Mem Cogn (2012) 40:326–338
DOI 10.3758/s13421-011-0165-1

Production benefits both recollection and familiarity

Jason D. Ozubko · Nigel Gopie · Colin M. MacLeod

Published online: 30 November 2011

Abstract In three experiments, we investigated the roles of recollection and familiarity in the production effect—the finding that words read aloud are remembered better than words read silently. Experiment 1, using the remember/know procedure, and Experiment 2, using the receiver operating characteristic procedure, converged in demonstrating that production enhanced both recollection and familiarity. Experiment 3 supported the role of recollection by demonstrating that specific episodic information—that is, whether a word had been studied aloud or silently—was stronger for items studied aloud. These findings fit with an explanation of the production effect as hinging on two factors: greater recollection of distinctive information from the study episode, and more familiarity due to greater attention allocated to the material studied aloud.

Keywords Production · Recollection · Familiarity

Very few encoding manipulations result in consistent, reliable memory benefits when a to-be-remembered stimulus is presented only once. Well-established mnemonics include imagery (Paivio, 1971), elaboration (i.e., levels of processing; Craik & Lockhart, 1972; Craik & Tulving, 1975), generation (Slamecka & Graf, 1978; see Bertsch, Pesta, Wiscott, & McDaniel, 2007, for a review and meta-analysis), and enactment (e.g., Cohen, 1981; Engelkamp & Krumnacker, 1980; see Engelkamp, 1998; Zimmer et al., 2001, for a review). To this list might be added lesser-known techniques such as narrative chaining (Bower & Clark, 1969), and newer techniques such as survival processing (Nairne & Pandeirada, 2008), but even then, the list is short.

Over the course of the last four decades, another mnemonic technique was discovered (Hopkins & Edwards, 1972)—and rediscovered (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; MacDonald & MacLeod, 1998)—but each time it was forgotten, despite clear and compelling data. Yet this technique is so very simple: When some words from a list are read aloud and others are read silently, memory is considerably enhanced for the words read aloud. MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010; see also Hourihan & MacLeod, 2008; Lin & MacLeod, 2011; MacLeod, in press; Ozubko & MacLeod, 2010) recently reintroduced this robust mnemonic and provided it with a name: the *production effect*.

The production effect represents a dependable and substantial enhancement of memory, measurable in both recognition (MacLeod et al., 2010) and recall (Conway & Gathercole, 1987; Gathercole & Conway, 1988; Lin & MacLeod, 2011). In the experiments to date, this straightforward manipulation has routinely boosted memory by 10% to 25%. Moreover, unlike other mnemonic techniques—in which the to-be-remembered information must be capable of being imagined, must be capable of being enacted, must be elicited from a well-matched cue, or must be sufficiently meaningful to permit elaborate processing—the production effect relies only on making a unique response to a stimulus—characteristically, simply saying it aloud. Furthermore, even

J. D. Ozubko (✉) · N. Gopie
Rotman Research Institute, Baycrest Centre,
3560 Bathurst Street,
Toronto, Ontario M6A 2E1, Canada
e-mail: jdozubko@rotman-baycrest.on.ca

C. M. MacLeod
University of Waterloo,
200 University Avenue West,
Waterloo, Ontario N2L 3G1, Canada
e-mail: cmacleod@uwaterloo.ca

following deep semantic encoding, MacLeod et al. observed a reliable production advantage. This points to another remarkable aspect of the production effect: It benefits retention whether the information is initially encoded weakly or strongly. MacLeod et al. demonstrated a production effect not only for items initially encoded by reading, but even for items initially encoded by generation or deep semantic processing (ruling out the "lazy reading hypothesis"; cf. Begg & Snider, 1987). In sum, the breadth of materials and encodings that can benefit from a production boost appears to be extensive, a point underscored by Forrin, MacLeod, and Ozubko (2011), who showed reliable enhancements for spelling, typing, and writing, among other modes of production.

The qualitative effect of production

A central question yet to be addressed with respect to the production effect concerns the extent to which production benefits recollection and/or familiarity. There now exists a wealth of data from cognitive, neuropsychological, and neuroimaging sources that converges on the idea that remembering can be dissociated into two distinct processes: recollection and familiarity (see Yonelinas, 2002, for a review; see also Eichenbaum, Yonelinas, & Ranganath, 2007; Skinner & Fernandes, 2007). *Recollection* refers to the experience of consciously re-creating or reimagining past events, and is akin to what most laypersons would refer to when they claim to have "remembered" something. *Familiarity*, on the other hand, refers merely to how familiar stimuli feel to an individual, regardless of whether any conscious recollection occurs.

In delineating the production effect, MacLeod et al. (2010) argued for a recollective-based distinctiveness account of that effect. In so doing, they adopted the theoretical perspective taken by earlier investigators, notably in the two most detailed studies reported previously (Conway & Gathercole, 1987; Gathercole & Conway, 1988). The idea behind the distinctiveness account is that a word read aloud has an additional source of discriminative information relative to a word read silently (i.e., relative to a word not produced aloud). From the proceduralist perspective (cf. Kolers, 1973; Kolers & Roediger, 1984), saying a word aloud results in a unique record of the operations used during encoding. Therefore, at the time of test, if the subject can recollect having said that word aloud, the subject can be certain that it was in fact studied. In contrast, neither words read silently nor distractor words will have a memory record of having been said aloud. The distinctiveness of a word studied aloud therefore provides a basis for a recollection boost during retrieval.

Corroborating the claim that the production effect benefits overall memory performance by incrementing distinctiveness, and thereby bolstering recollection, are three particularly relevant findings. First, it is now clear that the production effect is limited to within-subjects, mixed-list designs (Hopkins & Edwards, 1972; MacLeod et al., 2010; but see Dodson & Schacter, 2001, for a "production effect" in false alarm rates that occurs between subjects). Because distinctiveness is necessarily relative (see Hunt, 2006), only in within-subjects designs can distinctiveness express itself strongly. Second, the effect is observed only on explicit tests of memory, not on implicit tests (MacDonald & MacLeod, 1998; MacLeod et al., 2010). Recollection of a distinctive encoding is irrelevant on an implicit test, so there should be no production effect. And third, when the value of the distinctive record is undermined—for instance, by having subjects produce other, interfering information aloud—the production effect is eliminated (Ozubko & MacLeod, 2010). Despite these consistent findings, however, neither MacLeod et al. nor the earlier investigators provided direct evidence that conscious recollection drives the production effect. In fact, although these three findings suggest, indirectly, that familiarity may be unaffected by production, past work examining the recollection and familiarity benefits of similar mnemonic techniques has suggested that the production effect may not be restricted to recollection.

