Mem Cogn (2012) 40:693–702
DOI 10.3758/s13421-012-0182-8

The next generation: the value of reminding

**Colin M. MacLeod · Molly M. Pottruff ·
Noah D. Forrin · Michael E. J. Masson**

Published online: 31 January 2012

Abstract In two experiments, we investigated the influence of repeated processing in the context of the generation effect. In both experiments, participants studied words once or twice. Once-studied words either were read or were generated from a definition. Twice-studied words were read both times, generated both times, or read once and generated once. Free recall was best (in order of decreasing performance) after generating twice, after generating plus reading, and finally after generating once; any generation was better than purely reading. Recognition showed a similar pattern, except that the benefit of generating twice was not as striking as in recall and that reading plus generating was just as effective as generating twice. The overall pattern of results is accounted for by a simple model in which a second encoding results in a reminding of the first encoding, and this additional encoding supports subsequent recollection. This reminding is, consequently, more effective in recall than in recognition, and it operates in accordance with the principles of transfer-appropriate processing.

Keywords Memory · Recall · Recognition

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

M. M. Pottruff
e-mail: pottrumm@mcmaster.ca

N. D. Forrin
e-mail: nforrin@gmail.com

M. E. J. Masson
Department of Psychology, University of Victoria,
Victoria, British Columbia V8W 2Y2, Canada
e-mail: mmasson@uvic.ca

The *generation effect*, introduced by Slamecka and Graf (1978), is one of the best-known encoding manipulations. Simply put, it refers to the benefit on a later memory test of producing an item from a cue at study, without seeing the entire item, as opposed to simply seeing the complete item at study. Typically, this phenomenon has involved generating a word from a cue such as a definition or an antonym, in contrast to simply reading the word. The literature contains well over 200 studies exploring the generation effect, not taking into account the considerably greater number that have just used generation as a trustworthy encoding task. Not surprisingly, the original Slamecka and Graf article is a citation classic (Slamecka, 1992).

Generation is widely viewed as a powerful way to encode, competitive with imagery (Paivio, 1969) and deep, semantic processing (Craik & Lockhart, 1972). This perception of the power of generation is supported by a recent meta-analysis that included 86 studies. Bertsch, Pesta, Wiscott, and McDaniel (2007) reported the typical mean difference in favor of the generated items over the read items to be about 9%, or just under one half of a standard deviation (.40), leading them to refer to the effect as "robust and consistent" (p. 203), a fair evaluation of its significance. In fact, it even has everyday value: Mulligan and Lozito (2004, pp. 177–178) reviewed some of the applied studies in which generation has improved memory.

What causes the generation effect? A variety of accounts have been put forward, as briefly described by Bertsch et al. (2007). It has been suggested that memory benefits from greater effort being devoted to generated items (McFarland, Frey, & Rhodes, 1980), but Bertsch et al. demonstrated that generation difficulty is not a good predictor of memory performance, and we agree with them that the concept of effort is too broad and ill-defined (see also Mitchell & Hunt, 1989, on this issue). It has been suggested that generated items steal rehearsals from read items, a selective-rehearsal account (Slamecka & Katsaiti, 1987), but Bertsch et al. presented evidence against this view, too, including the fact that blocking trial types so that the read and generate trials are separated does not reduce (and, in fact, actually seems to increase) the generation effect. And it has been suggested that the effect is a relatively straightforward outgrowth of transfer-appropriate processing (see Morris, Bransford, & Franks, 1977), but again Bertsch et al. presented evidence against this explanation, including that the type of test that shows the largest generation effect often is not the one that would appear best matched to the type of generation that had been performed during study.

Currently, the most prominent explanation of the generation effect, according to Mulligan and Lozito (2004), maintains that differential processing is devoted to generated and read items. This account (Hunt & McDaniel, 1993; McDaniel, Waddill, & Einstein, 1988), sometimes called the *multifactor* account, holds that generation increments item distinctiveness: By emphasizing item-specific processing, generation differentiates items from each other. Not coincidentally, a very similar account has been offered for the *production effect*—the finding that producing a word makes it more memorable than does simply reading it—wherein production also is seen as enhancing distinctiveness at the time of encoding, which in turn provides a benefit at the time of retrieval (MacLeod, 2011; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; Ozubko & MacLeod, 2010).

Given the extensive literature on the generation effect, it is surprising that the influence of repeated processing of the same item by generating or reading has not been explored. In fact, the only relevant study was reported by Jacoby (1978) in the same year as Slamecka and Graf (1978) introduced the generation effect. Jacoby (1978) made the related distinction between solving a problem (construction/generation) and remembering a solution (reading). Presented with word pairs on each trial, his participants either completed the second word of the pair, which had some letters deleted, or simply read the second word, which was intact; trial types were presented randomly. There followed a cued-recall test, with the first word of each pair serving as the cue for the second word. For once-presented items, Jacoby reported a large generation effect, on the order of 30%. He also included a read–read condition and a read–generate condition, and he observed roughly a 30% benefit for generation here as well. Jacoby (1978) did not, however, contrast generate–read to read–generate trials and—most critically, from our standpoint—did not examine a generate–generate condition. So we do not know what would happen to a word that was generated twice.

