Generating makes words memorable, but so does effective reading

IAN BEGG, EDE VINSKI, LINDA FRANKOVICH, and BRIAN HOLGATE
McMaster University, Hamilton, Ontario, Canada

In many experiments, memorial benefits have been found when subjects generate items from fragments rather than read items in their complete forms. Does generation cause this difference, or are subjects disposed to adopt different strategies when generating as opposed to reading? If generating causes the difference, items processed in the same way apart from a generative stage should therefore benefit from that generative stage. Our experiments did result in benefits for generating as opposed to reading, but only when the readers processed the words poorly—by pronouncing them. When the readers processed the items well, by imagining them, generating was no better than reading. A new generation effect was found in metamemory. Subjects thought they would remember more generated items than read items; however, the act of making the prediction entailed meaningful processing, and the generated items were not actually remembered better than the read ones.

The generation effect occurs when people remember words presented as fragments better than words that are complete from the start (Jacoby, 1978; Slamecka & Graf, 1978). Since its initial demonstration, the generation effect has been found with different procedures, materials, and conditions of testing. Slamecka and Katsaiti (1987) examined the competing explanations of the generation effect, and concluded that each was unsatisfactory as a general account. The recent explanations have come from opposite directions. Some researchers have asked whether there really is a generation effect at all; generating and reading may dispose subjects to process items differently, but the difference in processing, rather than generation per se, may be responsible (Begg & Roe, 1988; Begg & Snider, 1987; Slamecka & Katsaiti, 1987). Others accept many memorial differences between generated items and read items as generation effects; the multifactor approach, for example, proposes that generation poses a complex problem that subjects may solve in different ways, each of which can influence performance on memory tests (McDaniel & Waddill, 1990; McDaniel, Waddill, & Einstein, 1988).

Our focus is on read items more than generated items. In most explanations of the generation effect, it is tacitly assumed that the locus of the effect lies in memory for generated items, probably because nobody dubbed the out-

come the "reading effect." Begg and Snider (1987) argued that generating has a defined objective—namely, the production of a word that meets the standard set by the initial fragment (cf. Begg, Snider, Foley, & Goddard, 1989). Therefore, generating is a discriminative encoding process that prepares the subject for some types of test. It is less clear what the objective is for readers, especially if they are not told about the upcoming test. Memory for read items will depend on how they are encoded, which in turn will depend on whatever strategies and preferences arise in meeting the perceived demands of the task. There will, obviously, be a "generation effect" on any test if readers do not engage in test-appropriate processing to the same extent that generators do. But are these advantages for generated items because they are generated, or because the tacit task for readers is off track?

In McDaniel et al.'s (1988) three-factor account, generation poses a puzzle that subjects may solve with reference to (1) information that is specific to the to-be-generated target, (2) information that relates the target to the informational context in which it is generated, or (3) information that relates some of the targets in a list to others. If the demand to generate alerts subjects to information that readers fail to encode, readers will fare poorly on tests that benefit from that information. If generation causes those differences in encoding, there should be an advantage for generated items over read items that are processed in the same way except in the generative stage.

We believe that the approaches differ more in how they define causality than they do in substance. To illustrate, contrast intentional learning, in which subjects are informed that a test will follow, with incidental learning, in which subjects are not informed that a test will follow. Intent to learn is correlated with improvement in memory, but is intent the cause of the improvement? The acid test was performed by Hyde and Jenkins (1973), who compared in-

This research was supported by Operating Grant OGP0008122 to Ian Begg, from the Natural Sciences and Engineering Research Council of Canada. Experiment 1 was part of an honors thesis by Ede Vinski, Experiments 2 and 4 were part of an honors thesis by Linda Frankovich, and Experiment 3 was part of an honors thesis by Brian Holgate, all supervised by Ian Begg. We are grateful to Ann Anas, Douglas Needham, and Michael Halldorson for their many insightful comments and criticisms of earlier versions of the article. Correspondence should be addressed to Ian Begg, Department of Psychology, McMaster University, Hamilton, Ontario L8S 4K1, Canada.

tentional and incidental procedures in which people were specifically told to process items meaningfully or superficially. The result was that processing, not intent, was associated with differences in memory. Intent disposes people to process items in memorable ways; the processes cause the improvements in memory, and they will yield good memory whether or not they are used with an intent to learn.

The generation effect is sensitive to intent. For example, Watkins and Sechler (1988) found a large generation effect in incidental learning conditions (.33 > .08), but the effect was greatly reduced in intentional conditions (.37 vs. .29). They concluded that incidental learning gave a clear picture of the generation effect. However, the generation effect is defined by a difference between two means; the generation effect is large under incidental conditions because very few read items are remembered. The numbers say that generating ensures reasonably good memory whether or not subjects study with a test in mind, but reading without intent does not dispose subjects to engage in processing that is especially useful.

McDaniel et al. (1988) found that intent eliminated the generation effect in free recall; if subjects wrote target items that were accompanied by category names, there was a generation effect in recall of targets in incidental conditions (.57 > .43), but not in intentional conditions (.56 vs. .54). Similarly, if each target was accompanied by a unique associate, there was a generation effect in incidental conditions (.44 > .30) but not in intentional conditions (.46 vs. .43). In each case, readers were at a disadvantage unless they were warned that a test would follow. Which differences reported by McDaniel et al. survive intent to learn? Begg, Snider, et al. (1989) concluded that only recognition of targets continued to show generation effects when readers processed items with a test in mind. They proposed that generation causes targets to be distinctively encoded, but that all other effects correlated with the difference between reading and generating result from different demands for readers and generators.

Multifactor approaches broaden our vision and enrich our experiments, leading to the knowledge that certain manipulations are correlated with differences in memory. Rather than enstate a new process for each accomplishment that the learner demonstrates, we propose that stringent requirements must be met before a correlation will be called a cause. In our view, most researchers have placed too little emphasis on reading, which is not a single, simple process. We are struck, each time we review the literature, with how little change from experiment to experiment there is in memory for generated items, compared with the changes for read items.

