an unhappy future in which psychology would grow into an overwhelming and incoherent database requiring researchers to develop local theories for local phenomena instead of searching for a coherent and unified explanation of behavior in general.

Our analysis takes Newell's (1973) warning to heart. Rather than contribute to the rapidly expanding empirical database on the production effect, we have paused to present a formal look at the data in hand from the perspective of standard theory. We are delighted to report that data and theory fit nicely. But, of course, our analysis has limits. First, the production effect is observed in recognition, recall, and source monitoring. Our analysis addresses recognition only. Second, we have represented words with random vectors. But, words are nonrandom events with semantic, orthographic, and phonological relations. A more complete model would use vectors that contained those relationships (see Chubala, Johns, Jamieson, & Mewhort, 2016). Third, a model is a simplification of an idea and, accordingly, our definition of distinctiveness might be impoverished relative to Forrin et al.'s (2012) ideas. We note, however, that the apparent problem is really a strength if it forces a more careful explication of the intuitive ideas about distinctiveness thought to support the production effect.

Our decision to use MINERVA 2 was deliberate. The theory has been used by a number of researchers in a number of laboratories to examine different psychological problems. Hintzman (1984, 1986, 1987, 1988) invented the theory to model recognition, frequency-judgment, classification, and cued-recall and to collapse the distinction between episodic and semantic memory. Jamieson and Mewhort (2009a, 2009b) used the theory to model implicit learning. Jamieson, Crump, and Hannah (2012) used the theory to model associative learning. Arndt and Hirshman (1998) used the theory to understand false recognition. Dougherty et al. (1999) used the theory to understand decision-making (Thomas, Dougherty, Sprenger, & Harbison, 2008). Kwantes and Mewhort (1999) used the theory to understand pronunciation in reading. Goldinger (1998) used the theory to understand speech normalization. Jamieson et al. (2010) used the theory to understand selective memory impairment in amnesia. Benjamin (2010) used the theory to investigate selective memory impairment in aging. Clark (1997) used the theory to explain confidence-accuracy inversions in recognition memory. Kwantes (2005) used the theory to explain semantic representation. And so on. Explaining the production effect using a model that is already known to explain other phenomena exemplifies the kind of productive and unified approach to understanding memory that Newell (1973, 1990) and others have called for (see Eliasmith et al., 2012; Surprenant & Neath, 2009). But is there value in building a unified account of memory?

In Newell's (1973) view, psychology had adopted a strategy of assuming nature can be understood by a series of binary divisions and decisions. Is the production effect due to distinctiveness or strength? Is recognition based on recollection or familiarity? Are categories represented by instances or prototypes? Although a strategy focused on resolving forced dichotomies is seductive, Newell warned that it yields a false impression of progress. To make the point, he asked an embarrassing rhetorical question: How many binary questions until we get to the core and truth about human behavior?

Rather than play 20 questions with nature, Newell (1973) argued that psychology should begin by specifying its principles and organizing them in a formal system. If the formal system's behavior matches that of the corresponding natural system (i.e., people), the principles can be regarded as sound. If not, they need revision or replacement. Of course, Newell's argument forces a larger reckoning. To build a complete behaving system, the theorist must consider how principles are coordinated and related on the whole. The strategy moves us away from dividing nature and playing 20 questions toward a more nuanced position on principles and behavior. Rather than ask whether the production effect is due to distinctiveness or strength, Newell's perspective asks how those concepts can be represented and how they should be mechanized in a way that translates into behavior. It also asks the hard question of how that can be accomplished in a theory that also explains other phenomena.

Of course, from Newell's (1973) perspective, we are not angels. The iterative retrieval idea that we used to acknowledge that subjects use information about production to assist in recognition is a special case of the usual retrieval process, forced by the production effect itself. That is, we explained the production effect by reconsidering how retrieval must work in the model. Yet, it is interesting to note that the change in the formal account that was needed to accommodate what appears to be an encoding effect (i.e., production at study) turned out to force a reconsideration of the mechanism for retrieval.

Résumé

Les gens se rappellent les mots qu'ils ont lus à haute voix mieux que ceux qu'ils ont lus silencieusement, un résultat appelé « effet de la production ». Cet effet serait attribuable au fait que lecture d'un mot à haute voix en accroît la distinctivité dans la mémoire, ce qui le rendrait plus facile à retenir. Selon un compte-rendu d'importance, la distinctivité est définie au moyen de la rétroaction sensorielle. Nous avons formalisé ce compte-rendu de rétroaction sensorielle au moyen de MINERVA 2, un modèle de la mémoire standard. Ce modèle s'adapte au résultat de base de la reconnaissance et au fait que l'effet de la production de la liste mixte est supérieur à celui de liste pure, que l'effet de la production est robuste face à l'oubli, et que les effets de la production et de la génération ont un effet cumulatif sur le rendement. Une dernière simulation est axée sur l'explication basée sur les forces et suggère qu'il sera plus difficile de faire la distinction entre une explication basée sur les forces et une explication basée sur la distinctivité, plus répandue. Nous concluons que l'effet de la production est conforme à la théorie actuelle, pour ensuite discuter de notre analyse en relation avec la critique d'Alan Newell (1973) de la psychologie tout en souhaitant que soit réalisée une analyse de principes psychologiques plutôt que de phénomènes de laboratoire.

Mots-clés : effet de la production, mémoire de reconnaissance, MINERVA 2, distinctivité.

References

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language, 39,* 371–391. http://dx.doi.org/10.1006/jmla.1998.2581

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19,* 487–497. http://dx.doi.org/10.3758/BF03199571

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review, 117,* 1055–1079. http://dx.doi.org/10.1037/a0020810

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

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. http://dx.doi.org/10.1037/0033-295X.114.3.539

Chubala, C. M., Johns, B. T., Jamieson, R. K., & Mewhort, D. J. K. (in press). Applying an exemplar model to the implicit rule-learning task: Implicit learning of semantic structure. *Quarterly Journal of Experimental Psychology, 69,* 1049–1055.

