Canadian Journal of Experimental Psychology
2010, Vol. 64, No. 4, 227–240

1196-1961/10/$12.00 DOI: 10.1037/a0021699

HEBB AWARD WINNER 2010 / GAGNANT DU PRIX HEBB 2010

When Learning Met Memory

Colin M. MacLeod
University of Waterloo

The first sentence of the introduction to Hebb's (1949) classic monograph, *The Organization of Behavior*, is "It might be argued that the task of the psychologist, the task of understanding behaviour and reducing the vagaries of human thought to a mechanical process of cause and effect, is a more difficult one than that of any other scientist" (p. xi). Nowhere is this more true than in the realm of human learning and memory, given our truly remarkable ability to acquire and retain prodigious amounts of information. This article is divided into two parts. The first part sketches my lifelong fascination with learning that led me to study first memory, then attention, and then their interplay, with examples of a few interesting findings along that path. The second part details recent work in my laboratory exploring a simple yet quite powerful encoding technique: Saying things aloud improves memory for them. This benefit, which we call the *production effect*, likely occurs by enhancing the distinctiveness of the things said aloud, and may constitute a beneficial study method. Understanding how we learn and remember is ultimately a crucial step in understanding ourselves.

Keywords: memory, attention, learning, production, distinctiveness

I am honoured—and absolutely delighted—to be the recipient of the 2010 Donald O. Hebb Distinguished Contribution Award from the Canadian Society for Brain, Behaviour, and Cognitive Science. As a Canadian, a Hebb student at McGill, and a "charter member" and former President of the Society, this has very special meaning to me. I am particularly pleased that this occasion also gives me an opportunity to recognise and thank many people without whom I certainly would not have received this tribute.[1]

In my undergraduate years at McGill University (1966–1971), I had no idea at the time what a privilege it was to have taken introductory psychology from Donald Hebb (and Ronald Melzack, Peter Milner, and Muriel Stern—a truly amazing line-up); indeed, by the time I graduated, Hebb had become Chancellor of the university, and it was he who "capped" me at convocation. My teachers at McGill, among them Don Donderi, who hired me as a research assistant in the summer after my third year, and Mike Corballis, who taught my first attention and memory course in my fourth year, were terrific, somehow seeing through my shyness to my emerging captivation with psychological research.

My introduction to actual cognitive research came in the summer of 1970 when Don Donderi hired me as a research assistant to help visiting scientist Yuji Baba carry out some studies on stabilized retinal images (see Pritchard, Heron, & Hebb, 1960). I built a dark adaptation room and constructed stimulus cards and did many other tasks that introduced me to careful methodology, all the while realising how much I was enjoying the work and looking forward to finding out what would happen in the project. The goal was to examine how meaningful Japanese Kanji characters versus nonmeaningful but equivalently complex artificial characters broke down for Japanese-speaking versus non-Japanese-speaking people when the characters were stabilized. [My recollection is that the characters broke down in such a way as to preserve meaning only when the real Kanji characters were viewed by

Colin M. MacLeod, Department of Psychology, University of Waterloo.
Most of the research reported here was supported by discovery Grant A7459 from the Natural Sciences and Engineering Research Council of Canada (NSERC), as was the preparation of this article. I thank Ramona Bobocel, Michael Masson, and Jason Ozubko for their comments on the previous draft. And "grazie mille" to Vince Di Lollo for nominating me for this award—that he saw me as worthy is an honour in itself. This article is dedicated to my mother, Anne MacLeod, and to the memory of my father, Clarence MacLeod, for their love and encouragement, and for their unswerving belief in me and in science and education.
Correspondence concerning this article should be addressed to Colin M. MacLeod, Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1. E-mail: cmacleod@uwaterloo.ca

[1] To my colleagues, both faculty and staff, at the University of Toronto, especially at Scarborough (1978–2003), and at the University of Waterloo (2003–present), thank you all for your support and friendship, and for the many things I have learned from you. I wish to single out one colleague at U of T at Scarborough—Bert Forrin—as my true mentor, the very essence of scholar and friend. At Waterloo, I have the great good fortune of having my wife, Ramona Bobocel, as a colleague, a wonderful sounding board, and my best friend. I cannot possibly mention by name all of my terrific collaborators, but I thank them all. One deserves special thanks because of our longstanding and productive series of projects, and as a career-long friend and advisor—Mike Masson, at the University of Victoria.

To the many undergraduate research and thesis students and to the graduate students who have made research always intriguing and fun, thank you for trusting me to be your supervisor and for your myriad contributions to my research. The older I get the more I believe that the most important research contribution I have made is through teaching, by fostering your research interests and careers. Thank you for that incredible opportunity, and may you all enjoy your careers and find them as fulfilling as I continue to find mine.

native Japanese speakers, but that could easily be a false memory based on my recollection of the hypothesis!] Sometime near the end of that summer, I told one of the graduate students with whom I shared an office how excited I had been to get the job; he smiled and told me that I had been the only applicant. And so began my research career.

When BA Met PhD

In graduate school at the University of Washington (1971–1975)—which I carefully selected on the basis of “adventure quotient,” it being the farthest away from Montréal—I was fortunate indeed to be Tom Nelson’s first graduate student, and to have Geoff Loftus and Earl (Buz) Hunt as my other committee members. From Tom I learned the significance of careful empirical work, from Geoff the centrality of theory, and from Buz the importance of the “big picture.” I had other wonderful teachers in graduate school, too, among them Bob Bolles and Beth Loftus. After a year away at the University of California in Isabel Birnbaum and Betsy Parker’s alcohol lab, I returned to Seattle where my postdoc (1975–1978) with Buz Hunt also enriched my appreciation for key elements of cognition (like intelligence and individual differences) that are too often neglected in mainstream cognitive psychology.

My first project at Washington was an exploration of the then new directed forgetting paradigm (see Bjork, 1972; MacLeod 1998a). At the time, there were two competing accounts of why instructions to forget actually worked, leading to poorer retention of to-be-forgotten [F] than to-be-remembered [R] information: (1) selective rehearsal—that the effect took place at encoding because F items were not rehearsed whereas R items were, and (2) selective search—that the effect took place at retrieval, with search focused more on the R set than on the F set of items. I reasoned that ability to focus search might break down over time but that rehearsal effects should be stable, and so undertook a study comparing immediate and 1-week delayed retention tests. The results, shown in Figure 1, indicated no change in the advantage of R over F items as a function of retention interval, which I took—rather naively, in retrospect—to be evidence in favour of selective rehearsal and in opposition to selective search (published as MacLeod, 1975a).[2]

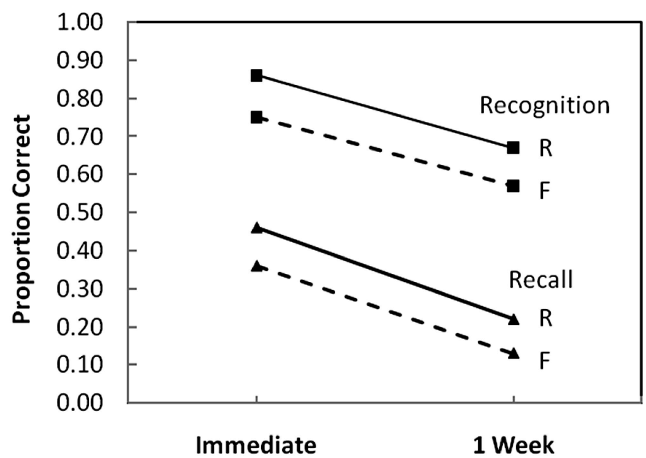

Figure 1. The directed forgetting effect does not change over a long retention interval. The data are replotted from MacLeod (1975a). Proportion correct in free recall and proportion hits in recognition are shown as a function of instructions (R = Remember vs. *F* = Forget) and time of test (immediate vs. 1-week delayed). [*SE* data were not available to construct error bars.]

That was my introduction to research on learning and memory. There were other related studies along the way, including one in which my fellow graduate student, Steve Poltrock, and I (Poltrock & MacLeod, 1977) had the temerity to name a phenomenon—the continuous distractor paradigm—that is still used today (e.g., Unsworth, 2007), although nobody seems to realise that we named it! But the next big step was of course the PhD, and for that I decided to combine the savings method that Tom Nelson had been working with since his dissertation (Nelson, 1971) with my own interest in bilingualism, stemming from growing up in Montréal. The Ebbinghaus (1885/1913) savings method, as updated by Nelson, involved learning a list of paired associates (e.g., 27-*house*) followed by an extended retention interval of weeks to permit adequate forgetting, and then followed by a single relearning trial where the relation of what was relearned to what had originally been learned could be manipulated (e.g., given inability to remember the original pair, how would relearning compare for the original response word *house* vs. a related word, such as *home*, vs. a control item, such as *sky*?).