In the present article, we present four experiments. In Experiments 1 and 2, we ask whether there is a true generation effect when readers and generators process information in the same way except in the generative stage. In both experiments, there were generation effects when the controlled process consisted of pronunciation, but not when it consisted of imagery. In Experiments 3 and 4, we ex-

plored subjects' cognitions about reading and generating; in both experiments, there were generation effects in people's predictions about memory, but no generation effects in memory.

EXPERIMENT 1

In Experiment 1, we measured recognition memory for read and generated items. Our major concern in Experiment 1 was the nature of the discrimination required at test. Begg, Snider, et al. (1989) proposed that generated targets were encoded as distinctive and prominent members of whatever family was activated by the context. Accordingly, they should be at an advantage if tests require subjects to discriminate them from other members of that family, unless the reading process provides as much discriminative information as generating does. In several experiments, generation effects have been found in recognition of targets studied in related contexts (e.g., Begg, Snider, et al., 1989; McDaniel & Waddill, 1990; Slamecka & Graf, 1978). However, there have been no reports of generation effects when recognition memory has been compared between pure readers and pure generators who have studied unrelated words.

If the distinctiveness account is right, it should pertain to unrelated words. If generating discriminates an item from subjectively generated alternatives, the item should be distinct from those alternatives on a test. But what is the family when single word fragments are the stimuli? The task imposes semantic and orthographic constraints that a candidate solution must satisfy. Do these two sources of candidates, which presumably are part of the tacit cognitive context of initial generation, leave any discriminatively useful residue? The study list in Experiment 1 consisted of words that were unrelated to each other, but a minority of them were chosen because they have very close relatives that we could use as distractors at test. For instance, will subjects be better at choosing which of *maiden* and *damsel* is "old" if they have seen *mai-en* rather than *maiden*? Will the choice between *irritation* and *irrigation* differ if subjects have seen *irritation* rather than *irr-tat-on*? We chose pairs of close synonyms and pairs of lookalikes (differing by one letter) and included one member of each pair in the study list, of which the majority of items were unrelated words. Then, at test, the subjects were asked which of two items was the old one; one member of each pair was old, and the other was its semantic or orthographic close relative.

Steps were taken to remove dispositional differences between readers and generators. First, generating and reading were manipulated between subjects. Previous researchers (e.g., Glisky & Rabinowitz, 1985) had found that subjects who read some words and generated others recognized more generated words than read words. However, the generation effect occurs because words read in mixed lists are recognized worse than words read in pure lists, which are at about the same level as words generated in pure and mixed lists (Begg & Roe, 1988; Begg & Snider,

1987). Begg and Snider (1987) have proposed that the presence of to-be-generated fragments in a list disposes subjects to process complete words more cursorily than they normally would. Even if target words are studied in the context of related words, the use of pure lists often eliminates the generation effect or reduces its magnitude. Slamecka and Katsaiti (1987), for example, found no generation effect in free recall of target words studied with related context words. Hirshman and Bjork (1988) also found relatively small generation effects in free recall, with much larger effects when the studied contexts were given as cues. Differences that truly reflect ease of learning will be effective between subjects.

Second, subjects were told there would be a test of memory. True differences will occur when subjects are informed about the test. In many experiments, the generation effect has been eliminated or attenuated under intentional conditions, usually because intent improves readers' memory; readers who do not expect a test may not process material in a way that is useful for them when they do the test.

Third, subjects were told how to process the items, to permit a comparison between generated items and read items processed in the same way except during the generative stage. We chose study procedures that are associated with differences in memory (Begg, 1982). One involved imagery: people were to imagine each word, and hold the image until the next word appeared. The other was pronunciation: people were to rehearse each word silently until the next word appeared. Begg and Roe (1988) had people imagine or pronounce words in pure or mixed lists, and they compared recognition of those words to recognition of generated words. Recognition of generated words fell between that for pronounced and imagined words. In Experiment 1, we compared generate-and-imagine, generate-and-pronounce, read-and-imagine, and read-and-pronounce conditions. We hypothesized that adding a generative stage to imagining would not increase recognition, whereas adding a generative stage to pronouncing would enhance retention. In other words, the addition of a generative stage to these processes should reduce their differential value in preparation for a test of recognition.

In summary, the subjects studied a list of single unrelated words or word fragments. The subjects either imagined the targets or pronounced them. After study, their recognition memory was tested. One test was a yes/no recognition test on which the new items bore no special resemblance to the studied items. The other test required a two-alternative forced choice between studied items and either close synonyms or close lookalikes.

Method

Subjects. A total of 175 students of introductory psychology at McMaster University participated by electing a participation option of the course; the procedure has been approved by the Ethics Committee. Testing was carried out with 20 groups of 8–13 subjects each; the subjects assigned themselves to groups by signing sheets. Four groups were assigned at random to each of four experimental conditions; for the imagery conditions, 43 subjects generated and 46 read; for the pronunciation conditions, 44 generated and 42 read.

Materials. The study list contained 135 words, of which the first 10 and last 11 were untested fillers. The middle 114 words were randomized in three blocks of 38, each with 25 unrelated words, 5 synonyms, and 8 lookalikes. Item selection will now be described for each type of item.

Unrelated words were selected from a pool of 171 nouns with moderate imagery (*I*) values ($3.60 < I < 5.99$), and Thorndike-Lorge frequency (*F*) values of 20 or greater; nouns were selected from Paivio, Yuille, and Madigan (1968) and from an extended norm list provided by A. Paivio. Twenty-one of these words were fillers. The other 150 were assigned at random to be studied words or distractors on the test, with mean *I* and *F* about equal for the two sets. The test had the 75 studied words and the 75 distractors, each with a typed "Y" and "N" to its left. Each block of 10 tested items had 5 old and 5 new items.