Studies on effective encoding techniques other than the production effect have often found that encoding manipulations that boost memory act by increasing both recollection and familiarity. For example, elaboration studies (i.e., manipulations of levels of processing) have found that deep processing enhances both recollection and familiarity (e.g., Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000; Rajaram, 1993; Toth, 1996; see Yonelinas, 2002, for a summary). Similarly, generation has also been observed to increase both recollection and familiarity (e.g., Dodson & Johnson, 1996; Jacoby, 1991; Jennings & Jacoby, 1993; Verfaellie & Treadwell, 1993; see Yonelinas, 2002, for a summary). Given that production seems especially related to generation (i.e., both involve the retrieval and pronunciation of a word), these findings suggest that production might well increase both recollection and familiarity, rather than just recollection.

On the one hand then, theoretical considerations to date and some indirect evidence converge to suggest that production enhances memory by boosting recollection. On the other hand, studies of apparently similar mnemonic phenomena suggest that production might actually increase both recollection and familiarity. The goal of the present research was to directly investigate the influence of production on recollection and familiarity. Although production has mainly been considered a recollective advantage to date, here we will show that production actually acts to increase both recollection and familiarity.

Experiment 1

Experiment 1 assessed the effects of production on recollection and familiarity as measured by the remember/know procedure. In the remember/know paradigm, introduced by Tulving (1985; see Gardiner, 1988), measures of recollection and familiarity are obtained by explicitly asking subjects to differentiate between them. That is, instead of being asked to indicate only whether a test word is "old" or "new," subjects are asked to divide their "old" responses into those that are associated with episodic detail ("remember") and those that are associated with familiarity in the absence of episodic detail ("know"). The experiencing of episodic detail presumably requires the recollection of that detail.

Although the remember/know procedure is the simplest and apparently most direct method for measuring recollection and familiarity, this procedure is often criticized or questioned as a valid approach for assessing recollection and familiarity (e.g., Donaldson, 1996; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Wixted, 2007; Wixted & Stretch, 2004). Other researchers, however, have come forward in support of the remember/know paradigm and the dual-process interpretation of it (e.g., Eichenbaum et al., 2007; Skinner & Fernandes, 2007; Yonelinas, 2002). We do not wish to digress into this debate extensively here, since it is beyond the scope of this article; however, Rotello, Macmillan, Reeder, and Wong (2005) have demonstrated that standard remember/know instructions often lead to "remember" responses that do not agree with estimates of recollection obtained via other methods, such as receiver operating characteristic (ROC) curves. Yet, when more conservative remember/know instructions were given, the recollective estimates derived from this paradigm did agree with those of other paradigms (see Yonelinas, 2002, for a review; see also Yonelinas, 2001; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996; although see Wixted & Mickes, 2010, for the argument that confidence must also be taken into account in the remember/know procedure before valid estimates of recollection and familiarity can be derived). These results demonstrate that some of the past criticisms or inconsistencies regarding the remember/know paradigm may have been due to the fact that the procedure was not always being carried out as strictly as it should. At a more general level, however, these results highlight the importance of careful and deliberate instructions whenever the remember/know paradigm is used, such that reliable estimates of recollection and familiarity can be observed. Therefore, to achieve reliable estimates of recollection and familiarity, conservative instructions were used in Experiment 1.

Finally, before proceeding to Experiment 1, one more methodological point regarding the remember/know paradigm must be discussed—specifically, the fact that "know" responses themselves do not provide ideal measures of familiarity, because they are limited by the proportion of "remember" responses. That is, one important aspect to keep in mind with regard to the remember/know procedure is that unadjusted "know" responses underestimate the influence of familiarity. In the remember/know procedure, subjects are instructed to respond "know" only if an item is familiar and they cannot recollect any details surrounding it. In essence, then, as the proportion of "remember" responses increases, the proportion of "know" responses must necessarily decrease, as there will be fewer opportunities to make "know" responses. If the raw proportion of "know" responses were taken as a measure of familiarity in this situation, it would appear that the influence of familiarity was decreasing, when in fact only recollection was increasing.

To compensate for this dependency, many researchers have advocated the use of the *independent remember/know method* (e.g., Jacoby, Yonelinas, & Jennings, 1997; Mangels, Picton, & Craik, 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995). In this method, whereas recollection is indexed by the proportion of "remember" responses, familiarity is measured as the proportion of "know" responses divided by the proportion of non-"remember" responses. Using this correction, researchers have demonstrated (as was mentioned above) that estimates of recollection and familiarity, as measured in the remember/know procedure, line up with estimates of recollection and familiarity as measured by other techniques (Yonelinas, 2002; Yonelinas & Jacoby, 1995; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). Hence, in Experiment 1, we used the independent remember/know method for calculating estimates of recollection and familiarity from "remember" and "know" responses. The goal of Experiment 1, once again, was to assess the degrees to which production affects recollection, familiarity, or both.

Method

Subjects A group of 27 students at the University of Waterloo received bonus credit in a course for taking part.

Stimuli The stimulus pool consisted of the 120 words in the Appendix of MacDonald and MacLeod (1998). The words were nouns from 5 to 10 letters long, with frequencies greater than 30 per million (Thorndike & Lorge, 1944). All stimuli were presented in 16-point lowercase Times New Roman font against a black background.

Apparatus The experiment was programmed in E-Prime version 1.2 (www.pstnet.com) and was carried out using a 17-in. color monitor and an IBM PC-compatible computer running Windows XP.

Procedure Subjects were asked to learn words for a later memory test, the nature of which was left unspecified. In the initial study phase, subjects studied a list of 80 words, 40 in blue and 40 in white, randomly intermixed and individually presented at the center of the screen. Subjects were instructed to read the words presented in blue aloud and the words presented in white silently. During study, a 500-ms blank screen preceded each word. If the word was in blue, the subject was to read it aloud into a microphone, and then a 500-ms blank interval followed; if the word was in white, it stayed on the screen for 2,000 ms before a 500-ms blank interval. Thus, the available study time for the words read silently was longer than that for the words read aloud, to which subject usually responded in under 700 ms. Despite this bias against the production effect, it still has been consistently observed (MacLeod et al., 2010).

A recognition test containing the 40 words studied aloud, the 40 words studied silently, and 40 new words immediately followed study. All words were presented in yellow font so that there was no color overlap for any subset between study and test. The 120 test words were presented in random order, and subjects were instructed to make a "new," "know," or "recollect" response for each test word by pressing labeled buttons ("N," "K," and "R") on the upper numeric row of the keyboard. The term "recollect" was used instead of the standard "remember" wording in order to avoid confusion due to the various meanings of the word "remember" outside of memory laboratories. The term "remember" will be used to refer to these responses in the results, however, to maintain consistency with the remember/know literature. There was a 500-ms blank interval before each word, and the word offset with the subject's key response.