To date, the Jacoby (1978) study is really the only one to have explored the influence of repetition at encoding on the generation effect, at least with words as the materials (see Pesta, Sanders, & Murphy, 1999, for repetition and generation in the context of arithmetic problems). A series of studies by Rabinowitz and colleagues (Glisky & Rabinowitz, 1985; Rabinowitz, 1990; Rabinowitz & Craik, 1986) did demonstrate that reinstating the generation operation at the time of test increased the generation effect when there was a good match between the generation operations at study and test. These findings imply that multiple generation opportunities during encoding might also augment the generation advantage, and they suggested to us that comparing repetition of the read and generate operations would be worthy of exploration.

In fact, we saw two possible outcomes of repeated generation. The first represents an incremental benefit due to repetition. It could be that, once generated, a word has received its maximum benefit from generation, in that it has been made distinctive by the act of retrieval required for generation. Then, either reading or generating that item again should result in only modest, and perhaps fairly equivalent, repetition-based improvements, corresponding to the benefit seen for a previously read word on its second reading (i.e., the typical repetition effect; see Hintzman, 1970; Nelson, 1977; Smith, 1896). This outcome would be consistent, for example, with the idea that our encoding of familiar items is reduced relative to our encoding of novel items (see Tulving, 2008; Tulving & Kroll, 1995). Under this view, the second occurrence of an item is discounted, resulting in a smaller benefit than is imparted by the first encoding.

Alternatively, the second generation might confer a benefit beyond mere repetition. In thinking about this possibility, we were influenced by the concept of reminding, which Hintzman (2004, 2008, 2010, 2011) has recently promoted. He has expressed the basic idea this way:

> The basic memory system encodes information automatically whenever we pay attention to something. Remindings—and recursive encodings of the experience of reminding—also arise automatically as a result of our interaction with the environment. An encoded reminding represents the relationship between two (or more) experiences that took place at different times. (Hintzman, 2011, p. 267)

If indeed such remindings happen routinely, a second experience with an item should routinely result in a reminding, which might be viewed as an additional trace—that is, a third trace to accompany the traces resulting from the initial encoding and the second encoding. This third trace could then produce a more-than-incremental benefit to repeated processing. Moreover, this likely would have its strongest influence on recall, where retrieval demands are greatest, such that a reminding would be most beneficial. The benefit ought to be less in recognition, in which the retrieval demands are considerably less than in recall. As well, in accord with the transfer-appropriate processing principle

(Morris et al., 1977), the benefit of a reminding should be greatest when encoding of the two occurrences of an item is done in the same way. This is the idea that we set out to test: Would the generation effect be only slightly increased the second time, or would a second generation opportunity contribute more dramatically to memory? Would repeated processing be discounted, or would it promote reminding—or might it do both?

Experiment 1

Our goal was to determine how repeated processing influences memory by comparing words generated or read once to words processed twice during study. We therefore included six conditions: read once, read twice, generate once, generate twice, read then generate, and generate then read. For all repeated words, the second processing opportunity was distributed rather than massed; in the only really relevant study, Jacoby (1978) showed distributed processing to have a much more powerful effect. In Experiment 1, we also tested memory by both recall and recognition, with the recognition test following the recall test to avoid having recognition provide a second study opportunity prior to recall. In Experiment 2, we tested separate groups on the two tests. As we have already laid out, we expected that if remindings did occur, their influence should be more evident on a recall test than on a recognition test.

We expected, of course, to replicate the generation effect, with once-generated words being better remembered than once-read words, as well as the standard repetition benefit, with twice-read words being better remembered than once-read words. These findings would provide baselines for our manipulations of primary interest, the most critical of which was repeated generation. We also included read-then-generate and generate-then-read conditions, expecting that their benefit would lie between those of generating once and generating twice: These items would produce repetition benefits, but those benefits would be smaller than when the same processing was carried out on the two occasions.

Method

Participants A group of 29 University of Waterloo students took part individually for bonus course credit; the data of 3 participants were rejected due to difficulty generating the words or to technical difficulties, resulting in final data for 26 participants.

Apparatus A PC-compatible microcomputer with a 15-in. color monitor was used for testing. The controlling program was written in QuickBASIC 4.5.

Stimuli The item pool consisted of 120 words, each with a corresponding generation cue (e.g., for the word "baby," the cue was "the tiny infant commonly put in a cradle – b?"). These generation cues were an extension of the set of 90 cues used by Masson and MacLeod (2002). A total of 48 word–cue pairs were randomly selected for each participant, 8 of which were assigned to each of the six study conditions: read once, read twice, generate once, generate twice, read then generate, and generate then read. For the read trials, a single word was presented to be read aloud. For the generate trials, a generation cue was presented, and the participant was asked to generate the target word aloud. The read-once and generate-once conditions each consisted of 8 trials; in the other four conditions, each item appeared on two separate trials, creating a total of 80 study trials. Words and generation cues were presented at the center of the screen in white, small DOS font against a black background.