Again, I hoped to contrast two prevalent views, this time concerning the organisation of bilingual memory: (1) independence—that there were two separate language systems with links between cognate words, and (2) interdependence—that there was one language system with two words attached to each unitary concept. I reasoned (again naively) that if one were to learn a pair such as 56-*horse*, then go away and forget it, and then return to relearn *horse*, *cheval*, or *table*, under the independence view, relearning of *cheval* ought to be no better than relearning the control item *table*, whereas under the interdependence view, relearning of *cheval* ought to be nearly as good as relearning of *horse*. The results (published as MacLeod, 1976) were in accord with the interdependence view, as Figure 2 shows.[3] Part of my dissertation research was conducted during an idyllic year at the University of California, Irvine, doing alcohol and memory research (see Parker, Birnbaum, & Noble, 1976), where I wrote much of the dissertation sitting on the beach in Newport Beach.

When PhD Met Postdoc

Upon completing my dissertation, I returned to the University of Washington to take up a postdoctoral position in the individual differences laboratory of Earl (Buz) Hunt. I continued my interest in memory, but began to think more in terms of how people use their cognitive skills differently in the performance of a wide variety of cognitive tasks. Our first major project investigated how

[2] Had I had the $25 processing fee and had it not been raining on the day that I had to walk across campus to do the necessary paper work, this would have constituted my Master’s thesis. I sometimes fantasize about offering to make a nice donation to the University of Washington if they would be willing to award me my MA now, 40 years later.

[3] Having had to use a 5-week retention interval to secure adequate forgetting, I realized that if I continued using longer and longer retention intervals, in a few years my data would take a very long time to collect!

people compare very simple pictures and sentences, using the sentence-picture verification task developed by Clark and Chase (1972). On each trial in this task, subjects read a short sentence like PLUS IS ABOVE STAR or PLUS IS NOT BELOW STAR and then verify a simple picture of either a plus above a star or a star above a plus. The dominant theory was that of Carpenter and Just (1975)—the constituent comparison theory—which held that we accomplish this task by breaking the sentence and the picture down into their component linguistic elements and then running a series of comparisons until we either exhaust the set (a match, or yes response) or determine a disagreement (a mismatch, or no response). Our initial goal had been to relate performance on this task to psychometrically measured verbal ability, but we discovered that there appeared to be two quite different subsets of subjects in terms of how they performed the task.

The larger group (about 70%) did appear to follow the constituent comparison model, but the smaller group (about 30%) did not. This is strikingly clear in Figure 3. Indeed, the smaller group appeared to be doing something wholly different. They took much longer to process the sentence, but then were much faster to process the picture, and showed a very different latency pattern in responding to the picture. We reasoned that the larger group was using a linguistic strategy and the smaller group was using a pictorial strategy. This pictorial strategy we saw as essentially constructing an expected picture from the sentence and then matching that to the actual picture, which accounted for the very long sentence processing times and the very short picture processing times. And the psychometric data aligned with this two-strategy account: Subjects using the pictorial strategy had very high spatial ability, whereas those using the linguistic strategy had average or low spatial ability.[4]

We reported these results in MacLeod, Hunt, and Mathews (1978) and then went on in Mathews, Hunt, and MacLeod (1980) to show that subjects could readily switch strategies when their nonpreferred strategy was explained to them. That is, subjects chose their strategy based on their spatial ability; strategy was not "hard wired." Intriguingly, Reichle, Carpenter, and Just (2000) demonstrated using functional MRI (fMRI) that the strategy difference was observable at the neural level as well. That strategy differences could be so dramatic and so linked to ability differences convinced me of the importance of individual differences in cognition. Of course, fitting with the theme of this article, it was apparent that these strategy differences were learned and could be quite readily revised and relearned. The flexibility of learning—admittedly, as well as the inflexibility—has always impressed me and was what led me to an interest not only in memory but also in attention, and to the controlled-automatic distinction (see Shiffrin & Schneider, 1977).

The other major project of my postdoctoral years focused on the cognitive costs of antiseizure medication in epileptic patients. This began when Anatole Dekaban, an expert on epilepsy, visited the University of Washington, and met initially with Buz Hunt and subsequently with me as well. He was intrigued, after a career of studying epilepsy from a medical model perspective, by how often patients reported what they called "mental slowing," and wondered what they meant and whether they were right. The plan was to study a group of patients who were coming into the Veteran's Administration Hospital on high levels of antiseizure medication and being titrated down to lower levels, a situation that regularly arose for patients who often had their medication levels increased as a consequence of suffering seizures until they reached quite toxic levels. Gradually, we came to design a large study involving

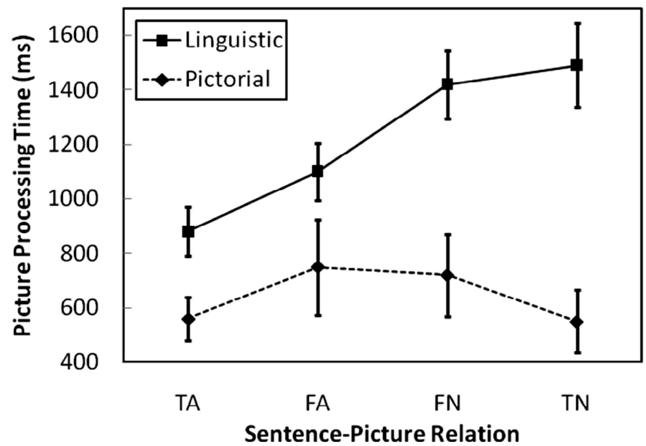

Figure 3. There are clear individual differences in sentence-picture verification strategies. The data are replotted from MacLeod, Hunt, and Mathews (1978). Mean picture processing time is shown as a function of sentence—picture relation (TA = True Affirmative; FA = False Affirmative; FN = False Negative; TN = True Negative) and the predicted sequence in the Carpenter and Just (1975) constituent comparison model. The two strategies—linguistic and pictorial—are very distinct. The error bars are the 95% confidence intervals for the respective means, replotted from the original figure.

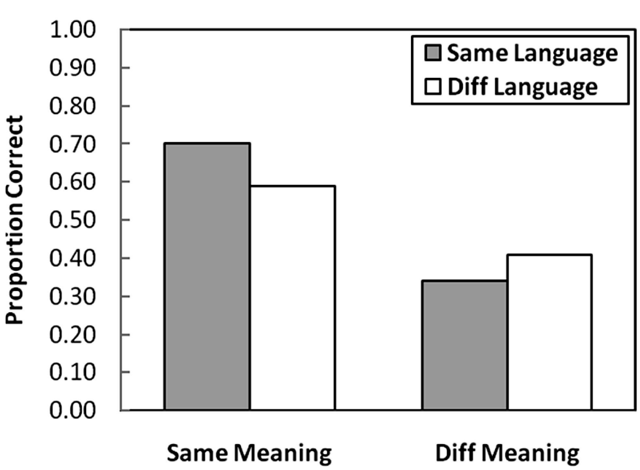

Figure 2. Long-term savings in bilingual subjects is equivalent between languages and within language. The data are from MacLeod (1976). Proportion correct on the relearning trial is shown as a function of language and meaning for items that had been forgotten after a 5-week retention interval. [*SE* data were not available to construct error bars.]

[4] Often when I see this study cited, it is described as the linguistic group being high verbal and the pictorial group being high spatial. This would seem to be a false memory based on what would have been a reasonably intuitive outcome . . . had it actually happened that way. But only spatial ability was related to strategy choice.

multiple cognitive tasks that would be administered to these patients in the first week under high dosages and in the second week under much lower dosages.

We anticipated that working memory tasks might be more affected than long-term memory tasks by drug level, on the view that working memory required more conscious control. Among the tasks that we included were Sternberg's (1966) short-term memory scanning task and Posner's (Posner & Mitchell, 1967) long-term memory access task, two of the then most well-studied tasks in cognitive psychology. The former measures time to search through working memory for a prespecified target; the latter measures time to retrieve a highly overlearned code from long-term memory. Not surprisingly, the medicated epileptic patients were overall slower than the normal controls. But what was most revealing was that the Posner long-term memory task was not reliably affected by dosage level whereas the Sternberg working memory task most definitely was: Scanning time through working memory was much slower under the high dosage. We reported these results as consistent with the drugs affecting working memory but not long-term memory (MacLeod, Dekaban, & Hunt, 1978).[5] This study made me realise how important it is to relate our cognitive tasks to "real world" phenomena, in the service of better understanding the challenges that can arise in learning and remembering. As a result, periodically throughout my career, I have studied cognitive performance in the context of alcohol, clinical disorders, and other factors.