Subjects were asked to make "recollect" responses when their recognition of the word was accompanied by episodic detail—by subjective awareness of recollection that the word had occurred in the study list. "Know" responses were to be made when they believed the word to have been studied but had no subjective awareness of it evoking any specific conscious recollection. "New" responses were to be made when the word was believed not to have appeared in the study list. Examples of each response type were given, and subjects were asked whether they understood the difference between "recollect" and "know" responses. If needed, clarification was provided, and subjects were asked to produce examples of each type of subjective state. Instructions were provided in sufficient detail as to be classified as conservative instructions, which Rotello et al. (2005) have shown to provide estimates of recollection and familiarity from remember/know paradigms that agree with estimates of recollection and familiarity as estimated from ROC curves.

Results and discussion

An alpha level of .05 was used as our criterion for significance in all significance tests. Effect size estimates were computed using partial eta-squared (η_p^2) or Cohen's d where appropriate. First, overall hit and false alarm rates were obtained by collapsing "remember" and "know" responses into "old" responses. The mean overall false alarm rate was .30 (SE = .04). The production effect was evident in the overall hit rates: Words spoken aloud at study had a mean hit rate of .76 (SE=.03), and words read silently at study had a mean hit rate of .59 (SE =.03); these two hit rates differed significantly, $t(26) = 5.56$, $d = 1.08$. Further attesting to the reliability of the production effect, this pattern of hit rate differences (aloud > silent) was evident for 25 of the 27 subjects (i.e., 93% of subjects).

Next, on the basis of the "remember" and "know" responses, measures of recollection and familiarity were calculated using the independent remember/know method (Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995). Recollection (R) was measured as the proportion of "remember" responses to an item set; familiarity (F) was measured as the proportion of "know" responses divided by the proportion of non-"remember" responses—that is, $F = p(\text{"know"})/[1 - p(\text{"remember"})]$. These estimates were calculated separately for each subject for words studied aloud, words studied silently, and new words.[1] These estimates of recollection and familiarity are shown in Fig. 1.

Recollection and familiarity estimates for aloud and silent words were analyzed in a 2 (aloud vs. silent) × 2 (R vs. F) within-subjects ANOVA. In general, familiarity estimates were higher than recollection estimates, $F(1, 26) = 29.15$, $MSE = 0.01$, $\eta_p^2 = .53$. More importantly, however, both recollection and familiarity estimates were higher for items studied aloud than for those studied silently, $F(1, 26) = 23.00$, $MSE = 0.01$, $\eta_p^2 = .47$, and the interaction was not reliable, $F(1, 26) = 1.35$, $MSE = 0.01$, $p = .26$, $\eta_p^2 = .05$. These results demonstrate that production increases memorability by increasing both recollection and familiarity. Thus, just as with previously studied, related encoding techniques such as elaboration, generation, and depth of processing, production appears to increase both recollection and familiarity.

[1] Note that we assume that R and F for new words represent false recollection/guessing and non-study-based familiarity/guessing, respectively.

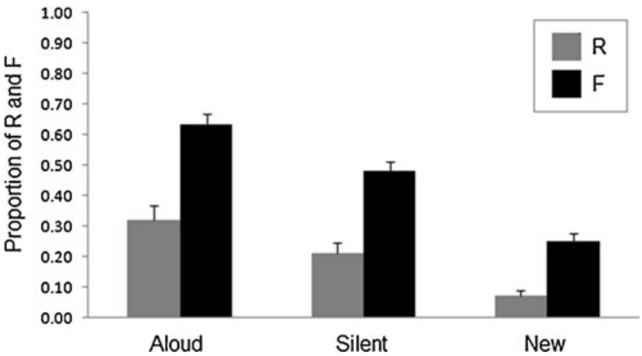

Fig. 1 Experiment 1: Mean proportions of recollection (*R*) and familiarity (*F*) responses. Error bars show the standard errors of the means. Note that *R* = *p*("remember") and *F* = *p*("know")/[1 − *p* ("remember")]. For aloud-studied, silent-studied, and new words, the mean *p*("know")s were .44 (*SE* = .04), .38 (*SE* = .03), and .22 (*SE* = .02), respectively

Experiment 2

In Experiment 1, improvements in both recollection and familiarity were observed for words read aloud versus words read silently, suggesting that production acts to enhance both processes. This result is consistent with past explanations of the production effect (see MacLeod et al., 2010) that have focused on the recollective benefit of production, but it further suggests that these accounts must incorporate the influence of familiarity. Before considering this point more thoroughly, however, we first wished to confirm the results of Experiment 1. That is, given that Experiment 1 was the first direct investigation of the influences of production on recollection and familiarity, a replication was warranted.

Although we have argued that the remember/know procedure is a valid procedure for measuring the influence of recollection and familiarity, it does make several key assumptions (e.g., that subjects can consciously differentiate between recollection and familiarity). We do not believe that these assumptions are invalid, and hence, we are confident in the results of Experiment 1. Nonetheless, replicating our results using a different methodology, one that makes different assumptions about recollection and familiarity, would be a powerful extension. Hence, in Experiment 2 we sought to provide converging evidence that production increases both recollection and familiarity by using a completely different methodology of gauging recollection and familiarity.

For Experiment 2, we used the ROC procedure and the dual-process signal detection model to estimate the influences of recollection and familiarity (Yonelinas, 1994, 1997). The ROC procedure does not involve subjects intentionally differentiating between recollection and familiarity. Instead, they are simply asked to respond on a 6-point confidence scale, ranging from *sure studied* to *sure new*, when recognizing words at test. From these data, 5-point ROC curves can be plotted and then fit using the dual-process signal detection model in order to obtain estimates of both recollection and familiarity (Yonelinas, 1994, 1997). Past work had demonstrated that the estimates of recollection and familiarity obtained using this procedure are consistent with estimates derived using the independent remember/know method (Yonelinas, 2002; Yonelinas & Jacoby, 1995; Yonelinas et al., 1998), so we expected Experiment 2 to replicate the findings of Experiment 1. The use of this alternative method nonetheless provided a powerful extension of Experiment 1: Using a completely different methodology from that of Experiment 1, we aimed to show that production does indeed increase both recollection and familiarity.

Method

Subjects A group of 32 students at the University of Waterloo received bonus credit in a course for taking part. One student's data were excluded from the study because that student failed to use all 6 points of the confidence scale, which resulted in a restriction of their ROC points.

Stimuli The stimulus pool consisted of 348 words drawn from the MRC (Medical Research Council) Linguistic Database (www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm). The words were nouns from 5 to 10 letters long, with Kučera and Francis (1967) frequencies between 30 and 847 and a mean frequency of 114.71 (*SD* = 120.59). For each subject, a subset of 240 words were randomly drawn from the stimulus pool. All stimuli were presented in 16-point lowercase Times New Roman font against a black background.

Apparatus The apparatus was the same as in Experiment 1.

Procedure The study phase was identical to that of Experiment 1, except that subjects studied a longer list of 120 words, 60 in blue and 60 in white. More words were used because ROC curves can be unreliable if too few observations are obtained per response category.