Procedure Each study trial began with a 250-ms blank screen, followed by a word or generation cue that remained on the screen until the participant responded aloud. The experimenter then pressed a key to score trial accuracy, after which the next study trial began immediately. A free recall phase immediately followed the study phase: Participants were given a blank piece of paper and were allowed an unlimited amount of time to write down as many words as they could recall having previously studied.[1] Following the recall test, there was a recognition test. The 48 words that were either read or generated during the study phase were randomly intermixed with 48 unstudied words and presented one at a time at the center of the screen. The participants responded "yes" (the "/" key) or "no" (the "z" key) as to whether a word had appeared in the study phase. The word then disappeared, there was a 500-ms blank screen, and the next trial began.

Results

Failure to generate a word, either once or twice from the same cue, resulted in removal of that word from both the recall and recognition data for that participant. The generation failure rates for the generate-once, generate-twice, read-then-generate, and generate-then-read conditions were .082,

[1] Following the study phase of Experiment 1, an additional 25 participants were given a speeded reading task, which served as an implicit test of memory. The stimuli consisted of the 48 studied words and 48 new words presented individually at the left center of the screen in a random order. One participant's data were discarded because of exceedingly slow responding, resulting in a final sample size of 24. Because the implicit test was not central to this research, we present these results in the Appendix.

.087, .029, and .091, respectively.[2] These failure rates were consistent with previous research on the generation effect (e.g., Masson & MacLeod, 1992).

Recall The top row of Table 1 displays the proportions of words correctly recalled in each of the six study conditions. A one-way analysis of variance (ANOVA) indicated a reliable difference in recall across the six conditions, $F(5, 125) = 28.12$, $MSE = .027$, $p < .001$, $\eta^2 = .53$. Pairwise comparisons revealed a data pattern that was consistent with past findings and with our novel predictions. Generating a word once resulted in better recall than did reading it once, $t(25) = 3.69$, $p < .01$, which is evidence of the familiar generation effect. Reading twice resulted in better recall than did reading once, $t(25) = 3.53$, $p < .01$, consistent with the standard benefit of repetition. Most importantly, generating twice resulted in better recall than did generating once, $t(25) = 7.15$, $p < .001$, with a much larger repetition benefit for generation than for reading. This increased benefit was confirmed by the finding that the difference in correct recall between generating once and twice ($M_{\mathrm{diff}} = .345$) was significantly larger than the difference between reading once and reading twice ($M_{\mathrm{diff}} = .106$), $t(25) = 5.06$, $p < .001$.

Both of the mixed conditions resulted in better recall than did generating the word once: $t(25) = 3.58$, $p < .01$, for generate–read, and $t(25) = 3.78$, $p < .01$, for read–generate. However, generating a word twice resulted in better recall than did either mixed condition: for generate–read, $t(25) = 2.92$, $p < .01$, and for read–generate, $t(25) = 3.43$, $p < .01$. There was no difference in recall between the generate–read and read–generate conditions, $t(25) = 0.25$. Thus, the combination of both reading and generating improved recall over correctly generating a word once, but not to the extent of two correct generations.

Recognition The third row of Table 1 displays the proportions of "yes" responses in each condition on the recognition test. These are hit rates for the six studied conditions; the false alarm rate is shown for the new (unstudied) condition at the far right. The low false alarm rate shows that memory in the studied conditions was quite good.

A one-way ANOVA comparing recognition accuracy for the six studied conditions indicated that there was significant variation across conditions, $F(5, 125) = 37.75$, $MSE = 0.024$, $p < .001$, $\eta^2 = .60$. As was the case for recall, generating once resulted in better recognition than did reading once, $t(25) = 6.81$, $p < .001$; reading twice resulted in better recognition than did reading once, $t(25) = 2.14$, $p < .05$; this time, generating twice resulted in marginally better recognition than did generating once, $t(25) = 1.83$, $p = .079$. Thus, there was again evidence of a repetition effect for words that were either read or generated. The difference in correct recognition between generating once and twice ($M_{\mathrm{diff}} = .061$) was not significantly different from the difference in correct recognition between reading once and reading twice ($M_{\mathrm{diff}} = .111$), $t(25) = 0.77$.

Both of the mixed conditions resulted in better recognition than did generating the word once, $t(25) = 3.45$, $p < .01$, for generate–read, and $t(25) = 2.66$, $p < .05$, for read–generate. Unlike in recall performance, however, generating a word twice did not result in better recognition than did either of the generate-plus-read conditions (*p*s > .15). (Although this might appear to have been the consequence of a ceiling effect, Exp. 2b will demonstrate that it was not.) As in recall, there was no difference in recognition between the generate–read and read–generate conditions, $t(25) = 0.93$.

[2] The generation failure rate was (consistently across experiments) lowest for the read-then-generate condition, presumably because only this condition provided participants with the word prior to having to generate it. As there were no errors in the two read-only conditions, to eliminate the possibility of differential error rates influencing the results, we conducted two separate analyses of the recall scores: One analysis excluded items that were incorrectly generated, and the other analysis included all studied items. In this experiment—and in each of the subsequent experiments reported in this article—these two analyses produced the same outcomes. Thus, we feel confident that item-selection effects did not contaminate our results.