The intent with synonyms was to require subjects to make a close semantic discrimination at test, but to have studied the items in the context of semantically unrelated words. First, we selected a pool of 15 pairs of close synonyms; the pairs were *answer-reply*, *argument-quarrel*, *captive-hostage*, *cellar-basement*, *courage-bravery*, *damsel-maiden*, *dinner-supper*, *error-mistake*, *freedom-liberty*, *lawyer-attorney*, *pupil-student*, *stream-brook*, *street-avenue*, *tavern-saloon*, and *victor-winner*. We consulted many reference sources, and chose pairs we felt were close in meaning, without glaring differences in frequency of usage. One word from each pair was selected at random to be the studied item. The test for these items was a two-alternative forced choice; each pair appeared on the test, and the subject was to circle which one had been in the study list.

The intent with lookalikes was to require subjects to make a close orthographic discrimination at test, but to have studied the items in the context of orthographically unrelated items. Lookalikes were selected from a pool of 24 pairs of words differing by one letter that a group of us came up with while at the pub on a Friday afternoon; the pairs were *addition-audition*, *altitude-attitude*, *banner-banker*, *carbon-carton*, *charity-clarity*, *crutch-crunch*, *demotion-devotion*, *department-deportment*, *detection-detention*, *farmer-former*, *hamper-hammer*, *irrigation-irritation*, *island-inland*, *jingle-jungle*, *manner-manger*, *notice-novice*, *phonograph-photograph*, *plaster-platter*, *plunder-plunger*, *prison-poison*, *protector-projector*, *squire-square*, *sunset-subset*, and *vocation-vacation*. One member of each pair was selected at random to be the studied item. The test for these items was a two-alternative forced choice; each pair was on the test, and the subject was to circle which one had been studied. The 24 lookalikes were sorted into two sets of 12 that were treated differently in generation conditions; for ambiguous items, the letter that distinguished the studied item from its distractor was absent from the fragment (*cru-ch* does not distinguish *crutch* from *crunch* until feedback is given), and for unambiguous items, the distinguishing letter was part of the fragment (*pri-on* can be *prison* but not *poison*).

The study list was recorded on videocassette from a computer-generated display. Two recordings were made, at the rate of 6 sec per item. For reading, each word was in its normal form for the entire presentation. For generation, each word was in fragmentary form for the first 2.5 sec of the presentation and was intact for the remaining 3.5 sec.

Procedure. People studied a list of words and then did two tests. The first test was a yes/no recognition test with 75 old words and 75 new words that bore no special relation to the old words. The second test was a two-alternative forced-choice test on which 15 of the tested pairs were close synonyms that bore no special orthographic relationship to each other, and 24 of the tested pairs were close orthographic neighbors that bore no special semantic relationship to each other.

The design was 2×2 between subjects: words were read or generated, and subjects were instructed to pronounce each word silently or to imagine its referent during the time available. The subjects were informed that there would be a test, so that they should

read all complete words carefully, and, for fragmentary stimuli, they should generate words and check them against the feedback.¹ Within subjects, each person completed a yes/no recognition test for unrelated items, for which 7 min were allowed, and then completed a two-alternative forced-choice test, circling the "old" item in each pair; the subjects were warned that the paired members would be very similar to each other in some way. The second test took less than 3 min to complete.

Results

Means in the article are expressed as proportions, and alpha was set at .05 for statistical inferences. Imagery correctly rejected more new unrelated items (read = .85, generated = .83) than pronouncers did (read = .74, generated = .75) [$F(1,171) = 17.6$, $MS_e = 0.014$]; there was no reliable effect of generating rather than reading.

We turn now to recognition of old items. In overview, there was a large advantage for imagery over pronunciation, but only for readers; with generators, the effect was greatly reduced. Generating was better than reading if the readers had pronounced the words, but generating was worse than reading if the readers had imagined the words. However, the extent to which generation aided pronunciation was not always reliable, nor was the extent to which generation hurt imagery.

Table 1 shows four indices of recognition for readers who imagined or pronounced complete words and generators who imagined or pronounced words that first appeared as fragments. To the right is the imagery effect, computed by subtracting the mean for pronouncers from the mean for imagers. At the bottom of each section is the generation effect, defined by subtracting the mean for readers from the mean for generators.

The interaction between processing instructions and whether the words were read or generated was reliable

for each measure. The first section shows yes/no recognition of the unrelated items [$F(1,171) = 5.41$, $MS_e = 0.025$, $d = .068$] (means that differ by .068 are reliably different). The second section shows two-alternative forced choices between old words and new synonyms (*maiden* vs. *damsel*) [$F(1,171) = 4.73$, $MS_e = 0.014$, $d = .051$]. The two bottom sections show two-alternative forced choices that correctly identified old words rather than new words that look like them (*irrigation* vs. *irritation*) [for the unambiguous lookalikes, $F(1,171) = 9.81$, $MS_e = 0.014$, $d = .051$, and for the ambiguous lookalikes, $F(1,171) = 7.22$, $MS_e = 0.011$, $d = .045$].

The interactions are easily interpreted in terms of the size of the difference between imagery and pronunciation. For each measure, there was a reliable advantage for imagery when words were read in their normal complete form. However, within generation conditions, the advantage for imagery was greatly reduced. Hence, generation closes the gap between a good way to study and a bad one. Does generation close the gap by bestowing a benefit on pronunciation, or by imposing a cost on imagery? In the next section, we will show that the answer depends on the task.

We now ask where the generation effect was reliable, within processing. For pronouncers, generating was superior to reading in every case, but the advantage was reliable only in discrimination between old words and close synonyms. This is the first demonstration of a between-subjects generation effect with items studied in unrelated contexts. If generation calls attention to the letters in a fragment, they may be encoded, aiding discrimination from semantically similar alternatives if the initial process includes only superficial attention to meaning. For imagers, generating was reliably inferior to reading only for the two kinds of lookalikes. Imagery encodes items meaningfully, and superficially similar orthographic neighbors, which are very different in meaning, are not confused with them. Generation, by calling attention to letters, may lead to the encoding of information that is shared by the old item and its new distractor.