When Postdoc Met Professor

At the end of my 3 years of postdoctoral work, it was time to move on. As I look back, I can readily recapture what a thrill it was to move to the University of Toronto in 1978, where the world's foremost memory group was assembled. Gus Craik, Paul Kolers, Bob Lockhart, Morris Moscovitch, Ben Murdock, Norm Slamecka, and Endel Tulving all inspired me with their research, in addition to providing superb models of scientists. The weekly Wednesday noon meeting of the "Ebbinghaus Empire" featured the most advanced research on memory and other aspects of cognition. And the graduate students and postdocs who went through the program—people like Gary Dell, Kevin Dunbar, Eric Eich, Bill Hockley, Steve Lewandowsky, Janet Metcalfe, and Dan Schacter—now represent a veritable who's who of leading researchers in the field. I had the benefit, too, of working in a smaller college environment at the Scarborough campus, while having the resources of the larger university nearby, an ideal setting for building a career.

At Toronto, I continued the lines of research begun at Washington, including work on savings (MacLeod, 1988), on directed forgetting (e.g., MacLeod 1989a), and on individual differences (MacLeod, Jackson, & Palmer, 1986). But two new interests arose as well, inspired by my colleagues and graduate students. First was work on implicit memory (e.g., MacLeod, 1989b), which tied back nicely to the savings work, in that relearning and savings really represent the first implicit memory testing procedure. Because of our shared interest in implicit memory, a longstanding and fruitful collaboration grew with Michael Masson at the University of Victoria (e.g., MacLeod & Masson, 2000; Masson & MacLeod, 1992). The idea that learning could be measured without the need for awareness inspired a great deal of research at Toronto in that period, and I was excited to be a part of a new way of thinking about memory.

Not long after I moved to Toronto, though, a quite different area of research captured my interest. This actually began one winter evening in 1981 when Kevin Dunbar (then doing his PhD in my lab) and I were talking about various cognitive tasks that might be worth considering to examine cognitive flexibility, an individual differences concept that I sought to revive. I had decided to write a grant to explore this concept, the core idea being to have subjects train on tasks until their performance was quite asymptotic and then put them in a situation where the task switched from trial to trial. My hypothesis was that individuals who had more difficulty with task switching would be less cognitively flexible, and we could then explore this difference. But in talking about various tasks, we accidentally combined Paul Kolers' reading upside down task (Kolers, 1973) with the venerable Stroop task (Stroop, 1935). It quickly occurred to us that this would be an interesting study in its own right, and the cognitive flexibility direction—and the soon to be very popular study of task switching—was immediately abandoned. It has always intrigued me how research projects and even entire research programs can change direction dramatically because of what is initially a seemingly casual or unrelated thought.

With that began a new research program on attention, in particular in the context of the Stroop effect and the factors that govern interference in attention. Our "start vector" was the idea of an upside-down Stroop task. In the basic Stroop task, colour words are presented in the wrong colour (e.g., the word red in green ink). Stroop (1935) showed that this mismatch or incongruency did not affect time to read the colour word, but that it drastically slowed time to name the colour relative to appropriate control conditions. Essentially, subjects could not turn off word reading, a skill that is much more practiced than colour naming—even automatic (see Cattell, 1886). What if reading were to be made much more difficult, even more difficult than colour naming?

We thought that presenting the colour words in novel orientations would slow reading, and should reverse the effect such that incompatible colours would now interfere with reading words, but incompatible words would not interfere with colour naming because the words would be processed so slowly. Our idea was that, once we had obtained this pattern, we would then teach subjects to read words in this new orientation, following Kolers' (1973) procedure, and expect to see the interference pattern return to normal for these transformed words as they became easier to read. Indeed, we thought that we might be able to use the point at which the reversal occurred as an empirical signature of the emergence of automaticity. It was an exciting idea, so we set out to explore it.

What happened surprised us: We obtained full blown interference for the transformed words without any training at all. As Figure 4 (taken from Dunbar & MacLeod, 1984, Experiment 3) shows, normally oriented words and transformed words produced equivalent interference in colour naming, despite time to read the transformed words being very slow, considerably slower even than colour naming. At the time, the dominant view of Stroop

[5] I was actually quite convinced that the effect was on controlled rather than on automatic processing, but in 1977 this was a very new idea so the more traditional memory store view held sway.

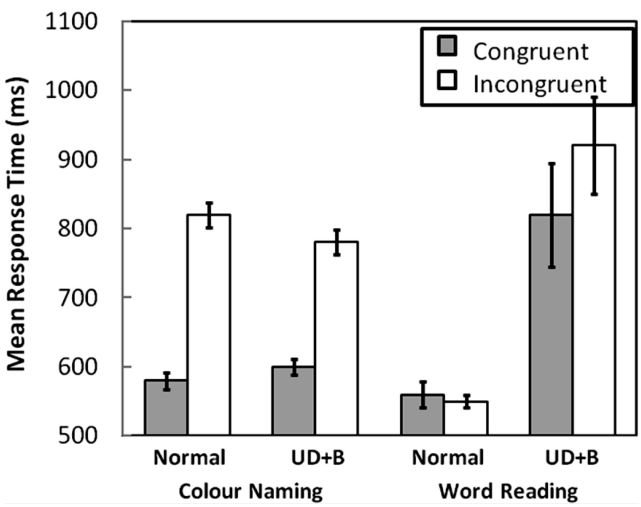

Figure 4. Stroop interference is equivalent for normally oriented and for transformed (in this case, upside down and backward) words. The data are replotted from Dunbar and MacLeod (1984, Experiment 3). Mean oral response time (in milliseconds) is shown as a function of stimulus condition (incongruent vs. congruent) and response dimension (colour naming vs. word reading). The error bars are the *SE*s for the respective means, replotted from the original figure.

interference was that it was caused by a faster process (reading) interfering with a slower process (colour naming)—the speed of processing account (see Dyer, 1973). So how could a slower process influence a faster one? We argued in favour of a parallel processing account where processing on the two dimensions (word and colour) was ongoing simultaneously, with crosstalk producing interference. But most critically, we argued that the sequential speed of processing account had to be wrong.

Initially, we had hoped to investigate how learning affected Stroop interference by manipulating amount of training in reading transformed text. Of course, this no longer made sense given full blown interference without training. Instead, we came up with a different way to investigate the development of automaticity through learning. We trained small groups of subjects to name four distinct random polygons, each with a unique colour name. I will describe an experiment in which training lasted for 20 days with several hundred trials per day. During training, the shapes were always presented in white to be named with their unique colour names. But on three critical days—Day 1, Day 5, and Day 20—we examined interference after the training was completed by presenting the shapes in colour (e.g., the shape called red presented in green). In one block, we asked subjects to name the newly learned shape colour (i.e., say “red”), in the other block, we asked them to name the familiar colour (i.e., say “green”). We were interested in how the interference pattern would change with practice.

Figure 5 displays the data from MacLeod and Dunbar (1988, Experiment 3). The training data were lovely, as shown in Panel 1—shape naming speed steadily improved.[6] Panel B shows the interference data. On Day 1, shape naming was very slow and familiar colours interfered with unfamiliar shape names, but not vice versa. By Day 5, there was bidirectional interference, certainly inconsistent with any speed of processing account, but understandable if the two dimensions were by then of roughly equivalent automaticity. And then on Day 20, the initial pattern had reversed, with interference now seen when naming colours but not when naming shapes. Note that there were no words at all in this study, demonstrating that reading is not a necessary element in Stroop interference. We had succeeded in making shape naming more automatic than colour naming via extended learning.

I went on to investigate other aspects of Stroop interference (e.g., MacLeod & Hodder, 1998; MacLeod & Bors, 2002), including further studies of training (MacLeod, 1998b). Somehow, I even found myself reviewing the vast Stroop literature (MacLeod, 1991b) and writing a brief sketch of John Ridley Stroop (MacLeod, 1991a), in keeping with my longstanding interest in the history of experimental psychology (see MacLeod, 1992). I also became interested in attention more generally (e.g., negative priming: MacLeod, Chiappe, & Fox, 2002; visual search: Wilson, MacLeod, & Muroi, 2008; Wilson, Muroi, & MacLeod, in press). For a long time, this attention work coexisted happily my continuing memory research without them ever actually meeting.

When Attention Met Memory

Then, in the late 1980s, I began to consider how these two cognitive domains might interact with each other. I suppose that my career-long interest in directed forgetting (see MacLeod, 1975a; MacLeod, 1998a) had always biased me to think about the degree of attention paid to material at the time of encoding. And I had occasionally explored the attention-memory interface from early on (e.g., MacLeod, 1975b; Hauer & MacLeod, 2006). But in thinking about the implicit-explicit distinction in memory, I had noticed that most often the research literature on attention and the research literature on memory were essentially disconnected, despite everyone realising that attention is an important determinant of memory. Out of this came several efforts to bring attention and memory together. It will help to set the stage for the topic of the rest of this article if I describe a couple of these studies.