Experiment 2

In this second experiment, we sought to replicate and extend the results of Experiment 1. Toward that end, in Experiment 2a we focused solely on the recall component of Experiment 1. In Experiment 2b, we focused on the recognition component, without the prior recall test of Experiment 1. Our goal in Experiment 2b was to reduce recognition accuracy so that we could observe performance without any possibility of a ceiling effect. To accomplish this, we incorporated a retention interval, putting a three-day gap between the study phase and the recognition test.

Experiment 2a

Method

Participants A group of 34 naive students from the same pool as in Experiment 1 took part individually for course credit. The data of four participants who had difficulty generating words (50% errors in at least one generate condition) were discarded, leaving 30 in the final sample.

Apparatus A PC-compatible computer with a 17-in. color monitor was used for testing. The controlling program was written in E-Prime 1.2, and the items were presented in 16-point Times New Roman font in black against a white background.

Table 1 Mean proportions correct in recall and mean proportions of "yes" responses in recognition in Experiments 1, 2a, and 2b

	Read Once	Read Twice	Generate Once	Generate Twice	Generate–Read	Read–Generate	New
Recall							
Exp. 1	.038	.144	.162	.507	.343	.330	–
	(.013)	(.026)	(.030)	(.041)	(.041)	(.045)	–
Exp. 2a	.071	.225	.262	.606	.436	.485	–
	(.024)	(.028)	(.038)	(.029)	(.035)	(.038)	–
Recognition							
Exp. 1	.457	.567	.806	.867	.903	.879	.078
	(.037)	(.047)	(.032)	(.029)	(.024)	(.031)	(.013)
Exp. 2b	.355	.429	.614	.751	.741	.747	.155
	(.031)	(.034)	(.030)	(.026)	(.029)	(.026)	(.016)

Standard errors are shown in parentheses below each corresponding mean

Stimuli The studied items and the study and recall test procedures were identical to those of Experiment 1, apart from the changes in apparatus and appearance.

Results

As in Experiment 1, generation failures were removed prior to all analyses. The generation failure rates for the generate-once, generate-twice, read-then-generate, and generate-then-read conditions were .042, .050, .025, and .050, respectively.[3]

The second row of Table 1 displays the proportions of words correctly recalled in each study condition. An overall one-way ANOVA indicated a reliable difference in recall across the six conditions, $F(5, 145) = 43.53$, $MSE = 0.027$, $p < .001$, $\eta^2 = .60$. The pattern was identical to that of Experiment 1. Indeed, when the recall data from Experiments 1 and 2a were combined, treating experiment as a between-subjects variable, both main effects were significant, but the 2 × 6 interaction was not, $F < 1$. (The main effect of experiment signified that, overall, performance was better in Experiment 2a than in Experiment 1.)

Unsurprisingly, then, pairwise comparisons revealed differences between the means that were consistent with Experiment 1. Once again, there was a generation effect: Generating a word once resulted in better recall than did reading it once, $t(29) = 4.97$, $p < .001$. There was also a strong effect of repetition for both reading and generating. Reading twice resulted in better recall than did reading once, $t(29) = 5.66$, $p < .001$, and generating twice resulted in much better recall than did generating once, $t(29) = 6.99$, $p < .001$. As with Experiment 1, the difference in correct recall between generating once and twice ($M_{\text{diff}} = .344$) was significantly larger than the difference in correct recall between reading once and twice ($M_{\text{diff}} = .154$), $t(29) = 3.15$, $p < .01$.

Consistent with Experiment 1, generating a word once resulted in poorer recall than did either generating and then reading a word, $t(29) = 3.23$, $p < .01$, or reading and then generating a word, $t(29) = 4.48$, $p < .001$. Generating a word twice, however, resulted in better recall than either the generate–read condition, $t(29) = 3.63$, $p < .01$, or the read–generate condition, $t(29) = 2.87$, $p < .01$. Again, there was no difference in recall between the two mixed conditions, $t(29) = 0.97$. Experiment 2a perfectly replicated the recall results of Experiment 1.

Experiment 2b

To remove any concern that in Experiment 1 the recognition pattern might have reflected a ceiling effect, or that prior recall might have contaminated recognition, we tested only recognition and incorporated a three-day retention interval to lower overall performance.

Method

Participants A group of 61 naive students from the same pool as in the previous experiments participated individually in exchange for bonus course credit. The data of 3 participants were rejected due either to difficulty generating the words or to technical difficulties, and 4 withdrew from the experiment after the study phase, resulting in final data for 54 participants.

[3] This decrease in the rate of generation failures relative to Experiment 1 may have resulted from a difference in procedure. In Experiment 2a, participants were given more time to correctly generate a word than they had been in Experiment 1, which may have produced fewer failed generations due to "giving up." Similarly, the mean recall scores were higher in Experiment 2a than in Experiment 1. This difference may have arisen because the experimenter in Experiment 2a encouraged participants to take their time and then left the room during the recall test, whereas the experimenter in Experiment 1 stayed in the room with the participants, possibly leading them to stop sooner.