Discussion

In Experiment 1, we examined recognition memory for words that were imagined or pronounced. The words initially appeared either in complete form or as fragments. In each task the order of means was the same, with best memory for read-and-imagine, followed by generate-and-imagine, then generate-and-pronounce, and read-and-pronounce. The negative generation effect in imagery conditions was reliable only when targets had to be discriminated from orthographic neighbors, and the positive generation effect in pronunciation conditions was reliable only when targets had to be discriminated from closer semantic neighbors at test than had been implied at study. A simple explanation is that subjects have some memory for the literal form of generated stimuli (cf. Johns & Swanson, 1988; Nairne & Widner, 1987). That little extra memory aids discrimination between semantically similar

Table 1
Recognition of Read or Generated Words Studied by Imagining Them or Pronouncing Them in Experiment 1

	Instructions		Imagery Effect (Imagine - Pronounce)
	Imagine	Pronounce	
Unrelated Words			
Read	.856	.712	.144*
Generate	.801	.769	.032
Generation effect	-.055	.057	
Synonyms			
Read	.897	.782	.115*
Generate	.879	.842	.037
Generation effect	-.018	.060*	
Lookalikes: Unambiguous			
Read	.946	.788	.158*
Generate	.847	.801	.046
Generation effect	-.099*	.013	
Lookalikes: Ambiguous			
Read	.944	.829	.115*
Generate	.897	.869	.028
Generation effect	-.047*	.040	

*Reliable difference; see text for specific critical differences.

but orthographically distinct alternatives, and hurts discrimination between orthographically similar but semantically distinct alternatives.

We now play "what if" with the data in Table 1. What if read-and-pronounce simulates the "lazy" reading attributed to subjects who read some items and generate others, but generate-and-imagine simulates the processing that would be received by generated items in the same lists? If so, we would have obtained generation effects on all our measures; for unrelated words, $.801 - .712 = .089$; for synonyms, $.879 - .782 = .097$; for unambiguous lookalikes, $.847 - .788 = .059$; for ambiguous lookalikes, $.897 - .829 = .068$. Thus, depending on the happenstantial choice of processing by readers, we could see benefits, costs, or null effects of generating compared with reading.

EXPERIMENT 2

In Experiment 2, as in Experiment 1, we examined memory for items that were generated or read, with processing controlled by instructions from the experimenter. This time, however, the stimuli were pairs of categorically related words, like *tiger-horse*. The first word provided a context for the second word, the target, which was either in complete or in fragmentary form. In several experiments, generation effects have been found in recall of targets cued by contexts, tested between subjects (Begg, Snider, et al., 1989; Hirshman & Bjork, 1988; McDaniel et al., 1988). As in Experiment 1, generating and reading were manipulated between subjects, and the subjects were informed that there would be a test of memory. They studied pairs of categorically related words by pronouncing them, by imagining them, or without particular directions about how to process the items.

Method

Subjects. A total of 135 students of introductory psychology from McMaster University were subjects. Twelve groups of 10-13 subjects were tested; 2 of the groups were assigned at random to each of six experimental conditions.

Materials. Categorical exemplars were chosen from 50 of Battig and Montague's (1969) categories. From each category, we chose two exemplars that were between the 3rd and 11th most frequently

produced members of the categories, excluding proper nouns and compound expressions. The exemplars from each category were paired for the study list, with one of them chosen at random to be the context, the other the target word.

The 50 pairs were recorded on videocassette from a computer-driven display, in two versions. For readers, all words were in their usual complete form. Each context appeared on its own for 2 sec, then the target appeared beneath it and the two remained on the screen for 2.5 sec, and then the target remained alone on the screen for 2 sec; there was a 1.5-sec blank period before the cycle began for the next pair. The generated version of the list was the same, except that the target members of all pairs were lacking one or two letters during the 2.5 sec in which both words were visible.

The test gave a cue for each of the 50 targets (e.g., *horse*). Twenty-five cues were the contexts (e.g., *tiger-?*), and the other 25 were the contexts augmented by the category labels, paraphrased as one or two words (e.g., *animal: tiger-?*). A second version of the test was constructed, so that targets cued by contexts in the first version received augmented cues in the second version, and vice versa. Each version was given to about half the subjects in every session.

Procedure. The design was 3×2 between subjects, in which instructions (control, pronunciation, imagery) were contrasted with formats (reading vs. generating). The subjects were encouraged to listen carefully to all instructions and to do their best on all tests. They were told that their memory would be tested, but they were not told the specific type of test that they would receive. They were told how the words would look on the monitor, and that the pairs would be related. Generators were told to guess silently each bottom word before its intact form appeared, and readers were told to read words carefully for memory. Between study and test there was a delay of about 15 min, during which time the subjects studied and were tested on other material that was unrelated to the materials in the list.

Results

Table 2 shows recall of read and generated items, with context cues and with augmented cues, after the pairs had been processed by means of pronunciation, imagery, or a control procedure. To make the results comparable to those in Experiment 1, we have computed the difference between imagery and pronunciation, as well as the size of the generation effect. It is obvious that generation removed the advantage for imagery, and that there was a generation effect only when generation was compared to poor reading by pronouncers. The interaction between instructions and format was reliable [$F(2,129) = 3.67$,

Table 2
Recall of Categorical Targets Cued by Their Contexts, or Their Contexts Augmented by the Category Names, in Experiment 2

	Instructions			Imagery Effect (Imagine – Pronounce)
	Controls	Imagine	Pronounce	
Contexts as Cues				
Read	.577	.613	.421	.192*
Generate	.502	.582	.556	.026
Generation effect	–.075	–.031	.135*	
Contexts Augmented with Category Names				
Read	.610	.651	.478	.173*
Generate	.573	.669	.629	.040
Generation effect	–.037	.018	.151*	

*Reliable differences; see text for specific critical differences.

$MS_e = 0.066$, $d = .08$], and recall was better with two cues than with one [.60 vs. .54; $F(1,129) = 30.4$, $MS_e = 0.0079$].