The first formed the Master’s thesis of Katrin Szymanski (Szymanski & MacLeod, 1996; see also MacLeod, 1996). In the study phase, we combined a Stroop-like task with a standard list learning memory procedure. There were two study blocks. In one block, subjects read words aloud ignoring their colour; in another block, they named the colours of words aloud ignoring the words themselves. They then performed one of two memory tests. On an explicit yes/no recognition test, subjects decided whether a word had or had not appeared in one of the two study blocks. Here, memory was much better for the words that had been read aloud than for those that had been colour-named aloud. On an implicit lexical-decision task, subjects decided as quickly as they could whether letter strings were words or nonwords. Some of the words had just been studied and some were new. Here, the priming advantage of having appeared in either of the study blocks was equivalent. Thus, attention to the words during study mattered on the explicit test but not on the implicit test, a result soon confirmed in other laboratories (e.g., Stone, Ladd, Vaidya, & Gabrieli, 1998).

For her Master’s thesis, Penny Macdonald pursued this finding (MacDonald & MacLeod, 1998). In Experiments 1 and 2, 80

[6] To demonstrate just how amazingly regular training day can be, there is even a little slowing each Monday after having a break over the weekend (Days 6, 11, and 16)!

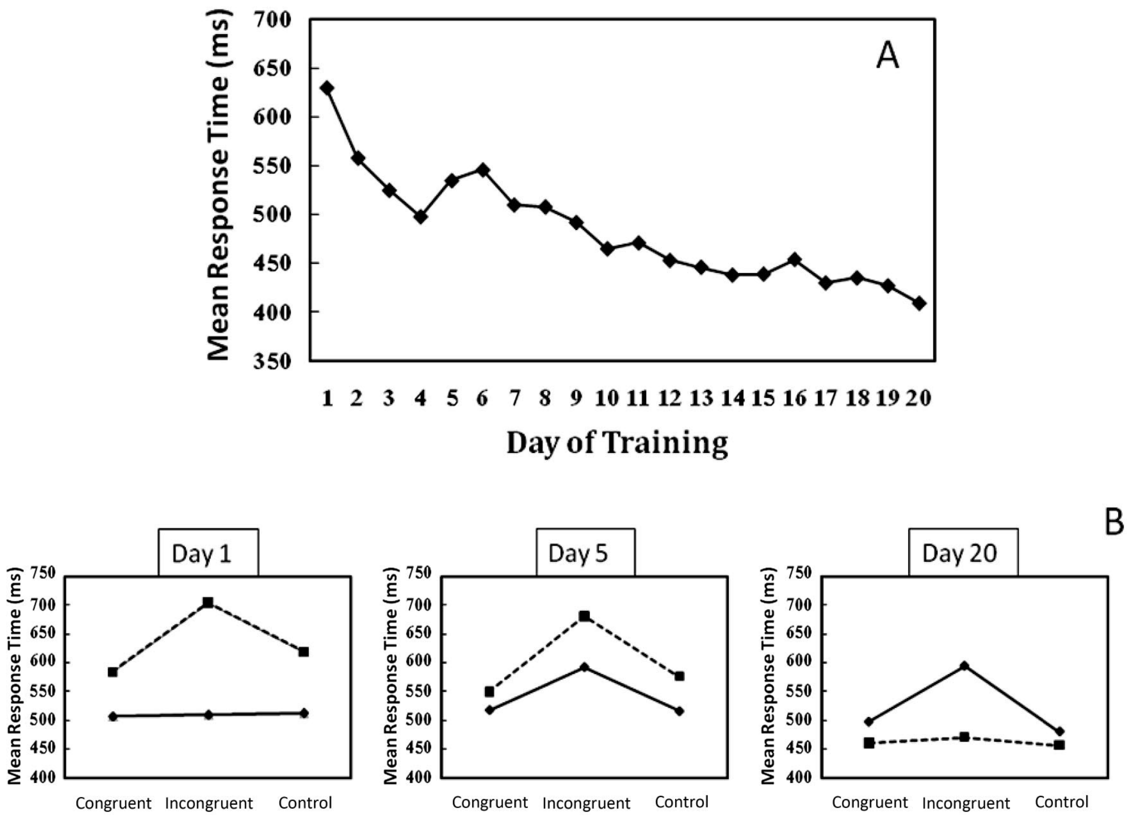

Figure 5. Stroop interference is governed by training leading to automaticity. The data are replotted from MacLeod and Dunbar (1988, Experiment 3). Panel A displays mean oral shape naming time (in milliseconds) as a function of day of training. Panel B displays oral shape naming time (dotted lines) and oral colour naming time (solid lines) as a function of condition (congruent, incongruent, and control). The three subpanels of Panel B present performance after the first, fifth, and twentieth day of practice. [*SE* data were not available to construct error bars.]

words appeared for study with 40 in blue—to be spoken aloud—and 40 in white—to be read silently. Aloud versus silent was the attentional manipulation. There were again two tests, this time speeded reading (also called naming, the implicit test where the task is simply to read a word aloud into a microphone as rapidly as possible) and yes/no recognition (the explicit test). Words read aloud were better recognised than those read silently on the explicit test but there was no difference on the implicit test. These results converged nicely with the Szymanski and MacLeod (1996) results. The novel finding was in Experiment 3 where there were two words on each trial, one in blue and one in white. Now, to attend to one word was essentially to ignore the other, and now even the implicit test showed a cost of ignoring. Even implicit memory requires some minimal amount of attention.

Some years later, Mike Dodd and I (Dodd & MacLeod, 2004) used this approach again to investigate false memory in the well known Deese-Roediger-McDermott paradigm (Roediger & McDermott, 1995). In this procedure, subjects study lists of words all of which are related to a critical unpresented word (e.g., study *tired, night, bed, pillow* . . .) and they subsequently are very prone to incorrectly remembering the critical unpresented word, in this case *sleep*. We wondered whether attention at study was crucial for this effect. So we had subjects study the lists with the words printed in colour and the subjects either read the words aloud (ignoring their colours) or identified the colours by pressing corresponding buttons (ignoring the words). They then did a yes/no recognition test. What we found was a huge cost for the actually studied words when they were colour-named but no cost at all for the critical unpresented word: False memories did not seem to require attention to be created quite robustly (and our subsequent research suggested that this really was an effect occurring at the time of study, not at the time of test; see Dodd, Sheard, & MacLeod, 2006).

By 2003, the Szymanski and MacLeod (1996) and MacDonald and MacLeod (1998) findings had had long enough to percolate, and it occurred to me that the very fact that reading a word aloud seemed to make it more memorable was in itself potentially important as a way to improve learning. That was the year I moved from the University of Toronto to the University of Waterloo so my new students and I began an exploration of this apparent memory benefit. We have now done some 40-plus experiments, and this work has convinced us that saying things aloud is in fact a quite powerful mnemonic technique. In the second half of this article, I will turn my attention to this new line of research.

The Production Effect in Memory

There are relatively few powerful learning tools. Any introductory text or first cognitive text—or indeed any book on memory improvement—will list imagery (Paivio, 1971) and elaboration (Craik & Lockhart, 1972). Textbooks will likely include the generation effect (Slamecka & Graf, 1978; for a review see Bertsch, Pesta, Wiscott, & McDaniel, 2007)—that retrieving something from memory (e.g., generating the word that fits a definition) leads to better retention than does simply experiencing it (e.g., simply reading the word). A few other less universal encoding techniques might be noted in any given source, but the list would remain very short. So any encoding procedure that enhances remembering, and therefore might be worthy of joining this short list, would be most welcome. We believe that we have found one.

The basic phenomenon. We began by replicating the MacDonald and MacLeod (1998) study, which had demonstrated that reading a word aloud improved yes/no recognition quite substantially over simply reading it silently. The basic procedure was to show subjects a list of words for study, half in blue (to be read aloud) and half in white (to be read silently). The list was quite long (80 items) in recognition experiments; in recall experiments, which I will describe later, the list was shortened (to about 36–48 words). Having easily replicated the MacDonald and MacLeod results,[7] we decided that the phenomenon deserved a name and so, by analogy to the generation effect, it became *the production effect*.

Around this same time, I started combing the literature for relevant articles other than MacDonald and MacLeod (1998) and found just a few, sprinkled over the years and rarely cited, as showing an advantage of aloud over silent study (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972). Precedence goes to Hopkins and Edwards (1972), who showed a clear production effect both in yes/no recognition and in two-alternative forced choice recognition but only when they used a within-subject design—the effect was absent when they used a between-subjects design. We also replicated this design specificity using both yes/no recognition—shown in Figure 6—and two-alternative forced-choice recognition (MacLeod et al., 2010, Experiment 3). This was the first clue to the cause of the production effect: The effect requires discrimination of aloud from silent items at the time of study.