Stimuli and apparatus These were identical to the stimuli and apparatus of Experiment 2a.

Procedure The study phase was identical to that in the previous experiments. The recognition test was nearly identical to that in Experiment 1, with the difference being the three-day retention interval between the study and recognition phases. Participants responded "yes" or "no" as to whether a word had been studied (i.e., either read or generated) by using either the "m" or the "x" key. The assignment of keys to responses was counterbalanced.

Results

Generation failures were removed prior to all analyses. The generation failure rates for the generate-once, generate-twice, read-then-generate, and generate-then-read conditions were .053, .053, .032, and .056.

The fourth row of Table 1 displays the proportions of "yes" responses (hit rates) in each condition on the recognition test. The false alarm rate for the new (unstudied) condition (shown at the far right) was again fairly low, demonstrating reasonable memory for the words in all of the studied conditions.

A one-way ANOVA comparing recognition accuracy for the six studied conditions showed significant variation across the conditions, $F(5, 265) = 61.81$, $MSE = 0.027$, $p < .001$, $\eta^2 = .54$. The pattern of the means was entirely consistent with that of Experiment 1. As was the case for recall in Experiments 1 and 2a, when we treated experiment as a between-subjects variable for the recognition data from Experiments 1 and 2b, both main effects were significant, but the 2 × 6 interaction clearly was not, $F < 1$. (The main effect of experiment signified that performance was worse in Experiment 2a than in Experiment 1, likely the result of the longer retention interval.) That the pattern of the condition results was the same in the two recognition experiments allays any possible concern about a ceiling effect having influenced the findings in Experiment 1.

Pairwise comparisons highlighted the expected differences between condition means. There was a strong generation effect: Generating a word once resulted in better recognition than did reading it once, $t(53) = 7.45$, $p < .001$. There were also repetition effects for both reading and generating: Reading twice resulted in better recognition than did reading once, $t(53) = 2.57$, $p < .05$, and generating twice resulted in better recognition than did generating once, $t(53) = 4.22$, $p < .001$. Although the repetition effect was numerically larger for generating ($M_{\text{diff}} = .137$) than for reading ($M_{\text{diff}} = .074$), this difference was nonsignificant, $t(53) = 1.44$, $p = .16$.

Generating and then reading a word resulted in better recognition than did merely generating the word once, $t(53) = 4.07$, $p < .001$. Reading and then generating a word also resulted in better recognition than did a single generation, $t(53) = 4.39$, $p < .001$. Consistent with the recognition data in Experiment 1, generating a word twice did not result in better recognition than did either of the generate-plus-read conditions (ts < 1). Once again, there was no difference in recognition between the generate–read and read–generate conditions, $t(53) = 0.23$. Experiment 2b perfectly replicated the recognition results of Experiment 1.

General discussion

In two experiments, we explored the consequences of reading and generating target words when participants process items twice. In so doing, we have replicated some well-known findings, and we have uncovered several interesting new observations. By way of replication, we have shown that, in recall, reading for a second time (repetition) improves memory, and does so to about the same extent as does generating once; and, in recognition, although reading twice is superior to reading once, generating once results in a considerably greater advantage than does simple repetition of reading. By way of new observations, we have shown that (1) in recall, generating twice confers a dramatically greater advantage than does generating once; (2) in recognition, the advantage of generating twice over generating once is reliable but considerably smaller than in recall; (3) in recall, reading and generating the same item, regardless of order, results in performance about halfway between generating once and generating twice; and (4) in recognition, reading and generating the same item, regardless of order, results in performance equivalent to generating twice. These patterns of findings were remarkably consistent across our two experiments.

It is interesting that under the multifactor account (Hunt & McDaniel, 1993; McDaniel, Waddill, & Einstein, 1988; see Mulligan & Lozito, 2004), generation is seen as enhancing item-specific representations and cue–target relational encoding, but as disrupting interitem relational encoding. Because recognition is usually thought to be mainly sensitive to item-specific representations, it would seem that generating twice ought to benefit recognition more than should either of the generate-plus-read combinations. We found the three conditions to be equal. In contrast, free recall is seen as sensitive to one form of information that generation enhances (item-specific) and another that it does not enhance, or may even disrupt (interitem relational). So it would seem that, as compared to the generate-plus-read combinations, there should be smaller benefits of generating twice than might be found in recognition. We found the opposite. These surprising findings led us to think about our results in a different way.

To help organize our pattern of findings, we initially considered a simple model (Model 1) in which each encoding episode was assumed to create a memory trace that could make an independent contribution to that item's later retrieval. To fit the data from the free recall and recognition tasks, this model had four free parameters, corresponding to the probability that an encoding episode (read or generate) would lead to a correct response on a memory task (free recall or recognition). Where two encoding episodes occurred for an item, there were two independent contributions toward a correct response for that item on a memory test. Model 1 was evaluated by attempting to fit the data from Experiments 2a and 2b, where the data were clearly free of any concerns about ceiling effects. The model failed to provide an adequate fit to either the recall or the recognition data (root-mean-squared error = .052, $\chi^2(8) = 27.66$, $p < .01$). Figure 1 shows the predicted mean performance based on this model in comparison to the observed data. The parameter values for the fit of Model 1 are shown in the top row of Table 2.