Discussion

Experiment 2 provided good confirmation of Experiment 1, despite the change from unrelated materials to categorically related materials. With pronunciation as the process used by readers, reading falls short of generating in recall of targets with their contexts as cues. However, if readers encode words distinctively by imagining them, memory is as good as it is for generated words.

The conclusion from the first two experiments is a compromise between two extreme positions. Generating demands meaningful processing that ensures a level of memory discriminability beyond the level achieved by readers who merely pronounce but not beyond the level achieved by readers who process words imaginally. Hence there is a true generation effect when all other demands of the experiment are controlled, provided that the level of control leads to inferior processing and mediocre memory. If the level of control is imaginal, however, the discriminability bestowed by generation fails to give an advantage in memory because imagery already distinguishes items precisely.

EXPERIMENT 3

Much of this article has been focused on subjects' ideas about task demands during reading and generating. In Experiment 3, we explored those ideas directly. The experiment was designed to be procedurally identical to other research we have done, except that the subjects were asked to predict whether they would remember or forget each word in a list of single, unrelated items. Readers saw all the words in their usual complete form. For generators, the initial stimuli were fragments of those words. Other subjects received a mixed list, in which half the items were words and half were fragments. In previous research with unrelated items (Begg & Roe, 1988; Begg & Snider, 1987), the generated words from a mixed list have been recognized better than the read words from the same list, without a difference between pure reading and pure generating.

In Experiment 3, we were interested in people's cognitions at the time of studying words or fragments, as indicated by predictions of memorability. There were two questions. First, would memory predictions differ for generated and read items—would people expect to remember more items after generating than after reading? And second, would the demand to predict memory influence recognition—would actual memory for generated and read items show the usual results?

Begg, Duft, Lalonde, Melnick, and Sanvito (1989) had people predict whether each item would succeed or fail on a future memory test. They concluded that memory predictions are implicit ratings of the ease of processing; people expect better memory for items that are easy to process than for items that are harder to process. The

predictions can be right; for example, concrete words are easier to process than abstract words and are also recognized better than abstract words. But the predictions can be wrong; common words are easier to process than rarer words but are recognized less well than rarer words. If rated memorability depends solely on ease of processing, subjects should expect better memory for read than for generated words because generating a word from a fragment is more difficult than reading a complete word.

However, Begg, Duft, et al. (1989) also compared memory predictions for pairs of words studied in different ways. They found that people expected more success for interactively than for separately imagined pairs if they studied half the items each way, but without a difference in predictions in a between-subjects contrast. That is, people believe that interactive imagery is a better means of study than separate imagery, but they can use that knowledge discriminantly only if some pairs have been processed each way. Our intuition is that people think generating is good for them. If so, subjects who read some items and generate others will predict better memory for the generated ones than for the read ones. Furthermore, the contrast between pure readers and pure generators is different from the contrast between interactive imagers and separate imagers, who did not differ in predicted levels of success. The difference is that subjects who are given fragments to complete encounter an implicit contrast with reading, which is the usual way they see words. Hence, if they believe that generating is a good thing to do, they can adopt a more lenient criterion for predicting memory than they usually would, and they will predict more success than the readers do.

The final question is whether the demand to predict memory will produce a different pattern of results than usual. We have reviewed research in which foreknowledge of a test eliminates the generation effect, presumably because readers who expect a test engage in more useful processing than readers who are just living through the experience of seeing a list. The question "How likely are you to remember this word later, compared to other words in the list?" entails discriminative processing of the item; Begg, Duft, et al. (1989) found better recognition after people had predicted memorability than in a control condition. In short, there should be less of a generation effect than usual, even in the fertile ground of a mixed list.

Method

Subjects. The subjects were 74 students of introductory psychology at McMaster University. They were tested in six groups; two groups were assigned at random to each of the three experimental conditions, with 22–26 subjects in each condition.

Materials and Procedure. The subjects studied a list of 100 words, of which the first 12 and last 4 were untested fillers. While studying the items, the subjects rated each item for memorability, using a prepared sheet on which each of the numerals from 1 to 100 appeared next to a 7-point scale; 1 stood for *least memorable* and 7 for *most memorable*. The studied words and 84 distractors for the test were selected from Paivio et al. (1968) to reflect a wide range of imagery and frequency. Words were assigned to be new or old at random, except that mean *I* and *F* were equal in the sets.

Four versions of the list were recorded on videocassette, each from a computer-generated display. For pure reading, all words appeared in their complete form for 6.5 sec, with 1 sec between them. For pure generating, all words had one or two letters missing for the first 4 sec; then the words appeared in complete form for 2.5 sec. There were two mixed lists in which half the words were complete and half were fragmentary for the first 4 sec; the complete words in one were the fragments in the other and vice versa. Each mixed list was shown to a group of subjects, and each pure list was shown to two groups. After study, subjects were given a test of recognition memory. The test had 84 old items and 84 new ones; subjects circled either "Y" (old) or "N" (new) for each tested item.²

Results

One result was that ratings of memorability revealed a generation effect both between and within subjects. Pure generators gave higher mean ratings than did pure readers [$4.82 > 3.72$; $F(1,46) = 19.4$, $MS_e = 2.96$]; with mixed lists, people gave higher ratings to the generated items than to the read items [$4.31 > 3.70$; $F(1,25) = 23.4$, $MS_e = 0.41$]. Therefore, our subjects believed that generating is a good thing for memory.

However, generating had very little effect on memory. The pure generators were unreliably worse than the pure readers [.84 vs. .87; $F(1,46) = 1.58$, $MS_e = 0.036$]; in the mixed lists, the generated items were recognized slightly better than the read items [.85 vs. .83; $F(1,25) = 3.34$, $MS_e = 0.0060$, $p < .08$]. Each of the three groups had the same false alarm rate (.11, $MS_e = 0.0052$).