In MacLeod et al. (2010), we went on to show that the production effect did not occur when the overt response was repeated. When subjects responded to all of the blue words by pressing the same key on the keyboard or by saying “yes” aloud (Experiments 4A and 4B), the blue words were not remembered better than the white ones, to which no overt response had been made. This was the second clue to the cause of the effect: Each produced response had to be unique. We did observe, however, that vocalizing was not crucial, because mouthing the words produced a robust production effect (Experiment 5), nor was it important that the stimuli be meaningful, because nonwords also displayed a large effect (Experiment 6).

One concern we had was that production might only work for relatively weak encodings, but Experiment 7 in MacLeod et al. (2010) dispelled that notion. We had subjects generate all of the words at study, using definitional cues such as “the tiny infant commonly put in a cradle - b?” for the word “baby.” For blue cues, they were to generate aloud; for white clues, they were to generate silently. On the left side, Figure 7 shows the reliable production effect obtained, with overall performance much improved by generation, so clearly even strong encodings can benefit from production. This was reinforced by Experiment 8, using deep semantic encoding—an initial living/nonliving judgment, followed by saying the word aloud if in blue, silently if in white. On the right side, Figure 7 shows that overall performance was very good compared with the earlier experiments, yet there was still a reliable production effect. These data also lay to rest the concern that the production effect might actually be a cost, not a benefit, with subjects ignoring the silent white words and attending preferentially to the aloud blue words. In the context of the generation effect, this has been referred to as the “lazy reading hypothesis” by Begg and Snider (1987), but the initial deep processing insures that the white words have been well encoded, not ignored.

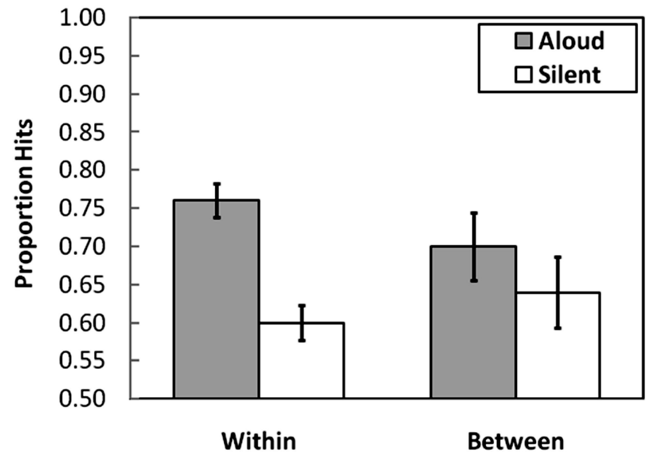

Figure 6. Reading words aloud produces a benefit over reading words silently—the production effect—under a within-subject but not under a between-subjects design. The data are from MacLeod et al. (2010). The left side shows the combined recognition hit rate data from Experiments 1a and 1b using a within-subject design (false alarm rate = .22); the right side shows the recognition hit rate data from Experiment 2 using a between-subjects design (false alarm rate = .26 for the aloud condition and .19 for the silent condition). The error bars are the *SE*s for the respective means.

The distinctiveness explanation. From this initial set of experiments, we came to an explanation of the production effect in terms of distinctiveness (see Conway & Gathercole, 1987, for the first mention of this explanation in this context). The basic idea of distinctiveness as an explanatory mechanism is that information which is made to stand out from other information at the time of encoding will show enhanced memory. It is an old idea (see, e.g., Murdock, 1962) that has been the subject of renewed interest of

[7] My view is that replication is one of the most powerful tools that we have in experimental science. I have often joked to my students that I would like my tombstone to read “Nothing he ever did failed to replicate.” [I should note here, too, that we have of course counterbalanced the colours that signal the subject to respond aloud versus silent; I have stayed with the blue = aloud and white = silent description here for consistency and clarity.]

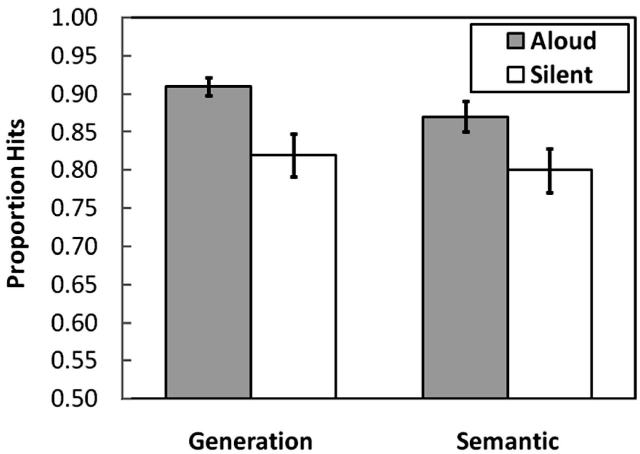

Figure 7. Even strong initial encodings show the production effect. The left side shows the recognition hit rate data from Experiment 7 where items were generated from definitional cues (false alarm rate = .08); the right side shows the recognition hit rate data from Experiment 8 where items were semantically encoded (false alarm rate = .14). The data are from MacLeod et al. (2010). The error bars are the *SE*s for the respective means.

late, as is evident in a recent book on the subject (Hunt & Worthen, 2006), in which Hunt (2006) provides a thorough overview of the concept (see also Hunt & McDaniel, 1993). Our argument is simply that, during study, words read aloud are discriminated from those read silently such that, at the time of test, those read aloud have available as part of their encoding that they were in fact read aloud. This unique information can be used as diagnostic that an item was in fact studied.

Dodson and Schacter (2001, p. 155) suggest a distinctiveness heuristic whereby people are thought to "demand access to [the distinctive] information as a basis for judging items as previously studied; the absence of memory for this distinctive information indicates that the test item is new." We think of this in the proceduralist framework espoused by Kolers (1973; Kolers & Roediger, 1984; see also Kirsner & Dunn, 1985) as attempting to replay the encoding at the time of retrieval. If that replay reveals encoding to have been aloud, this confirms the item as having been studied, supporting an "old" response. If, however, there is no evidence of its having been read aloud, the status of the item is ambiguous: Was it studied silently or was it not studied at all (i.e., is it a lure)? The answer depends on recollection.

Generalising and applying the production effect. Along the way, we have learned a good deal more about the production effect. I will just broadly sketch some of the recently conducted studies here. We know that there is a reliable production effect if the study list is blocked, with all of the aloud items preceding all of the silent items or vice versa. This suggests to us that the distinctiveness is "global," not "local": An aloud item need not be embedded among silent items to produce the benefit. We know that a production effect is observed even if subjects just imagine saying the blue words aloud and the white ones silently, suggesting that distinctiveness does not require an actual motor component. We know that there is a reliable production effect if the subject writes or types—or even spells out loud—the blue words but not the white ones, so the modality of production is not critical, nor must the response produced at study be the entire item.

For any encoding technique to be broadly valuable, it must work for various types of materials and for different types of tests. On the materials side, we know that there is a robust production effect for simple line-drawing pictures, just as there is for nonwords. Word pairs as well as sentences show a consistently reliable production effect. And because all of the prior studies had used recognition tests of one form or another, we now have investigated recall as well. With words as stimuli—in this case, one word from each of 32 distinct categories—both free and cued recall (using the category names) show nice, strong production effects, as shown in Figure 8. This production benefit extends to recall of the names of simple pictures as well.