Following the study phase, the subjects were given a recognition test containing the 60 words studied aloud, the 60 words studied silently, and 120 new words. All words were presented in yellow font to prevent color overlap for any subset between study and test. The 240 words were presented in random order, and subjects were instructed to make a confidence judgment from 1 to 6 for each word. They were instructed to press 4, 5, or 6 if they believed a word to have been studied—6 if they were *absolutely sure*, 5 if they were *very sure*, and 4 if they were *somewhat sure*. Similarly, they were instructed to press 1, 2, or 3 if they believed a word to be new. Subjects pressed 1 if they were *absolutely sure*, 2 if they were *very sure*, and 3 if they were *somewhat sure*. Subjects were asked to think carefully about their confidence ratings and to try to use the entire scale over the course of the recognition test.

Results and discussion

ROC analysis involves plotting hit rates against false alarm rates across various response criteria. ROC curves were plotted separately for words studied aloud and studied silently; these are shown in Fig. 2. In ROC analyses, the first data point is obtained by considering only the strictest response criterion (i.e., only "6" responses are taken to be hits or false alarms). For both read-aloud and silently studied words, then, the hit rates at this level of confidence were plotted at the level of false alarms. These points are plotted as the two far left points in Fig. 2 (i.e., the first empty and filled circles). The next point was obtained by considering a slightly less strict criterion (i.e., either "5" or "6" responses are taken to be hits or false alarms). For both aloud- and silent-studied words, the hit rates at this level of confidence were plotted at the level of false alarms and can be seen as the second two points from the left in Fig. 2. This process was repeated until the most lax criterion had been plotted (i.e., when a response from "2" to "6" was taken to be a hit or a false alarm). ROC analysis does not plot a point where responses from "1" to "6" are considered to be hits or false alarms because these points would always add to 1.0, and thus would provide no useful information. Hence, from a 6-point confidence scale, 5-point ROC curves can be plotted.

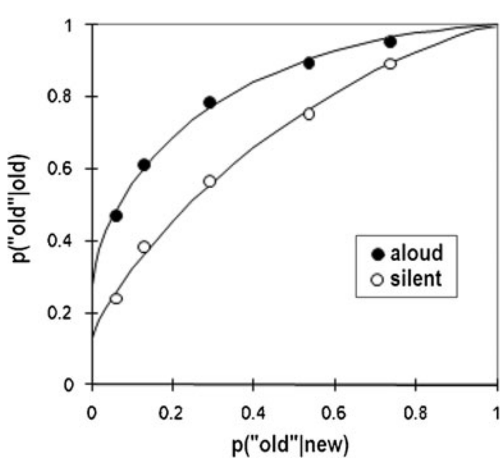

Fig. 2 Experiment 2: The recognition receiver operating characteristics (ROCs) for aloud- and silent-studied words. The displayed curves were fit using the dual-process signal detection model. The model provided an accurate account of the observed ROCs for both aloud- and silent-studied words

To test whether an overall production effect was present, hit rates for read-aloud and silently studied words were obtained by collapsing the "4," "5," and "6" responses. That is, subjects had previously been told to press 4, 5, or 6 if they believed that a word was studied, therefore we treated each of these as an "old" response and calculated hit rates for aloud- and silent-studied words. Overall, a production effect was observed, with greater hits to aloud (M = .76, SE = .02) than to silent words (M = .55, SE = .03), $t(31)$ = 9.14, d = 1.64, and once again the production effect in hit rates was reliable (aloud > silent) for 31 of the 32 subjects (i.e., 97% of subjects).

The dual-process signal detection model can be used to fit ROC curves and yields estimates of both recollection and familiarity (Yonelinas, 1994, 1997). This model has been successfully used to obtain measures of recollection and familiarity that can be compared to the output of remember/know procedures (Yonelinas & Jacoby, 1995; Yonelinas et al., 1998). Although alternative models for analyzing ROC curves exist (e.g., the unequal-variance signal detection model; see Wixted, 2007), our goal here was not to evaluate different models but simply to use a relatively standard model to obtain estimates of recollection and familiarity. Thus, we used the dual-process signal detection model to obtain these measures from our data so that they could be compared with those from Experiment 1. We will return to the issue of alternative analysis methods of our ROC data in the General Discussion.

The dual-process signal detection model assumes that recollection is a threshold process, and therefore estimates of recollection (R) are provided on a probability scale, the same as in the remember/know procedure. However, the model further assumes that familiarity is best described as a signal detection process, not a threshold process. In other words, the probability of accepting an item as "old" on the basis of familiarity is a function of how much more familiar studied items are relative to new items. Hence, the model produces estimates of familiarity in terms of d', not of raw probability. Although in Experiment 1 familiarity estimates were dealt with in terms of raw probability and, hence, are not analogous to the familiarity estimates derived here, it is a trivial matter to convert raw probabilities for old and new items into d' measures. Therefore, the familiarity estimates from Experiment 1 were converted into d's so that they could be compared directly to the values from Experiment 2. The dual-process signal detection model was therefore fit to the ROC curves for read-aloud and silent words separately, to yield separate estimates of recollection and familiarity for the two conditions. These estimates are

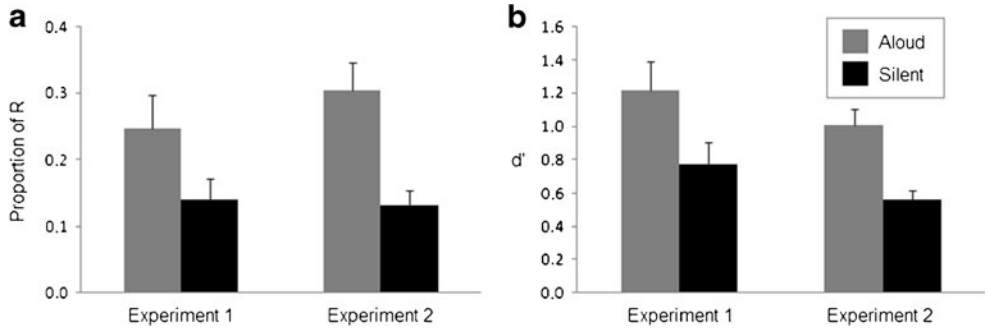

Fig. 3 Experiments 1 and 2: Recollection (*R*) and familiarity (*d*′) estimates for aloud- and silent-studied words. Error bars show the standard errors of the means. (**a**) In Experiment 1, *R* = *p*(“remember”|aloud) – *p*(“remember”|new) for read-aloud words, and *R* = *p*(“remember”|silent) – *p*(“remember”|new) for silent-studied words. This measure of *R* is equivalent to that derived for the aloud and silent words in Experiment 2 using the dual-process signal detection model on the ROC curves. (**b**) For Experiment 1, familiarity was measured as *d*′, which is the difference in *z*-space between *F* responses for aloud (or silent) words and new words. This measure is equivalent to the *d*′ measure obtained for aloud and silent words in Experiment 2 by using the dual-process signal detection model on the ROC curves

plotted in Fig. 3, along with the corresponding recollection and familiarity estimates recalculated from Experiment 1.