Discussion

The subjects expected that generating would lead to better memory than would reading. However, these same people, when tested for recognition, showed no generation effect between subjects who saw pure lists, and only a tiny effect (.02) within subjects who saw mixed lists. Hence specific attention to memorability was sufficient to eliminate the generation effect. Although we did not include a group whose subjects did not rate memorability, we did another experiment shortly after Experiment 3, in which we included unrelated words tested for recognition. In that experiment, we found that pure readers were about the same as pure generators (.66 vs. .69, $MS_e = 0.037$), and that in a mixed list, the read items fell short of the generated items [.60 < .70, $F(1,12) = 16.4$, $MS_e = 0.0040$].

EXPERIMENT 4

Experiment 4 was much like Experiment 3, except that the materials were pairs of categorically related exemplars, and memory was tested by recall, cued with category names. Begg, Snider, et al. (1989) found reliable generation effects when people studied pairs of related exemplars and then received the category names as cues. For the present experiment, the only intended difference that we introduced was to have subjects rate memorability of the targets with the contexts as cues.

Method

Subjects. The subjects were 72 students of introductory psychology from McMaster University. They were tested in six groups;

two groups were assigned at random to each of three conditions, with 22–28 subjects in each.

Materials and Procedure. The subjects saw 50 pairs of related exemplars, selected from 50 different categories (Battig & Montague, 1969). The exemplars were 5–11 letters in length, and none was a member of any other category. The exemplars were from the 3rd to the 11th most frequently produced members of their categories. They were assigned at random to be contexts or targets in their pairs.

Four versions of the list were recorded on videocassette. For pure reading, all words were complete during the entire presentation. For pure generation, each target lacked one or two letters.³ Two mixed versions of the list had half the targets complete and half fragmentary; the fragments from one list were the complete words in the other. In the mixed lists, half the targets in each 10-item block were fragments. Each pair appeared in three stages, with 1.5 sec between pairs. The context appeared alone for 2 sec, then the target appeared below it in fragmentary or complete form for 2.5 sec, and then the complete target appeared alone for 2 sec.

While studying the items, the subjects rated each item for memorability, using a sheet on which each of the numerals from 1 to 50 appeared next to a 7-point scale; 1 stood for *least memorable* and 7 for *most memorable*. The instructions were to rate items with regard to a test in which the contexts would be the cues for the targets. The extralist cues for the test were the names of the categories, written as single words or short phrases. Each had two blank spaces in which subjects were to recall the two studied exemplars.

The subjects were encouraged to pay attention to all instructions and to try hard on all tests. They were to read all complete words carefully and to try to generate any fragmentary words before feedback was given. One group of subjects saw each mixed list and two saw each unmixed list, for a total of six groups.

Results

As in Experiment 3, pure generators expected more success than readers did [$4.70 > 4.06$; $F(1,42) = 7.13$, $MS_e = 1.26$], and subjects with mixed lists expected more success for generated than for read items [$4.49 > 4.20$; $F(1,27) = 5.45$, $MS_e = 0.18$]. Despite their confidence in generation as a way to study, their recall revealed no generation effect ($MS_e < 0.02$). Readers recalled .64 of the targets and generators recalled .62; with mixed lists, the subjects recalled .57 of the generated targets and .55 of the read ones.

Discussion

The ratings indicated that the subjects expected generation effects, but recall with extralist cues revealed no generation effects. Although there was not a condition in Experiment 4 in which subjects studied without rating memory, we did obtain the usual result in another experiment conducted within days of Experiment 4; with category names as cues, pure generators recalled .58 of generated targets and pure readers recalled .40 [$F(1,22) = 7.26$, $MS_e = 0.051$], and the difference was about the same with a mixed list, .60 vs. .43 [$F(1,9) = 9.07$, $MS_e = 0.016$].

GENERAL DISCUSSION

First we will summarize the results of our experiments. Where did we find advantages favoring generated items over read items? Between subjects, we found only two memory advantages for generated items. First, generated targets were more discriminable from close synonyms

(*saloon* vs. *tavern*) than read targets were, provided that the readers had pronounced the items when they studied them. Second, recall of generated targets cued by their contexts (*tiger* → *horse*) exceeded recall of read targets, provided that the readers had pronounced the items when they studied them. We found no benefits of generating over reading if the readers imagined the items.

A new generation effect was found in people's ratings of memorability of generated and read items. People were more optimistic about their future success in remembering generated as opposed to read items. However, the process of making those predictions rendered them inaccurate. Generation had more of an effect on cognition than on memory.

What does it all mean? The easiest way to incorporate all the data into one idea is to shift focus from trying to explain the generation effect to considering what generating does and what reading doesn't do. Reading does not ensure that items will be processed in a way that is appropriate for the test. Generating has a better chance of doing just that. Although there is some question about whether the good thing that generating does is to distinguish the items, or to engage other useful skills that subjects have in their bag of cognitive tricks, it seems clear that it is the result of applying these tricks, not the reason for doing so, that is important. Generating might make items distinctive, and might even guarantee it, but reading can do every bit as good a job, although it is unlikely to do so spontaneously.

What Does Generating Do?

Encoding, by our account, is the progressive discrimination of an item until it is specifically identified in the context of whatever information and experimental demands are in evidence. Memory is the record of those discriminations, and memory performance will be maximal on a test that requires the same discriminations that were made initially. Generating is a particular kind of discriminative encoding. A fragmentary stimulus, together with other contextual information, is the source of response candidates, and it sets the standard that the winning candidate must satisfy. The winner is chosen instead of other candidates and is distinct from them. In summary, generation is a discriminative encoding process, and a generated response is prominent among other candidates brought to mind by the context.

Even if generation does yield distinct memories, we cannot expect ubiquitous memorial advantages for generated items. For one thing, not all tests reflect the particular type of distinctiveness that generation bestows. For another, generation is not the only encoding process that gives distinct memories. In order to decide whether generation does what we think it does, we need results from different memory tests and from different encoding processes. Generation effects are most easily obtained on tests that benefit from item-specific distinctiveness, but even then the effects are eliminated or reduced if readers use

encoding processes that discriminate items as well as generating them does.