All of these findings reassure us that the production effect is powerful and readily obtained under a wide variety of circumstances. In fact, this led us to begin considering production as a potentially useful study technique in "the real world." Two additional conditions needed to be met for production to be valuable in studying. First, the production effect needed to survive a considerably longer retention interval. Jason Ozubko, Kathleen Hourihan and I have shown, using word lists at study, that the effect is just as strong one week after study as it is immediately after study. So survival over a long retention interval has been demonstrated (see Figure 9). What of a production benefit for more meaningful materials? We have also carried out an experiment using the text materials from Chan, McDermott, and Roediger (2006), short passages of a couple of pages about interesting topics (e.g., Hong Kong, toucan birds). This time, we highlighted some of the paragraphs, and these were to be read aloud. Once again, the production effect was readily apparent, this time on a fill-in-the-blanks test, further generalising the type of test that shows the effect. The upshot is that the production effect does appear to have real

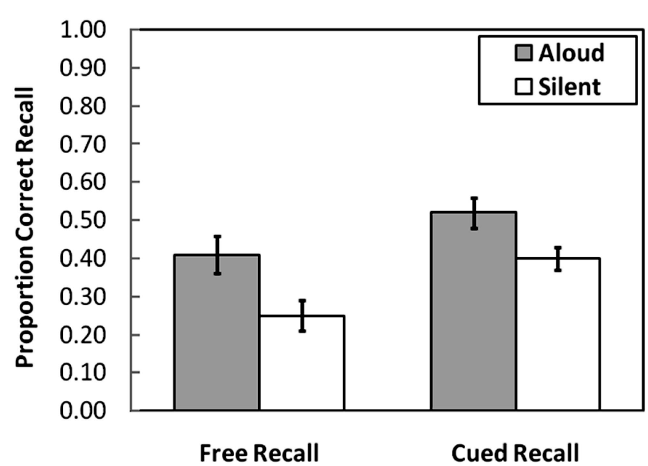

Figure 8. The production effect is evident in free recall and in cued recall. Panel A shows free recall proportion correct for a list of words selected one from each category; Panel B shows cued recall proportion correct for the same list with the category names as cues. The experiments from which these data are drawn are unpublished. The error bars are the standard *SE*s for the respective means.

potential as a study technique.[8] To this can be added the finding from another unpublished study where Jason Ozubko and I found that repeated production does not alter (in particular, repetition does not undermine) the benefit: The production effect is stable over repetition.

Of course, if they are to be believed, students often study together. Does the production advantage accrue when someone else does the producing? Recently (MacLeod, 2010), I have examined this question using subject dyads and a different way of indicating what to produce. Pairs of subjects sat together facing one screen on which words could appear in one of four locations: the bottom signalled that both subjects should read the word silently, the top signalled that both should read the word aloud, the left signalled that only the subject on the left should read the word aloud, and the right signalled that only the subject on the right should read the word aloud. The results (Figure 10) showed precisely the same pattern for both recall (A) and recognition (B)—a small but reliable production effect for the words that the other person produced, a reliably larger production effect for those that both subjects produced, and a reliably larger still production effect for the words produced by oneself. So the answer is that hearing another person produce is better than no production at all, but it is best to do the production alone oneself: There appears to be a cost to doing it together!

Most recently, Olivia Lin and I have carried out a study (as yet unpublished) comparing the production effect in older and younger adults. As we expected, older adults—like the younger adults that we typically have studied—do show a reliable benefit of production. The effect is, however, significantly smaller for older subjects than for younger subjects, both in recall and in recognition. [In recognition, this is true even despite the fact that we did not observe a reliable overall reduction in performance for the older subjects relative to the younger subjects.] Interestingly, the form of this interaction is consistent in recall and recognition: Performance for the older and younger subjects is almost identical for the silent items, but the gain for the aloud items is greater for the younger subjects. This suggests that the additional recollection that supports the production benefit is compromised in older adults.

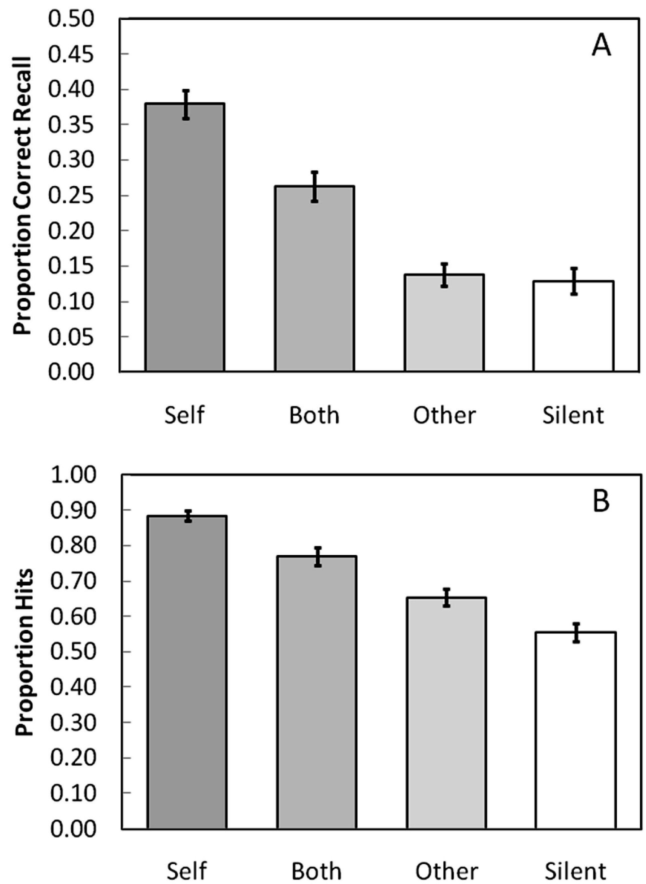

Figure 10. Production helps when you listen to someone else do it, but it helps most when you do it yourself. Panel A shows free recall proportion correct; Panel B shows recognition hit rate (false alarm rate = .12). The data are from MacLeod (2010, Experiment 2). The error bars are the *SE*s for the respective means.

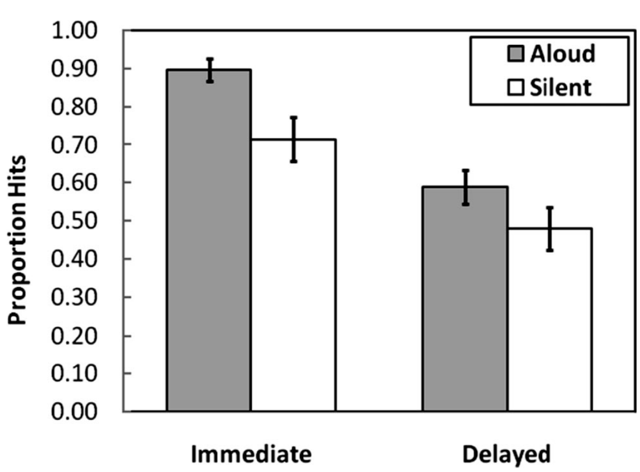

Figure 9. The production effect endures over a longer retention interval. The proportion of hits for recognition of words over a 1-week retention interval shows the stability of the production effect (false alarm rates = .23 immediate, .36 delayed). These data are drawn from an unpublished experiment. The error bars are the *SE*s for the respective means.

One other result that we have observed repeatedly is worthy of note before turning to some more direct tests of the distinctiveness explanation. In MacDonald and MacLeod (1998), in Hourihan and MacLeod (2008), and in most of the experiments in MacLeod et al. (2010), we included an implicit memory test as well. This test was speeded reading (also called naming), wherein subjects simply have to read test words as quickly as possible into a microphone, some of which are studied words and some of which are unstudied words. Priming—faster reading of studied words—would be evidence of an effect of production on implicit memory (see MacLeod & Masson, 2000). Over a great many experiments, we have never found differential priming in speeded word reading for words studied aloud versus silently. We think that this is as it should be:

[8] Occasionally I have been asked whether production really is useful as a study technique, given that it only works in within-subject designs. In fact, I see this as a virtue, not a drawback. When studying, we need to identify the important material and emphasize it, and the production effect seems like a really effective tool for this purpose.

The benefit of production is based on the recollection of a distinctive encoding, which should only be relevant on an explicit test where the subject is trying to remember a prior study episode. This was our third clue: The production benefit is restricted to explicit remembering.

Testing the distinctiveness explanation. How can we ascertain whether distinctiveness is in fact the "active ingredient" in the production effect? Some evidence consistent with this account has already been sketched out: (1) that the effect occurs only within subject and not between subjects; (2) that the response to each item must be unique; (3) that the benefit is not present on a test of implicit memory; (4) that the produced response does not have to be audible or even involve motor activity; and even (5) that the effect varies as a function of personal involvement. We wanted, however, to put the distinctiveness account to more direct tests. I will describe three.

First, Kathleen Hourihan and I (Hourihan & MacLeod, 2008) combined the production effect with the item method directed forgetting procedure. The standard finding in directed forgetting studies is that recall (and recognition, for the item method) is poorer for items designated as to-be-forgotten than for items designated as to-be-remembered (for a review, see MacLeod, 1998a). We reasoned that were an item to be made distinctive prior to a forget instruction, that item would be difficult if not impossible to intentionally forget. Therefore, the silent items should show a directed forgetting effect, but the aloud items should not. That is precisely what we found. This pattern fit nicely with a study in which Karen Daniels and I (MacLeod & Daniels, 2000) had combined the generation effect with the directed forgetting procedure, and found a directed forgetting effect for the items that had been read but no directed forgetting effect for those that had been generated. Distinctiveness apparently inoculates material against intentional forgetting (see Golding, Long, & MacLeod, 1993, for a related result).