The dual-process signal detection model was fit to ROC curves using Yonelinas’s DPSD solver (available from http://psychology.ucdavis.edu/Labs/Yonelinas/PWT/). The solver allows hit and false alarm rates at five levels of confidence to be entered, which places five points on a ROC graph. Hit and false alarm rates were entered separately for each subject. Subsequently, using the Excel data solver, r_o (i.e., recollection for studied items), *d*′, and the criteria at each level of confidence could be solved for. This involves the Excel solver changing these parameters, which changes the ROC function, until a ROC curve is obtained that passes through or near the five points on the ROC graph. The solver attempts to find the best fit for the curve by minimizing the sum-squared error of the points from the curve. After the ROC curve is fit, r_o is used as our estimate of recollection and *d*′ as our estimate of familiarity for that particular subject. This process is repeated until all estimates have been obtained for all subjects.

Because recollection and familiarity are measured on different scales in ROC analyses (i.e., probability vs. *d*′), direct comparisons between recollection and familiarity were not performed. Instead, recollection and familiarity were analyzed separately between Experiments 1 and 2: Separate 2 (aloud vs. silent) × 2 (Exp. 1 vs. 2) mixed ANOVAs were conducted for *R* and *d*′. In terms of recollection, *R* estimates were greater for aloud than for silent words, $F(1, 56) = 31.44$, $MSE = 0.02$, $\eta_p^2 = .36$, and this effect did not interact with experiment, $F(1, 56) = 1.74$, $MSE = 0.02$, $p = .19$, $\eta_p^2 = .03$, nor was there a main effect of experiment, $F < 1$. Hence, in Experiment 2 as in Experiment 1, a recollection advantage was observed for words read aloud at study.

Turning to familiarity, *d*′ estimates were also greater for aloud than for silent words, $F(1, 56) = 49.25$, $MSE = 0.12$, $\eta_p^2 = .47$, and this effect did not interact with experiment, $F < 1$, nor was there a main effect of experiment, $F(1, 56) = 1.91$, $MSE = 0.66$, $p = .17$, $\eta_p^2 = .03$. Hence, Experiment 2 also replicated Experiment 1 in observing a familiarity advantage for words read aloud at study. Although there might have been a trend toward greater *d*′ estimates in Experiment 1, paired-samples *t* tests found no difference between experiments for the *d*′s of either aloud or silent words, $t(56) = 0.28$, $p = .82$, $d = 0.08$, and $t(56) = 0.11$, $p = .82$, $d = 0.03$, respectively. Of course, even had such an effect been significant, all it would signify was that Experiment 1 was easier than Experiment 2. And given that 50% more words were studied in Experiment 2 (to obtain reliable ROC curves), this would not be surprising, nor would it undermine the relative difference that was found between aloud and silent words.

Thus, Experiment 2 demonstrated both that production increases recollection and familiarity, and that the remember/know procedure provides estimates of recollection and familiarity that are very much in line with those derived from ROC curves using the dual-process signal detection model. Once again, the strength of our findings here rests on the fact that Experiments 1 and 2 used very different methodologies for estimating the influences of recollection and familiarity, yet they produced almost identical results. For this reason, we are confident in asserting that production does indeed enhance both recollection and familiarity.

Experiment 3

Across two experiments, using two very different methodologies, we have demonstrated that production leads to an enhancement of both recollection and familiarity. Experiment 3 sought to assess the recollective benefit of production using a different approach. Previous work on the remember/know procedure has shown that items that are remembered are often accompanied by specific contextual details, whereas items that are merely known often lack such details (Perfect, Mayes, Downes, & Van Eijk, 1996). Although subsequent work has shown that known items are accompanied by above-chance levels of contextual details, these levels are still below those of recollected items and are not accompanied by a subjective sense of mentally reliving those details (Eldridge, Engel, Zeineh, Bookheimer, & Knowlton, 2005; Hicks, Marsh, & Ritschel, 2002; Wixted, 2007). In our experiments, we have found that words read aloud at study show a greater degree of recollection than do words read silently at study. Hence, it would be reasonable to expect that words that had been read aloud would more often be accompanied at the time of retrieval by contextual details than would words that had been read silently.

Experiment 3 sought to test this expectation by examining the degree to which subjects can accurately identify words at test as "aloud" or "silent." Because words spoken aloud demonstrated greater recollection than did words read silently in Experiments 1 and 2, there should be better aloud/silent differentiation for aloud words than for silent words. That is, the greater recollective estimates should indicate that subjects have more recollective detail for aloud than for silent words. Subjects should therefore be more likely to remember contextual or item information about aloud words that would then permit the proper identification of study modality. Hence, subjects should be relatively likely to call words said aloud "aloud" and relatively unlikely to call them "silent." Silently read words, because they are less recollectable than spoken-aloud words, should be accompanied by less contextual or item detail. Lacking such information, subjects should have more difficulty in correctly identifying silently read words as "silent" versus "aloud." Hence, the difference between "aloud" and "silent" judgments should be smaller for words read silently than for words read aloud. To address this prediction, in Experiment 3 we asked subjects at the time of test to make judgments about which encoding condition an item had appeared in.

Before continuing, it should be noted that when examining source discrimination, it is relatively common to conditionalize source responses on hit rates—for example, by dividing the proportion of "aloud" responses to old aloud words by the total hit rate for aloud words. In our analyses, we will perform this conditionalization on responses that have been corrected for false alarm rates. Correcting for false alarm rates is not common practice in source discrimination; however, in most source discrimination experiments, the false alarm rates among different sources are comparable and therefore this practice is not needed. In our data, a bias will be observed in the false alarms, whereby subjects are more likely to false alarm with a "silent" than with an "aloud" response (see Dodson & Schacter, 2001, for a similar finding). In this scenario, conditionalizing responses on hit rates without taking into account false alarm rates would distort our data, as it would not take into account the base-rate differences in responses. For example, if subjects were saying "silent" very often to new items, we would necessarily expect them to say "silent" very often to silent studied items as well, even if they do not remember those studied items whatsoever. Hence, the absolute "silent" hit rate is not informative. What is informative is the difference between the "silent" hit rate and the "silent" false alarm rate, with a greater difference indicating stronger source memory. False alarm rates were therefore subtracted from hit rates in our analyses, before any conditionalization was performed. Beyond that first step, however, the analyses performed were typical for source discrimination.

Method

Subjects A group of 26 students at the University of Waterloo received bonus credit in a course for taking part in the experiment.

Stimuli and apparatus These were the same as in Experiment 1.