We now review and organize much of the background literature and document key points of our account.

Evidence for Item-Specific Distinctiveness

The generation effect often occurs on tests that entail item-specific discriminations, even when there is no generation effect on other tests. The most obvious item-specific test is item recognition, which is especially sensitive to the difference between generating and reading. For example, Jacoby (1983) found a generation effect in recognition memory even though the read words were superior on a test of perceptual identification. Graf (1980) found a generation effect in recognition of the words from anomalous sentences but not in memory for the sentences as units. Begg and Snider (1987) found a generation effect in recognition of words even when the generated ones fell short of read words in cued recall. Begg, Snider, et al. (1989) found a generation effect in recognition memory even when relational measures showed negative effects of generating. McDaniel et al. (1988) found a generation effect in recognition even when there was no generation effect in a prior test of free recall. Thus, recognition memory is sensitive to whatever generation does.

Recognition explicitly requires discrimination between studied and unstudied items, but other tests do so implicitly. If recall of exemplars is cued by category names, for example, the number of related intrusions is equal to the number of false recognitions of related items, implying that intrusions are subjective false alarms (Begg, 1978). Begg, Snider, et al. (1989) included tests of recognition memory and also tests of extralist cued recall, and found similar generation effects in both cases. As in recognition, subjects cued by *animal* were more likely to recall *horse* after seeing *tiger-h-r-e* rather than *tiger-horse*; generated *horses* were especially likely to be recalled alone, rather than with *tiger*, indicating that the difference is item-specific and cannot reflect context-target relations.

In contrast, no published experiments have resulted in a generation effect on a relational measure of memory, but that is absent from an item-specific measure of the target. In many experiments, generation effects have been found that could reflect enhanced context-target relations. For example, Rabinowitz and Craik (1986) used strongly related pairs (*knee-shin*; *knee-sh-n*), for which recall of targets (*shin*) showed a larger generation effect if cued by the old contexts (*knee*; $.72 - .64 = .08$) than if cued by extralist semantic information (*part of leg*; $.57 - .53 = .04$). However, the results were different for weakly related pairs (*kick-shin*), for which there was a generation effect in recall of *shin* with general semantic cues but not if cued by the old contexts (*kick*), and they used only one measure of memory. Begg, Snider, et al. (1989) used both types of measure, and found that some apparently relational effects turn out to be item-specific. In an early experiment, Schwartz (1971) found an advantage

in cued recall if people had generated mediators rather than read them. The effect occurred because cues were more likely to elicit retrieval of generated mediators than of read mediators; once retrieved, the two types of mediator were equally able to retrieve the responses.

In conclusion, the generation effect is readily obtained on item-specific tests, even when the effect is absent with other tests and with other measures.

Generated Targets Benefit from Generation

The general picture from previous research is that the effect of generation is restricted to its targets, not to the contextual information that initiated the process. There are cases in which generated targets benefit from generation but their contexts do not; there are no cases in which contexts benefit from generation but the targets do not. For example, recognition memory for *horse* is better if the original stimulus was *tiger-h-r-e* rather than *tiger-horse*, but recognition of *tiger* is equally good in both cases (Begg, Snider, et al., 1989; see also Slamecka & Graf, 1978). In two recent articles, generation effects have been reported for contexts (Greenwald & Johnson, 1989; McDaniel & Waddill, 1990): Greenwald and Johnson found a generation effect in recognition of contexts, but the recognition test followed a recall test; McDaniel and Waddill tested recognition of contexts unconfounded by a previous test, and did not find a generation effect. Hence, the generation effect in recognition is item-specific, and the target is the item.

Begg, Snider, et al. (1989) found no cases in which recall of contexts cued by generated targets exceeded recall of contexts cued by read targets. Greenwald and Johnson (1989) found a generation effect in recall of contexts, but the small effect was reliable only when experiments were combined, and there was a strong generation effect in memory for the targets; perhaps the generation effect in recall of contexts occurred because more of the targets were recognized when they were given as cues. McDaniel and Waddill (1990) found a more substantial effect, but there was no unique generation effect in recall of the contexts; there were strong effects of generation on recall of targets. We cannot rule out the alternative hypothesis that the readers failed to process items in a way the tests would reflect. This problem is especially acute because the experiments consisted only of incidental learning conditions, to the extreme of replacing subjects who anticipated a test. Furthermore, the results may reflect the materials. McDaniel and Waddill used their list of 36 targets (6 from each of 6 categories), in which each target has a unique context that the experimenters choose for it; *strum-guitar* is one of the pairs. The finding of better recall of *strum* if cued by generated *guitar* rather than read *guitar* may be a case in which the demand to generate calls attention to a strong relation that the experimenters share with the subjects.

In conclusion, there is clear evidence that generated targets exceed read targets on tests on which there is no memorial difference for contexts; there are no cases in

which contexts are at benefit on tests on which targets are not.

To-Be-Generated Fragments Benefit from Generation

A common finding is that the generation effect is larger for coherent than for incoherent sets of information. Incoherent sets in which investigators have failed to find generation effects include nonwords (McElroy & Slamecka, 1982; Nairne, Pusey, & Widner, 1985), unfamiliar phrases and acronyms (Gardiner & Hampton, 1985), anomalous sentences (Graf, 1980, 1982), scrambled passages (Einstein, McDaniel, Bowers, & Stevens, 1984), and pairs of unrelated words (Begg & Snider, 1987). In contrast, there are clear generation effects for words, familiar phrases and acronyms, meaningful sentences and passages, and pairs of related words. Begg, Snider, et al. (1989) concluded that the requirement to generate may call attention to coherence so that sets of items may be selectively processed coherently, but it is not sufficient to establish coherence if none is there to begin with. However, the contrast between coherent and incoherent sets is ambiguous, because it is not clear that the unit for analysis is the same in the two cases. For example, generated sentences exceed read sentences in recall, but not if the generated sentences are anomalous; however, the words in anomalous sentences do benefit from generation (Graf, 1980). Similarly, unrelated pairs do not show generation effects at the level of pairs, but they do in measures of word recognition (Begg & Snider, 1987). It may be the case that the generation task defines the initial fragment as the to-be-remembered unit. The case is clearest with nonwords.