It is apparent that, for the free recall data, Model 1 substantially overestimated performance in the read-once and generate-once conditions. In this simple, two-independent-trace model, elevated estimates of the effectiveness of a single read or generate encoding were required to enable the model to produce sufficiently strong recall performance in the conditions involving two encodings. Had the parameter values for a single read or generate encoding more closely matched the actual recall data, the predictions for the other four

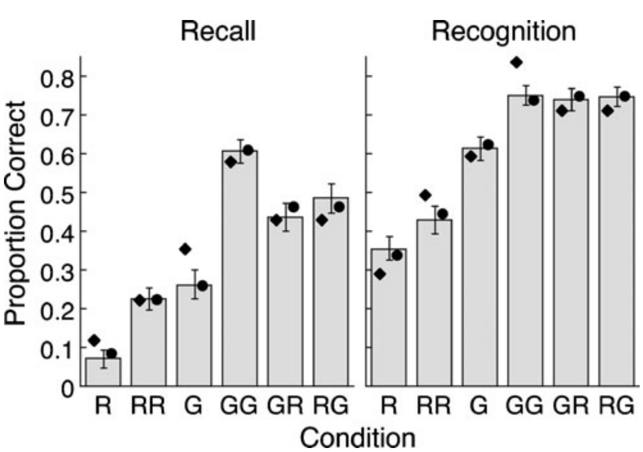

Fig. 1 Observed proportions of correct recall and recognition in Experiment 2 (bars) and the predicted performance from our two models. Conditions are designated by the encoding task(s) performed during study (R, read; G, generate). Diamond symbols indicate the predictions for Model 1, and circle symbols the predictions for Model 2. See the text for the details of the models. Error bars represent one standard error of the mean

Table 2 Best-fitting parameters for Models 1 and 2, applied to the data from Experiments 2a and 2b

	r_F	g_F	r_R	g_R	α	β	γ
Model 1	.118	.352	.289	.593	–	–	–
Model 2	.081	.259	.335	.621	.492	1.460	.208

r_F, probability that the first read encoding will support successful free recall; g_F, probability that the first generate encoding will support successful free recall; r_R, probability that the first read encoding will support successful recognition; g_R, probability that the first generate encoding will support successful recognition; α, proportionate amount by which memory retrieval parameters r and g are reduced to establish the probability that the memory trace for the second encoding of an item will support later recall or recognition when that item is encoded with the same task both times; β, proportionate amount by which memory retrieval parameters r_F and g_F are increased to establish the probabilities that the reminding memory trace produced by the second encoding of an item in the read-twice and generate-twice conditions will support successful free recall; γ, probability that the reminding memory trace produced by the second encoding of an item using a new task will support successful free recall. Note that both β and γ take on values of zero in recognition, where reminding is assumed to play no role

conditions would have substantially underestimated the observed performance. For the recognition data, Model 1 clearly underestimated the read-once condition and overestimated the read-twice and generate-twice conditions. This overestimation indicated that two encodings of an item using the same task failed to produce as much benefit as would be expected from two equally strong and independent memory traces.

To improve our theoretical account of the data, Model 1 was modified to address two fundamental issues. First, in the recognition data, the problem for Model 1 was that observed performance following two identical encodings of an item fell below what an independence model would predict. We therefore considered the possibility that when encoding an item using the same task for a second time, participants may have relied to some extent on memory for the prior encoding episode with that item, leading to a more efficient but less memorable encoding experience (e.g., Jacoby, 1978).

In Model 2, then, we added a parameter, α, to capture the reduction in memory trace strength on trials on which an item was encoded for a second occasion using the same encoding task as on the first occasion. This parameter represented a proportionate decrease in the value of the memory trace strength associated with a read or a generate encoding episode. For example, if we designate the probability of correct recognition based on the first read encoding episode for an item as r_R, the probability of a second read encoding episode supporting later recognition of that item would be given as $\alpha(r_R)$. This reduction of trace strength yielded by a second encoding under an identical task can also be interpreted as nonindependence between two memory traces. Under independence, if one trace fails to support remembering, the other trace may still support remembering with unaltered probability. Under nonindependence, however, failure of the first trace implies that the item's second trace will likely be relatively weak, hence the application of the α parameter. We suggest that it is reasonable to suppose that two identical encodings of an item may be correlated, given that much the same cognitive operations would be applied on the two occasions. Furthermore, we assumed that this violation of independence would apply only when the same encoding task was used for both presentations of an item. For items in the read–generate and generate–read conditions, then, the two encodings were assumed to produce two independent memory traces, each with strength corresponding to the type of encoding task executed.