Procedure The procedure was identical to that of Experiment 1, except that instead of making "remember," "know," and "new" judgments, subjects made "aloud," "silent," and "new" judgments. That is, subjects were instructed to choose either "aloud" or "silent" when they believed that a word had been studied, and to make their response on the basis of the believed modality of study. If subjects believed that a word was studied aloud, they were to select "aloud"; if they believed that a word was studied silently, they were to select "silent." Responses were made by pressing labeled buttons ("A," "S," and "N") on the upper numeric row of the keyboard.

Results and discussion

Overall hit and false alarm rates were obtained by collapsing "aloud" and "silent" responses into "old" responses. The mean overall false alarm rate was .29 (*SE* = .03). The production effect was strongly evident in the overall hit rates. Words spoken aloud at study had a mean hit rate of .81 (*SE* = .02), whereas words read

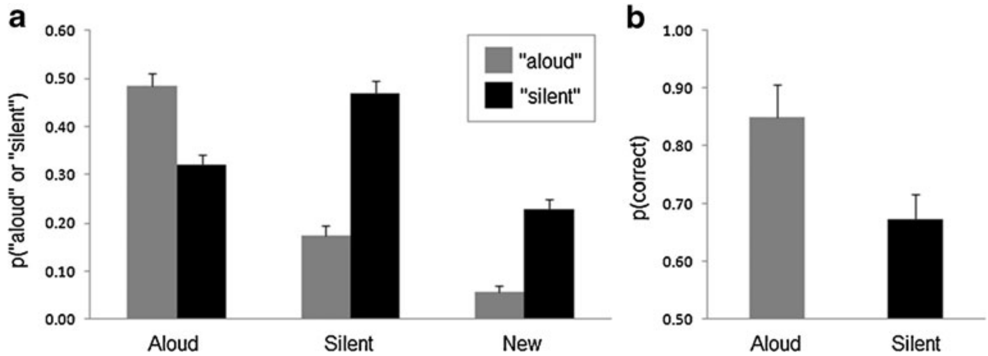

Fig. 4 Experiment 3: Mean modality judgments for words studied aloud, words studied silently, and new words (**a**), as well as adjusted accuracy scores (**b**). Error bars show the standard errors of the means. Note that .50 represents chance performance for accuracy in panel B

silently at study had a mean hit rate of .64 (*SE* = .03); these two hit rates differed significantly, $t(25) = 7.14$, $d = 1.46$. Further attesting to the reliability of the production effect, this pattern of hit rate differences (aloud > silent) was seen for 24 of the 26 subjects (i.e., 92% of the subjects) in Experiment 3.

The unadjusted "aloud" and "silent" responses can be seen in Fig. 4A. Upon first glance, it appears as if subjects called silent words "silent" and aloud words "aloud" to the same degree, and thus one might conclude that source memory was equivalent for aloud and silent words (or, if "aloud" and "silent" responses were conditionalized on hit rates, one might even conclude that source memory was better for the silent than for the aloud words). However, it needs to be stressed here that these two measures cannot be directly compared or interpreted in this manner, because false alarm rates were not equal for "aloud" and "silent" responses. That is, turning to "aloud" and "silent" responses for new items, we see that when subjects are assigning the source for an item that was never studied, they are more likely to respond "silent" than "aloud," $t(34) = 7.13$, $d = 1.24$. Thus, when subjects actually have no memory for items, they would be expected to produce many "silent" responses. Therefore, the finding that subjects said "silent" more than "aloud" to silent words might not indicate strong source memory at all, but merely reflect a response bias toward saying "silent" in general. To correct for this response bias, then, the false alarm rates were first subtracted from the hit rates for both aloud and silent words, yielding adjusted "aloud" and "silent" responses, which could be summed to obtain adjusted hit rates. This procedure could be considered the mathematical equivalent of setting the corresponding false alarm rates of the hit rates to zero.

Next, these adjusted "aloud" and "silent" responses were conditionalized on adjusted hit rates to obtain correct-response scores (see Fig. 4B). That is, simply looking at adjusted "aloud" or adjusted "silent" responses in our data could be misleading unless those responses were conditionalized on the adjusted overall hit rates. For example, if a subject called aloud words "aloud" 30% of the time but called aloud words "silent" 30% of the time, then in actuality they would only be responding "aloud" in 50% of their "old" responses (i.e., in 50% of the adjusted hit rates, subjects correctly identified the source of aloud words). However, if another subject called aloud words "aloud" 30% of the time but called aloud words "silent" only 10% of the time, this subject would have responded "aloud" in 75% of their "old" responses (i.e., in 75% of their adjusted hit rates). Thus, although in both cases aloud words were called "aloud" 30% of the time, the second case represents stronger source memory than the first. Consequently, adjusted accuracy scores for aloud words were calculated by dividing the adjusted proportion of "aloud" responses by the adjusted hit rates for aloud words. Similarly, adjusted accuracy scores for silent words were calculated by dividing the adjusted proportion of "silent" responses by the adjusted hit rates for silent words. These conditionalized scores are presented in Fig. 4B as the probability of correct source identification.

Due to the nature of conditionalizing, the number of incorrect responses was always linearly related to the number of correct responses [i.e., $p(\text{incorrect}) = 1 - p(\text{correct})$]. As a result of this dependency, incorrect-response rates were disregarded, and only correct responses were compared between aloud and silent words. Subjects were able to accurately identify modality more often than chance (i.e., 50%) for both aloud and silent words, $t(25) = 6.20$, $d = 1.22$, and $t(25) = 3.99$, $d = 0.78$, respectively. However, most importantly, subjects did indeed make more accurate modality judgments for aloud than for silent words, $t(25) = 2.27$, $d = 0.46$.

When subjects were asked to identify the modality of study, they were better able to accomplish this for words that were read aloud than for those that were not read aloud. Given that Experiments 1 and 2 demonstrated that words read aloud at study were also more recollectable than words read silently at study, the results of Experiment 3 are in line with past work that has shown that recollected items are accompanied by more contextual and item detail than are items that are not recollected (Eldridge et al., 2005; Hicks et al., 2002; Perfect et al., 1996; Wixted, 2007).

Finally, the tendency to false alarm with "silent" more than with "aloud" here was similar to the findings of Dodson and Schacter (2001). Those researchers found that subjects were more likely to false alarm to new items that were related to studied items if the studied items had all been heard at study. If the studied items had all been read aloud at study, the false alarm rate was reduced. This finding was observed both between and within subjects and was interpreted in the context of a distinctiveness heuristic, whereby subjects who had read words aloud demanded more distinctive information at test—specifically, some record of having said a word aloud—before they would endorse it as a studied word. The authors argued that this effect provides evidence that reading words aloud enhances recollective distinctiveness.