Second, if production hinges on distinctiveness, it should do so by calling on something like a distinctiveness heuristic (see Dodson & Schacter, 2001). Of interest, subjects in our experiments have told us that they actively try to remember—to recollect—saying a word aloud to help in diagnosing whether it was studied. So subjective reports align with such an explanation. To pursue this, Nigel Gopie and I incorporated the "Remember/Know" decision (see, e.g., Gardiner, 1988; Rajaram, 1993; Tulving, 1985) into our testing procedure. For each test item, subjects were to decide whether they were basing their positive recognition judgments on (1) actually recollected information from the study episode (the "Recollect" response) or (2) just a sense that the item had been studied, unaccompanied by any recollective experience (the "Know" response). Collapsing the "Recollect" and "Know" responses into a single hit rate for each of the aloud and silent conditions, we observed the usual strong production effect. We reasoned that if recollection of the distinctive information that an item had been said aloud during study was what made retention superior following production, then the production effect advantage should be largely—or even entirely—restricted to the "Recollect" judgments. As Figure 11 shows, that is precisely what we found: The production benefit was apparent only for "Recollect" responses and was absent for "Know" responses, fitting well with the distinctiveness explanation.

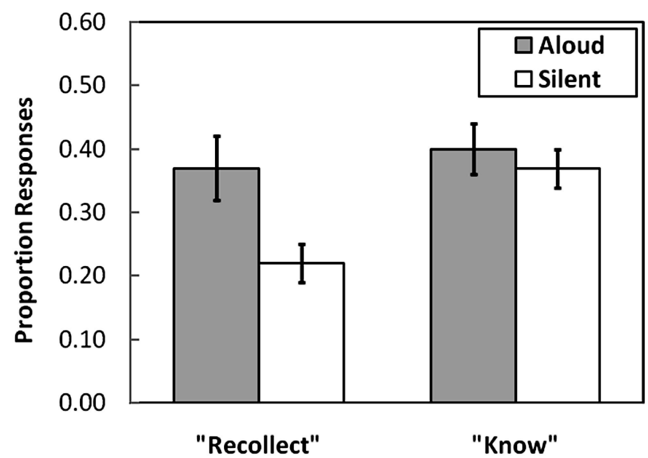

Figure 11. The benefit of production largely relies on recollection. The hit rates are shown separately for the two type of responses—"Recollect" and "Know"—for both Aloud and Silent items. The experiment from which these data are drawn is unpublished. The error bars are the *SE*s for the respective means.

The third test was the most direct. What is distinctive in all of the experiments that I have described is that some items were read aloud and some silently. What would happen if we were to tamper with this dimension of distinctiveness? Jason Ozubko and I (Ozubko & MacLeod, in press) thought of a way to accomplish this. The procedure is somewhat different from the previously described studies. Here, subjects studied two lists. One was the typical half-aloud, half-silent list. The other list was either all aloud or all silent. The test was list discrimination: On the test, all of the items had been studied, and the question was which list they were studied on. We reasoned that the distinctive "aloud" information would still be useful in the face of another list that was all silent, but that its utility would diminish and perhaps even vanish in the face of another list that was all aloud. That is precisely what we found in two experiments (see Figure 12 for the data from one of them)—no production effect when the other list was all aloud.[9] We see this as strong evidence in favour of the distinctiveness explanation, especially when coupled with the evidence described previously.

The relation between production and generation. The production effect and the generation effect seem more than superficially similar.[10] In each case, something must be retrieved from memory—either the item itself or something related to the item. So, spurred by the finding in Experiment 7 of MacLeod et al. (2010) that a production effect can occur even for generated items, I have begun to explore their relation, too. In one unpublished experiment, subjects were presented with a list of words, one from

[9] We did the experiment once with the critical mixed list before the pure list, and once with the critical mixed list after the pure list. The same pattern of results was obtained in both cases.

[10] I would actually include in this same class the enactment effect—where memory is better for an action phrase like "break the match" if the action is actually carried out than if it is simply read (for more on enactment, see Engelkamp, 1998; Engelkamp & Jahn, 2003). It will be interesting to examine the connections among these phenomena.

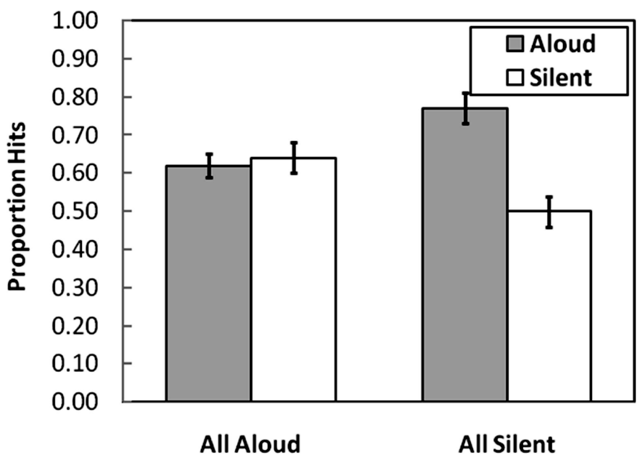

Figure 12. Undermining the distinctiveness of having said some words aloud eliminates the production effect. The data are recognition hit rates as a function of the other studied list (false alarm rates = .35 for the all-aloud other list, .33 for the all-silent other list). The production effect in the critical mixed list is intact when the other list is studied all silently but disappears when the other list is studied all aloud. The data are from Ozubko and MacLeod (2010, Experiment 2). The error bars are the *SE*s for the respective means.

each category and half of them in blue. One group of subjects read the word aloud for the blue ones and silently for the white ones; this should lead to a production effect. The other group of subjects generated the category name aloud for the blue ones and silently for the white ones; this should lead to a generation effect. Both effects were very clearly present in the data for both free recall and category-cued recall, as shown in Figure 13A. The data even suggest— contrary to intuition—that the production benefit may be larger for generated items than for read items, although this will remain speculative until we can gather further evidence.

A more direct attack on the production-generation relation comes from another unpublished experiment crossing production with generation in a within-subject design. Here, subjects studied a list just as in MacLeod et al. (2010). Half of the items were presented as generation clues, with half of these in blue to be generated aloud and half in white to be generated silently. The other half of the studied items were presented as words, with half of these in blue to be read aloud and half in white to be read silently. As Figure 13B shows, both the generation effect and the production effect were robust whether the test was free recall or recognition. Of particular interest, there was no interaction between the two effects, suggesting that they make independent contributions to memorability. Further studies are under way to explore this relation.

Production: The "big picture." In this second section, I have detailed one domain of memory research that has captured the attention of my students and me over the past several years—the value of actually producing something, rather than experiencing it more passively. We have long known that active experience is important (see, e.g., the classic studies of active vs. passive experience in kittens, Held & Hein, 1963). The production effect fits with that tradition—and indeed with the modern tradition of embodied cognition (see, e.g., Anderson, 2003; Robbins & Aydede, 2009). Like generation and enactment, production focuses attention on the information that is most important to learn, and provides a useful means for helping to remember that information later. Production makes the produced information distinctive, and may even prove to be a useful study technique. To enhance memory of the information judged to be most important, one effective strategy would appear to be to produce that information. In the coming years, one goal for research in my laboratory will be to develop a more complete picture of this aspect of learning, both empirically and theoretically.

Conclusion

Since childhood, I have been fascinated by learning and memory. This fascination began with the other passion of my life,

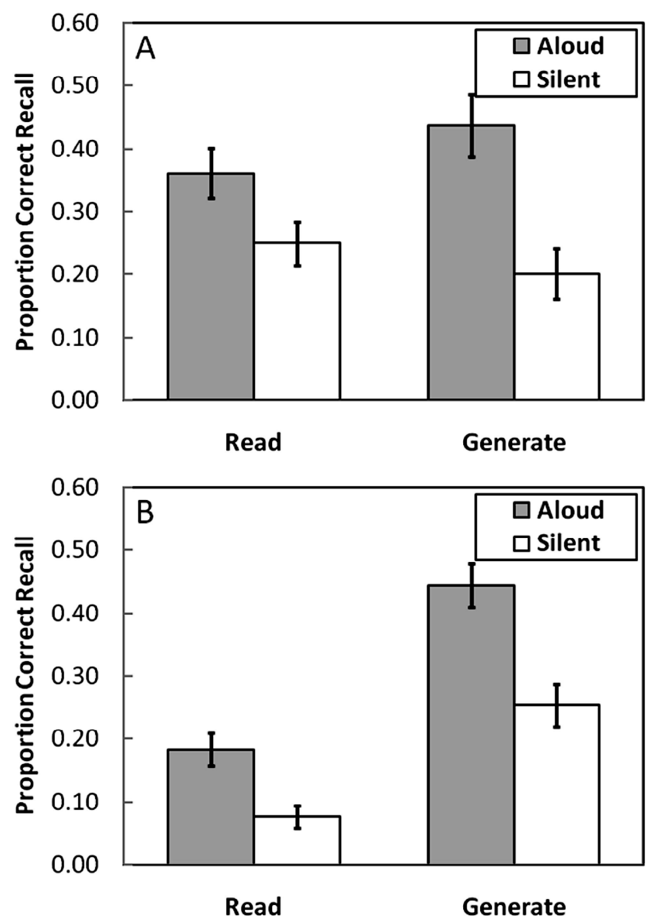

Figure 13. The production and generation effects appear to be independent. Panel A presents the free recall data from a between-subjects experiment in which one group responded by saying the blue words aloud and the white ones silently (production) and another group responded by generating the blue words aloud and the white ones silently (generation). Panel B presents the free recall data from a within-subject experiment in which trials involved either words to be produced or phrases from which words were to be generated, with subjects responding aloud half of the time and silently the other half of the time. Both sets of data are unpublished. The error bars are the *SE*s for the respective means.