Because this nonindependence assumption for read-twice and generate-twice encodings was meant to reflect operations taking place in the study phase, both recognition and free recall performance would have to be influenced in the same way. Therefore, the α parameter was also applied when determining the trace strength for the second identical encoding of an item with respect to later free recall. For example, if we designate the probability that the first generate encoding of an item will lead to successful free recall as g_F, the probability that a second generate encoding of that item would support recall was $\alpha(g_F)$.

The second issue that was addressed in constructing Model 2 pertains to Model 1's overestimation of free recall in the read-once and generate-once conditions. That overestimation was the consequence of bringing predicted performance in the conditions involving two encodings up to the observed levels of recall. We propose that for the free recall task, retrieval operations make use of an additional aspect of encoding that transpired when an item was presented for a second time. Hintzman (2010, 2011) has argued that a second presentation of an item leads to a reminding of the earlier presentation of that item, and this reminding is itself encoded into memory. We incorporated this idea into Model 2 by assuming that the second presentation of an item created not only a second memory trace for that item, but also a third trace, independent of the first two traces, representing the reminding that occurred during the second presentation.

To establish a reminding trace for conditions in which the same encoding task was used on both presentations, we took the parameter representing the strength of the trace for a single encoding in a given task and modulated this strength by multiplying it by a new parameter, β. For example, for the generate-twice condition, the probability that a reminding trace would support subsequent free recall was $\beta(g_F)$. For the read-twice condition, the reminding memory trace for recall would have strength equal to $\beta(r_F)$. The cases involving a different encoding task on each of the two presentations posed an interesting question as to the strength of the resulting reminding memory trace. With one read and one generate encoding episode for such items, it was unclear how the two episodes would combine to determine the strength of a reminding trace. Therefore, rather than starting with a base memory trace strength and multiplying it by the β parameter, we introduced one final parameter, γ, to specify the strength of the reminding memory trace for the read–generate and generate–read items.

As compared to what was seen in the recall data, the relatively small amount of benefit accruing to recognition memory for read-twice and generate-twice items suggested that participants were not relying on a reminding memory trace when making recognition judgments. Indeed, when we fit a version of Model 2 in which reminding memory traces were free to influence performance in the recognition task as well as in free recall, both β and γ were estimated to be equal to 0 for the recognition task. Moreover, Hintzman (2004) showed that judgments of frequency, which rely heavily on recollection of remindings in his account, can be dissociated from confidence-based ratings of recognition memory. We suggest that in the recognition task, participants did not engage in the recollection operations associated with frequency judgments or free recall that were likely to bring reminding memory traces into play. Given these considerations, the fit of Model 2 that we present here restricted the involvement of reminding memory traces to the free recall task. The best-fitting parameter values for Model 2 are shown in Table 2, and the predicted probabilities of recall and recognition are shown in Fig. 1.

Model 2 produced a very good fit, with all predicted values within one standard error of the observed values, and most within just half a standard error [root-mean-squared error = .014, $\chi^2(5) = 1.86$, $p > .75$]. The parameter values are sensible, in that they establish a higher probability of correct responding for recognition than for free recall, and for generate than for read encoding. In addition, the parameter that determined the reduction in strength of a second encoding with the same task, α, dropped the strength of the second trace to about half that of the first trace. The β parameter was larger than 1, indicating that the reminding memory trace is a substantial embellishment of the second encoding event when the same task is used on both presentations of an item. Finally, γ, the reminding trace parameter for mixed encoding items, resided between the trace strength values for the first read and first generate encoding of an item. According to Model 2, then, reminding traces for read–generate and generate–read items are weak in comparison to those of items encoded in the same way on both presentations. This difference may reflect a reduced likelihood of a successful reminding when the encoding task is changed between presentations, caused by a relatively low degree of overlap in processing operations applied to the item on the two occasions. An interesting prediction of the model arises from this speculation—namely, that judgments of frequency and recency, which rely on reminding memory traces (Hintzman, 2004, 2010), should be less accurate when items are presented under different encoding tasks, rather than under the same encoding task, on their successive presentations.

In conclusion, our study of the situation in which processing of words via reading and generation is repeated, in the various possible combinations, has been instructive in two regards. First, empirically, it has revealed a hitherto unknown benefit of repeated generation, which is especially notable in recall. Second, theoretically, it has been informative with respect to the processes underlying the repeated encoding of an item. The free recall data, in particular, are consistent with the proposal that a second encoding yields a reminding of the earlier study episode that can support later recollection processes (Hintzman, 2011). Moreover, this reminding appears to be sensitive to the principles of transfer-appropriate processing (Morris et al., 1977), as it is less effective when the encoding task changes across presentations. Overall, then, the value of repetition depends on both the extent of processing overlap and the extent of retrieval, with reminding apparently assuming a key role.

Author note This research was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grants A7459, to C.M.M., and A7910, to M.E.J.M. M.M.P. is now at McMaster University. We thank Anthony Marley for helpful discussion about the modeling work reported here.