By this argument, then, the false alarm rate effects we have observed here could be viewed as providing further evidence that production enhances recollection. That is, whereas subjects would have false alarmed "silent" so long as a new item felt familiar, they would require a recollective record of having said a word aloud before they would false alarm "aloud." Because such false recollections for new items should be rare to nonexistent, "aloud" false alarms would be expected to be at floor, and they were. The false alarm effects observed are therefore consistent with the work of Dodson and Schacter (2001) and the claim that production enhances recollection. Although this result alone would not make a compelling case, we have now shown across three experiments, using a variety of methods, that production enhances both familiarity and recollection.

General discussion

The production effect is the simple finding that reading some words aloud and other words silently at study leads to a consistent memory boost for the words read aloud (Hourihan & MacLeod, 2008; Lin & MacLeod, 2011; MacLeod, in press; MacLeod et al., 2010; Ozubko & MacLeod, 2010). The present work demonstrates that this memory enhancement occurs due to boosts from production in both recollection and familiarity. Indeed, using both the remember/know procedure (Exp. 1) and the ROC procedure (Exp. 2), we have demonstrated a consistent recollective-based and familiarity-based advantage for words read aloud versus those read silently at study. These procedures are very different from each other, making quite different assumptions, yet both yielded the exact same results: Production enhanced both recollection and familiarity. Furthermore, Experiment 3 provided converging evidence for the recollective advantage of words read aloud, by demonstrating that subjects were better able to identify qualitative information about words read aloud—specifically, the form of encoding used at study. Taken together, then, these experiments are consistent with previous research that has supported an account of the production effect hinging on recollection and distinctiveness. Critically, these are the first experiments to demonstrate that production benefits both recollection and familiarity.

The possibility of a single-process account

So far, we have focused on interpreting our results from a dual-process perspective. Indeed, the results of Experiment 2 were even fit using a dual-process model, despite the fact that a single-process model could have been adopted. Although we have previously cited evidence that is inconsistent with a single-process strength interpretation of the production effect (Hopkins & Edwards, 1972; MacDonald & MacLeod, 1998; MacLeod et al., 2010; Ozubko & MacLeod, 2010), it remains a possibility that a single-process account could have been used to explain the data reported in this article. Specifically, if it is assumed that "remember" and "know" responses do not reflect a qualitative difference in memory (in line with single-process accounts), then they must represent a quantitative difference along a single dimension of memory. A single-process account would therefore need to posit two criteria along a single dimension to give rise to both "remember" and "know" responses. The results of Experiment 1 are consistent with such a model, so long as it is predicted that aloud items are represented more strongly in memory than are silent items. Experiment 3 could similarly be viewed within such a single-process strength framework, if it is assumed that identification of a word as "aloud" correlates with memory strength for that word. Indeed, few "aloud" false alarms were observed in Experiment 3, which is consistent with the claim that "aloud" responses merely correlate with memory strength. Hence, Experiment 3 is not incompatible with a strength view of the production effect.

More informative, however, are the results of Experiment 2. A formal single-process alternative to the dual-process signal detection model used in Experiment 2 is the unequal-variance signal detection (UVSD) model (Wixted, 2007). Although the UVSD model has been criticized on the basis of the results of behavioral studies (see Yonelinas & Parks, 2007, for a review), it nonetheless represents an alternative single-process account that can be fit to ROC data. Re-examining the results of Experiment 2 using the UVSD, we see some consistency with our previous conclusions. Namely, d' was lower for words read silently

(M = 0.85, SE = 0.09) than for words read aloud (M = 1.86, SE = 0.16), $t(30) = 6.78$, $d = 1.36$, $p < .01$, indicating that memory strength (or "familiarity") was less for silent- than for aloud-studied words. The second component of UVSD is the ratio of the variances of the "old" and "new" distributions. Here, words studied aloud produced a more variable old distribution (M = 1.49, SE = 0.11) than did words studied silently (M = 1.24, SE = 0.07), $t(30) = 2.64$, $d = 0.51$, $p < .05$. Thus, in some sense, the results from the UVSD model parallel our previous findings, in that both components are changing, despite assuming only one strength dimension underlying memory. Indeed, the only real difference between the results of these two models is that the dual-process signal detection model would claim that the increased variance in the old distribution is due to high-confidence recollected items, whereas the UVSD would claim that it is due to the encoding variability of the studied items. Otherwise, the two accounts tell very similar stories.

Interestingly, however, if we were to believe in the single-process framework, Experiment 1 could also be fit to the UVSD model. That is, as described previously, a single-process account would need to assume that "remember" and "know" responses arise as a result of two different criteria. If this is so, then "remember," "know," and "new" responses should be more-or-less equivalent to 3-point confidence ratings (see Donaldson, 1996). From this perspective, two-point ROC curves could be plotted for each subject and fit to an UVSD model, the same as was done for the data from Experiment 2.

Fitting the results of Experiment 1 to the UVSD, we see that, as in Experiment 2, d' was lower for words read silently (M = 1.06, SE = 0.16) than for those read aloud (M = 1.65, SE = 0.20), $t(26) = 3.65$, $d = 0.72$, $p < .01$. However, unlike in Experiment 2, no difference between the variance ratios could be observed between words read aloud (M = 1.22, SE = 0.16) and words studied silently (M = 1.21, SE = 0.14), $t(26) = 0.032$, $d = 0.01$, $p = .97$. Thus, the results of the UVSD analysis are inconsistent across experiments. In Experiment 1, the UVSD describes the aloud and silent distributions as equally variable, but in Experiment 2, the UVSD describes the aloud distribution as more variable than the silent distribution. A single-process account would only be able to reconcile these results if a coherent reason was put forth for why the underlying distributions would change across Experiments 1 and 2. However, the only definite change across these two experiments, from a single-process perspective, was that the confidence scale had 3 points in Experiment 1 and 6 points in Experiment 2. It is not clear why this manipulation at test would affect distributions of the aloud or silent items, as the UVSD presumes that the old distributions' variability arises at encoding, as a result of encoding variability (Wixted, 2007).

Preference for the dual-process account

The inconsistency of the UVSD across experiments represents an additional reason to adopt a dual-process interpretation of the production effect over a single-process strength explanation and, indeed, provides us with a basis for rejecting a single-process account for the data reported here. It remains possible that the methodological differences between Experiments 1 and 2 could have contributed to the different conclusions of the UVSD across experiments; however, given the identical encoding conditions between Experiment 1 and 2, it is hard to see exactly why a difference would have arisen if that approach were valid. Instead, a simpler conclusion is that the dual-process signal detection model and the dual-process interpretation of the remember/know paradigm indeed provide a more coherent account of our data than does a single-process approach. And, importantly, according to a dual-process account, production affects both recollection and familiarity.