popular music, and with the realisation, as I became an expert in that domain, of just how remarkable these cognitive skills are—to learn an artist's entire repertoire, to be able to recognise a song in its first couple of notes, and to have quick access to millions of facts about music. That my career has been spent studying how we learn and remember has, consequently, been a privilege. To study what may well be our most prevalent activity and to try to understand its intricacies is exciting, as is knowing that we have only just begun to uncover these most fundamental skills. With those thoughts in mind, I leave the last word to Theodore Geisel (Dr. Seuss), with whom I completely agree that "The more that you learn, the more places you'll go" (Seuss, 1978). It is a wonderful ongoing journey.

Résumé

La première phrase de l'introduction de la monographie classique de Hebb (1949), *The Organisation of Behaviour*, est « *It might be argued that the task of the psychologist, the task of understanding behaviour and reducing the vagaries of human thought to a mechanical process of cause and effect, is a more difficult one than that of any other scientist* » (p. xi). Cette affirmation est particulièrement vraie dans un contexte d'apprentissage et de mémoire chez l'humain, étant donné notre habileté remarquable à acquérir et retenir une quantité prodigieuse d'information. Cet article se divise en deux parties. La première partie traite de ma fascination de longue date d'abord pour la mémoire, ensuite pour l'attention, et finalement pour leur interaction, tout en donnant des exemples de trouvailles intéressantes en cours de route. La deuxième partie décrit les travaux récents réalisés dans mon laboratoire portant sur une technique d'encodage simple, mais puissante : prononcer les choses à voix haute améliore leur mémorisation. Ce bénéfice, que nous appelons l'effet de production, est probablement dû à une amplification du caractère distinctif des éléments prononcés à voix haute, et pourrait constituer une technique d'étude efficace. Comprendre comment nous apprenons et nous rappelons est ultimement une étape cruciale afin de comprendre qui nous sommes.

Mots-clés : mémoire, attention, apprentissage, production, trait distinctif

References

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149,* 91–130.

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

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

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Washington, DC: Winston.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82,* 45–73.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472–517.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26,* 341–361.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671–684.

Dodd, M. D., & MacLeod, C. M. (2004). False recognition without intentional learning. *Psychonomic Bulletin & Review, 11,* 137–142.

Dodd, M. D., Sheard, E. D., & MacLeod, C. M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory, 14,* 115–126.

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.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 622–639.

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition, 1,* 106–120.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology.* New York: Teachers College, Columbia University.

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

Engelkamp, J., & Jahn, P. (2003). Lexical, conceptual and motor information in memory for action phrases: A multi-system account. *Acta Psychologica, 113,* 147–165.

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

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

Golding, J. M., Long, D. L., & MacLeod, C. M. (1994). You can't always forget what you want: Directed forgetting of related words. *Journal of Memory and Language, 33,* 493–510.

Hauer, B. J. A., & MacLeod, C. M. (2006). Endogenous versus exogenous cuing effects on memory. *Acta Psychologica, 122,* 305–320.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology, 56,* 872–876.

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. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York: Oxford University Press.

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

Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory.* New York: Oxford University Press.

Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention? In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 547–565). Hillsdale, NJ: Erlbaum.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition, 1,* 347–355.

Kolers, P. A., & Roediger, H. L., III (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23,* 425–449.

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. (1975a). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 271–279.

MacLeod, C. M. (1975b). Release from proactive interference: Insufficiency of an attentional account. *American Journal of Psychology, 88,* 459–465.

MacLeod, C. M. (1976). Bilingual episodic memory: Acquisition and forgetting. *Journal of Verbal Learning and Verbal Behavior, 15,* 347–364.

MacLeod, C. M. (1988). Forgotten but not gone: Savings for pictures and words in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 195–212.

MacLeod, C. M. (1989a). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 13–21.

MacLeod, C. M. (1989b). Word context during initial exposure influences degree of priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 398–406.

MacLeod, C. M. (1991a). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

MacLeod, C. M. (1991b). John Ridley Stroop: Creator of a landmark cognitive task. *Canadian Psychology, 32,* 521–524.

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General, 121,* 12–14.

MacLeod, C. M. (1996). How priming affects two speeded implicit tests of remembering: Naming colors versus reading words. *Consciousness & Cognition, 5,* 73–90.

MacLeod, C. M. (1998a). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.

MacLeod, C. M. (1998b). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26,* 201–211.

MacLeod, C. M. (2010). *I said, you said: The production effect gets personal.* Manuscript submitted for publication.

MacLeod, C. M., & Bors, D. A. (2002). Presenting two color words on a single Stroop trial: Evidence for joint influence, not capture. *Memory & Cognition, 30, 789*–797.

MacLeod, C. M., Chiappe, D. L., & Fox, E. (2002). The crucial roles of stimulus matching and stimulus identity in negative priming. *Psychonomic Bulletin & Review, 9,* 521–528.

MacLeod, C. M., & Daniels, K. A. (2000). Direct vs. indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review, 7,* 354–359.

MacLeod, C. M., Dekaban, A. S., & Hunt, E. B. (1978). Memory impairment in epileptic patients: Selective effects of phenobarbital concentration. *Science, 202,* 1102–1104.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 126–135.

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., & Hodder, S. L. (1998). Presenting two color words on a single trial does not alter Stroop interference. *Memory & Cognition, 26,* 212–219.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior, 17,* 493–507.

MacLeod, C. M., Jackson, R. A., & Palmer, J. (1986). On the relation between field dependence and spatial ability. *Intelligence, 10,* 141–151.

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. J., & MacLeod, C. M. (1992). Re-enacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General, 121,* 145–176.

Mathews, N. N., Hunt, E. B., & MacLeod, C. M. (1980). Strategy choice and strategy training in sentence-picture verification. *Journal of Verbal Learning and Verbal Behavior, 19,* 531–548.

Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review, 67,* 16–31.

Nelson, T. O. (1971). Savings and forgetting from long-term memory. *Journal of Verbal Learning & Verbal Behavior, 10,* 568–576.

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.* Oxford, United Kingdom: Holt, Rinehart & Winston.

Parker, E. S., Birnbaum, I. M., & Noble, E. P. (1976). Alcohol and memory: Storage and state dependency. *Journal of Verbal Learning & Verbal Behavior, 15,* 691–702.

Poltrock, S. E., & MacLeod, C. M. (1977). Primacy and recency in the continuous distractor paradigm. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 560–571.

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review, 74,* 392–409.

Pritchard, R. M., Heron, W., & Hebb, D. O. (1960). Visual perception approached by the method of stabilized images. *Canadian Journal of Psychology, 14,* 67–77.

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

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence–picture verification. *Cognitive Psychology, 40,* 261–295.

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 3–10). New York: Cambridge University Press.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Seuss, Dr. (1978). *I can read with my eyes shut.* New York: Random House.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.

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

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Stone, M., Ladd, S. L., Vaidya, C. J., & Gabrieli, J. D. E. (1998). Word-identification priming for ignored and attended words. *Consciousness and Cognition, 7,* 238–258.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 12,* 242–248.

Szymanski, K. F., & MacLeod, C. M. (1996). Manipulation of attention at study affects an explicit but not an implicit test of memory. *Consciousness & Cognition, 5,* 165–175.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1020–1034.

Wilson, D. E., MacLeod, C. M., & Muroi, M. (2008). Practice in visual search produces decreased capacity demands but increased distraction. *Perception & Psychophysics, 70,* 1130–1137.

Wilson, D. E., Muroi, M., & MacLeod, C. M. (in press). Dilution, not load, affects distractor processing. *Journal of Experimental Psychology: Human Perception and Performance.*

Received September 28, 2010
Accepted September 29, 2010 ■