Clark, S. E. (1997). A familiarity-based account of confidence–accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 232–238. http://dx.doi.org/10.1037/0278-7393.23.1.232

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26,* 341–361. http://dx.doi.org/10.1016/0749-596X(87)90118-5

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 17,* 671–684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X

De Haan, E. H. F., Appels, B., Aleman, A., & Postma, A. (2000). Inter- and intramodal encoding of auditory and visual presentation of material: Effects on memory performance. *The Psychological Record, 50,* 577–586.

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. http://dx.doi.org/10.3758/BF03196152

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106,* 180–209. http://dx.doi.org/10.1037/0033-295X.106.1.180

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science, 338,* 1202–1205. http://dx.doi.org/10.1126/science.1225266

Engelkamp, J. (1995). Visual imagery and enactment of actions in memory. *British Journal of Psychology, 86,* 227–240. http://dx.doi.org/10.1111/j.2044-8295.1995.tb02558.x

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 671–682. http://dx.doi.org/10.1037/0278-7393.26.3.671

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22,* 34–39. http://dx.doi.org/10.3758/BF03202759

Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology, 12,* 263–282. http://dx.doi.org/10.1016/0022-2496(75)90025-5

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9,* 3–25. http://dx.doi.org/10.3758/BF03196254

Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science, 19,* 329–335. http://dx.doi.org/10.1177/0963721410386677

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

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

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

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. http://dx.doi.org/10.1037/0033-295X.105.2.251

Higham, P. A., Vokey, J. R., & Pritchard, J. L. (2000). Beyond dissociation logic: Evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General, 129,* 457–470. http://dx.doi.org/10.1037/0096-3445.129.4.457

Hintzman, D. L. (1984). MINERVA-2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16,* 96–101. http://dx.doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428. http://dx.doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1987). Recognition and recall in MINERVA-2: Analysis of the recognition failure paradigm. In P. E. Morris (Ed.), *Modelling cognition* (pp. 215–229). London, England: Wiley.

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

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth". *Perspectives on Psychological Science, 6,* 253–271. http://dx.doi.org/10.1177/1745691611406924

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and 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. Worthern (Eds.), *Distinctiveness and memory* (pp. 1–25). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195169669.003.0001

Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior, 40,* 61–82.

Jamieson, R. K., Holmes, S., & Mewhort, D. J. K. (2010). Global similarity predicts dissociation of classification and recognition: Evidence questioning the implicit-explicit learning distinction in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1529–1535. http://dx.doi.org/10.1037/a0020598

Jamieson, R. K., & Mewhort, D. J. K. (2009a). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology, 62,* 550–575. http://dx.doi.org/10.1080/17470210802055749

Jamieson, R. K., & Mewhort, D. J. K. (2009b). Applying an exemplar model to the serial reaction-time task: Anticipating from experience. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62,* 1757–1783. http://dx.doi.org/10.1080/17470210802557637

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: String completion and performance on individual items. *Quarterly Journal of Experimental Psychology, 63,* 1014–1039. http://dx.doi.org/10.1080/17470210903267417

Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 64,* 209–216. http://dx.doi.org/10.1080/17470218.2010.537932

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

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 180–190. http://dx.doi.org/10.1037/0278-7393.14.1.180

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2014). The Memory Tesseract: Distributed MINERVA and the unification of memory. In P. Bello, M. Gurarini, M. McShane, & B. Scassellayi (Eds.), *Proceedings of the 38th annual Cognitive Science Conference* (pp. 2483–2488). Austin, TX: Cognitive Science Society.

Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience, 13,* 648–669. http://dx.doi.org/10.1162/089892901750363217

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review, 110,* 728–744. http://dx.doi.org/10.1037/0033-295X.110.4.728

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

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

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review, 12,* 703–710. http://dx.doi.org/10.3758/BF03196761

Kwantes, P. J., & Mewhort, D. J. K. (1999). Modeling lexical decision and word naming as a retrieval process. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimental, 53,* 306–315. http://dx.doi.org/10.1037/h0087318

Lewandowsky, S. (1993). The rewards and hazards of computer simulation. *Psychological Science, 4,* 236–243. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00267.x

Macleod, C. M. (2010). When learning met memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimental, 64,* 227–240. http://dx.doi.org/10.1037/a0021699

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

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16,* 343–352. http://dx.doi.org/10.3758/BF03197045

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269. http://dx.doi.org/10.3758/BF03213879

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York, NY: Academic Press.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255. http://dx.doi.org/10.1111/1467-9280.00051

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

Peterson, D. J., & Mulligan, N. W. (2010). Enactment and retrieval. *Memory & Cognition, 38,* 233–243. http://dx.doi.org/10.3758/MC.38.2.233

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

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language, 68,* 10–25. http://dx.doi.org/10.1016/j.jml.2012.09.001

Russell, B. (1931). *The scientific outlook*. London, England: Allen and Unwin.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1–29. http://dx.doi.org/10.1146/annurev.ps.41.020190.000245

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

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems support learning and memory. In D. Schacter & E. Tulving (Eds.), *Advances in the study of memory and memory systems* (pp. 203–231). Cambridge, MA: MIT Press.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82,* 171–177. http://dx.doi.org/10.1016/j.nlm.2004.06.005

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York, NY: Psychology Press.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115,* 155–185. http://dx.doi.org/10.1037/0033-295X.115.1.155

Received June 29, 2015
Accepted March 10, 2016 ■