To date, explanations of the production effect have focused on the recollective advantage of production. However, as we have shown here, production also acts to increase familiarity. Interestingly, in a review of phenomena that affect recollection and familiarity, Yonelinas (2002) observed that other mnemonic techniques, such as levels of processing, generation, and focused attention (as opposed to divided attention), tend to increase both recollection and familiarity but have a larger impact on recollection. In Experiment 1, production appears to increase both recollection and familiarity, but to similar degrees. Thus, perhaps production has less of a recollective advantage than do levels of processing, generation, and focused attention. This is a reasonable possibility, in the sense that production is so much simpler and easier than more elaborate encoding strategies. It might, therefore, be expected that production would be less effective in terms of boosting recollection than are other, more elaborate strategies. That is, if levels of processing and generation produce large recollective benefits because they produce elaborate mental images or memories regarding the details of semantic generation, then production might be expected to produce a smaller recollective benefit, since the details of vocalization are likely less distinctive than mental imagery or semantic generation. However, before making too much of the size of the recollective benefit of production, it should be noted that this is the first empirical investigation of the influences of production on recollection and familiarity; whether this pattern will hold up across future experiments remains to be seen.

Whether or not production boosts recollection and familiarity to similar degrees, it nonetheless does increase both recollection and familiarity. This finding is hardly surprising, because work with levels of processing and generation has shown similar qualitative enhancements of both recollection and familiarity (see Yonelinas, 2002, for a review). However, our finding does raise the question of *why* production enhances familiarity. At this point, our best suggestion is that production may indeed have an attentional component, an idea that was at the core of the explanation offered by MacDonald and MacLeod (1998). That is, words that are read aloud may seem more important to subjects, and hence, subjects focus more on those words at study. For words read silently at study, subjects do not pay as much attention to encoding those items, and this contributes to the mnemonic difference between aloud and silent words.

If production has some attentional component, then this component could underlie the familiarity boost seen in our experiments. We would still have expected to see a production effect when words are encoded deeply or generated (see MacLeod et al., 2010), because production could still provide distinctive information regarding these items that could later be recollected. However, when both aloud and silent words were encoded deeply or generated, the familiarity component of the production effect might have been eliminated. Hence, that production acts to increase both recollection and familiarity may indicate that there are two processes in play that give rise to the production effect. Although this issue must remain speculative at the moment, it will no doubt become a more central focus to theorizing about production as more research on the recollective- and familiarity-based influences of production becomes available.

In sum, reading words aloud is an effective method to improve memory performance. It both increases hits and decreases false recognitions as compared to reading words silently. MacLeod et al. (2010; see also Hourihan & MacLeod, 2008; Lin & MacLeod, 2011; MacLeod, in press; Ozubko & MacLeod, 2010) have provided evidence that having read a word aloud becomes part of the encoding record, making those words more distinctive against the backdrop of words not read aloud. MacLeod et al. have also proposed that the conscious recollection of this distinctive encoding record provides for the enhancement stemming from production. Here we have shown that conscious recollection does indeed play a role in the production effect. We have also found clear evidence that production further increases familiarity, as well. Such a finding is consistent with other mnemonic phenomena, such as levels of processing and generation, that also enhance both familiarity and recollection. Indeed, the present work provides a clear demonstration that production might be a valuable mnemonic technique, as much so as the predominant techniques commonly in use today.

Author note This research was supported by Discovery Grant A7459 from the Natural Sciences and Engineering Research Council of Canada. We thank Domenica De Pasquale and Keehan Bailey for assistance with data collection in Experiment 1, and Bri Lock for assistance with data collection in Experiment 2. Additionally, we thank Mariam Aly for helpful discussions regarding the ROC data. N.G. is now at IBM, New York, New York.

References

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 553–563. doi:10.1037/0278-7393.13.4.553

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

Bower, G. H., & Clark, M. C. (1969). Narrative stories as mediators for serial learning. *Psychonomic Science, 14,* 181–182.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22,* 267–281.

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

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

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology. General, 104,* 268–294. doi:10.1037/0096-3445.104.3.268

Dodson, C. S., & Johnson, M. K. (1996). Some problems with the process-dissociation approach to memory. *Journal of Experimental Psychology. General, 125,* 181–194.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8,* 155–161. doi:10.3758/BF03196152

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30,* 123–152.

Eldridge, L. L., Engel, S. A., Zeineh, M. M., Bookheimer, S. Y., & Knowlton, B. J. (2005). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience, 25,* 3280–3286.

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

Engelkamp, J., & Krumnacker, H. (1980). Imaginale und motorische Prozesse beim Behalten verbalen Materials [Imagery and motor processes in memory of verbal material]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 27,* 511–533.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2011). *Widening the boundaries of the production effect*. Manuscript submitted for publication

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313. doi:10.3758/BF03197041

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

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 503–508.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. *Memory & Cognition, 25,* 345–351.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior, 11,* 534–537.

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology, 62,* 242–246.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). Oxford: Oxford University Press.

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition, 26,* 299–308.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541. doi:10.1016/0749-596X(91)90025-F

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Hillsdale: Erlbaum.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8,* 283–293.

Khoe, W., Kroll, N. E. A., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes–no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia, 38,* 1333–1341.

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

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

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence: Brown University Press.

Lin, O., & MacLeod, C. M. (2011). *Aging and the production effect: A test of the distinctiveness account*. Manuscript submitted for publication

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.

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.

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11,* 77–95.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science, 17,* 239–243. doi:10.1111/j.1467-8721.2008.00582.x

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology. General, 129,* 242–261.

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.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology, 49A,* 797–813.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102. doi:10.3758/BF03211168

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12,* 865–873. doi:10.3758/BF03196778

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia, 45,* 2163–2179.

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

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College.

Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels-of-processing effects on familiarity. *Canadian Journal of Experimental Psychology, 50,* 123–138.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12. doi:10.1037/h0080017

Verfaellie, M., & Treadwell, J. R. (1993). Status of recognition memory in amnesia. *Neuropsychology, 7,* 5–13. doi:10.1037/0894-4105.7.1.5

Wixted, J. T. (2007). Signal-detection theory and the neuroscience of recognition memory. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 67–82). New York: Psychology Press.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117,* 1025–1054. doi:10.1037/a0020874

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11,* 616–641. doi:10.3758/BF03196616

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. doi:10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology. General, 130,* 361–379. doi:10.1037/0096-3445.130.3.361

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517. doi:10.1006/jmla.2002.2864

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5,* 418–441. doi:10.1006/ccog.1996.0026

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34,* 622–643. doi:10.1006/jmla.1995.1028

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember–know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 323–339. doi:10.1037/0894-4105.12.3.323

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133,* 800–832. doi:10.1037/0033-2909.133.5.800

Zimmer, H. D., Cohen, R. L., Guynn, M. J., Engelkamp, J., Kormi-Nouri, R., & Foley, M. A. (2001). *Memory for action: A distinct form of episodic memory?* New York: Oxford University Press.