Appendix

Table 3 displays the mean response times for speeded reading—simply reading the words aloud as quickly as possible—which was the implicit test included in Experiment 1. A one-way ANOVA comparing the seven conditions (six studied and one unstudied) was not significant overall, $F(6, 138) = 0.99$, $MSE = 1{,}952.60$, $p > .40$. Nonetheless, a single planned comparison was conducted, contrasting the six studied conditions to the one unstudied condition, and this was significant, $F(1, 23) = 9.00$, $MSE = 35{,}118.96$, $p < .01$. Clearly, priming did occur for studied words—their mean of 507 ms showed 19 ms of priming relative to the unstudied mean of 526 ms—but that priming was unaffected by study condition.

This pattern of equivalent priming on an implicit speeded reading test for all studied items conceptually replicates the pattern that has consistently been observed in studies of the production effect (MacDonald & MacLeod, 1998; MacLeod et al., 2010), in which items spoken aloud are recalled and recognized better than those read silently on explicit tests but do not show differential priming on implicit tests. The finding also agrees with several studies (Masson & MacLeod, 1992, 1996, 2002; MacLeod & Masson, 1997, 2000) using various implicit measures (including speeded reading), all of which have shown equivalent priming for read and generated items, contrary to influential earlier work (Jacoby, 1983). Our view is that the distinctiveness of produced or generated items is useful on explicit tests, which emphasize remembering whether an item was previously studied, but not on implicit tests, in which episodic experience is irrelevant. The absence of differential priming therefore represents a correspondence between generation and production.

Table 3 Experiment 1: Mean response times (in milliseconds) in speeded reading

Read Once	Read Twice	Generate Once	Generate Twice	Generate–Read	Read–Generate	New
509	500	510	506	515	500	526
(14.01)	(17.96)	(14.91)	(15.15)	(15.64)	(11.56)	(12.10)

Standard errors are shown in parentheses below each corresponding mean

References

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684. doi:10.1016/S0022-5371(72)80001-X

Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 193–205.

Hintzman, D. L. (1970). Effects of repetition and exposure duration on memory. *Journal of Experimental Psychology, 83,* 435–444. doi:10.1037/h0028865

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition, 32,* 336–350.

Hintzman, D. L. (2008). Memory from the outside, memory from the inside. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A festschrift for Gordon H. Bower* (pp. 15–30). Mahwah, NJ: Erlbaum.

Hintzman, D. L. (2010). How does repetition affect memory? Evidence form judgments of recency. *Memory & Cognition, 38,* 102–115. doi:10.3758/MC.38.1.102

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. doi:10.1177/1745691611406924

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421–445.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667. doi:10.1016/S0022-5371(78)90393-6

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98,* 291–310.

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

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.

MacLeod, C. M., & Masson, M. E. J. (1997). Priming patterns are different in masked word identification and word fragment completion. *Journal of Memory and Language, 36,* 461–483. doi:10.1006/jmla.1997.2501

MacLeod, C. M., & Masson, M. E. J. (2000). Repetition priming in speeded word reading: Contributions of perceptual and conceptual processing episodes. *Journal of Memory and Language, 42,* 208–228.

Masson, M. E., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General, 121,* 145–176. doi:10.1037/0096-3445.121.2.145

Masson, M. E. J., & MacLeod, C. M. (1996). Contributions of processing fluency to repetition effects in masked word identification. *Canadian Journal of Experimental Psychology, 50,* 9–21. doi:10.1037/1196-1961.50.1.9

Masson, M. E. J., & MacLeod, C. M. (2002). Covert operations: Orthographic recoding as a basis for repetition priming in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 858–871. doi:10.1037/0278-7393.28.5.858

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521–536.

McFarland, C. E., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19,* 210–225.

Mitchell, D. B., & Hunt, R. R. (1989). How much "effort" should be devoted to memory? *Memory & Cognition, 17,* 337–348.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. doi:10.1016/S0022-5371(77)80016-9

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 45, pp. 175–214). San Diego, CA: Academic Press.

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 151–171. doi:10.1016/S0022-5371(77)80044-3

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

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76,* 241–263. doi:10.1037/h0027272

Pesta, B. J., Sanders, R. E., & Murphy, M. D. (1999). A beautiful day in the neighborhood: What factors determine the generation effect for simple multiplication problems? *Memory & Cognition, 27,* 106–115.

Rabinowitz, J. C. (1990). Effects of repetition of mental operations on memory for occurrence and origin. *Memory & Cognition, 18,* 72–82.

Rabinowitz, J. C., & Craik, F. I. M. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language, 25,* 226–237. doi:10.1016/0749-596X(86)90031-8

Slamecka, N. J. (1992). Delineation of the generation effect—A citation-classic commentary on "The generation effect: Delineation of a phenomenon", by Slamecka, N. J., & Graf, P. *Current Contents/Social and Behavioral Sciences, 18,* 10.

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

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26,* 589–607.

Smith, W. G. (1896). The place of repetition in memory. *Psychological Review, 3,* 21–31. doi:10.1037/h0070984

Tulving, E. (2008). On the law of primacy. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A festschrift for Gordon H. Bower* (pp. 31–48). Mahwah, NJ: Erlbaum.

Tulving, E., & Kroll, N. E. A. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review, 2,* 387–390.