A nonword stimulus is the experimenter-defined unit of analysis, but the initial fragment may be the unit for generating subjects. Generation enhances memory for fragments that have been stimuli for generation; words show larger generation effects if the original fragments are included on the test (Glisky & Rabinowitz, 1985), and nonwords show a generation effect if the test requires identification of original fragments (Nairne & Widner, 1987). Tests of recognition of nonwords often give readers an unfair advantage. For example, if generators see *ohrven* and transpose letters mentally to produce "horven" as a response, whereas readers see *ohrven* and *horven* side by side, then a test for *horven* involves a stimulus that the readers have actually seen, but that the generators have not. Johns and Swanson (1988) found that testing with the complete forms decreased recognition for generated nonwords; there was a generation effect when the fragments were the tested stimuli, or when the complete form had been given as feedback after the generative attempt. To avoid a study-test confound, we can test with the studied fragments, or we can provide feedback after the attempt, to define the unit that the experimenter has in mind. Results indicate that it does not matter whether the solution is achieved by the subject's efforts or is provided as feedback by the experimenter (Kane & Anderson,

1978; McElroy & Slamecka, 1982; Slamecka & Fevreski, 1983); if the generation task is easy, the magnitude of the effect is about the same whether people are given feedback or not (Begg & Snider, 1987).

In conclusion, a piece of the reason why generated targets enjoy item-specific benefits is that a little extra discriminability may be derived from memory for the initial fragment.

Dispositional Effects of Generation

If generating and reading create different perceptions about what subjects should do, there may be strategic differences between the conditions. If so, differences in memory are caused by the processes that those strategies recruit, not by generating. The idea that some ways to read result in memory as good as that for generated items is not new. For example, Donaldson and Bass (1980) found no generation effect in recall of targets cued by contexts when subjects were explicitly told to attend to the relation between targets and the contexts studied with them. Ghatala (1981, 1983) found similar results by using a sentence completion procedure, but the elimination of the effect occurred only when the sentence stems were highly predictive of their completions. Thus, generating may call attention to strong relations, establishing a coherent unit. Reading may or may not call attention to these relations, but if it does, they are remembered. Similarly, generating entails meaningful discrimination of the target in some domain, but reading may or may not. If differential memory after reading and generating reflects differences in processing strategy imposed by the two demands, then the generation effect should be removed if processing is controlled rather than left to the subjects' preferences or guesses as to what is going on. As we saw in Experiments 1 and 2, generated words were recognized better than pronounced words, but they were not recognized better than imagined words.

The Last Word

Generation is a useful addition to the processing armory. Generative processing is a good thing to do and it is a sensible thing to investigate. Reading is also a good thing to do, although under some conditions our subjects do it better than they do under other conditions. In consequence, many instances of the generation effect arise because readers are not disposed to process items in a way that is useful for the test that the experimenter has in mind.

REFERENCES

- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, **80** (3, Pt. 2).
- BEGG, I. (1978). Similarity and contrast in memory for relations. *Memory & Cognition*, **6**, 509-517.
- BEGG, I. (1982). Imagery, organization, and discriminative processes. *Canadian Journal of Psychology*, **36**, 273-290.
- BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, **28**, 610-632.
- BEGG, I., & ROE, H. (1988). On the inhibition of reading by generating. *Canadian Journal of Psychology*, **42**, 325-336.
- BEGG, I., & SNIDER, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 553-563.
- BEGG, I., SNIDER, A., FOLEY, F., & GODDARD, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 977-989.
- DONALDSON, W., & BASS, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning & Verbal Behavior*, **19**, 26-35.
- EINSTEIN, G. O., MCDANIEL, M. A., BOWERS, C. A., & STEVENS, D. T. (1984). Memory for prose: The influence of relational and proposition-specific encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 133-143.
- GARDINER, J. M., & HAMPTON, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 732-741.
- GHATALA, E. S. (1981). The effect of internal generation of information on memory performance. *American Journal of Psychology*, **94**, 443-450.
- GHATALA, E. S. (1983). When does internal generation facilitate memory for sentences? *American Journal of Psychology*, **96**, 75-83.
- GLISKY, E. L., & RABINOWITZ, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 193-205.
- GRAF, P. (1980). Two consequences of generating: Increased inter- and intraword organization of sentences. *Journal of Verbal Learning & Verbal Behavior*, **19**, 316-327.
- GRAF, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning & Verbal Behavior*, **21**, 539-548.
- GREENWALD, A. G., & JOHNSON, M. M. S. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, **17**, 673-681.
- HIRSHMAN, E., & BJORK, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 484-494.
- HYDE, T. S., & JENKINS, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning & Verbal Behavior*, **12**, 471-480.
- JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**, 649-668.
- JACOBY, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 21-38.
- JOHNS, E. J., & SWANSON, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 180-190.
- KANE, J. H., & ANDERSON, R. C. (1978). Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology*, **70**, 626-635.
- MCDANIEL, M. A., & WADDILL, P. J. (1990). Generation effects for context words: Implications for item-specific and multi-factor theories. *Journal of Memory & Language*, **29**, 201-211.
- MCDANIEL, M. A., WADDILL, P. J., & EINSTEIN, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory & Language*, **27**, 521-536.
- MCELROY, L. A., & SLAMECKA, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning & Verbal Behavior*, **21**, 249-259.
- NAIRNE, J. S., PUSEN, C., & WIDNER, R. L., JR. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, **13**, 183-191.
- NAIRNE, J. S., & WIDNER, R. L., JR. (1987). Generation effects with

- nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 164-171.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, 76 (1, Pt. 2).
- RABINOWITZ, J. C., & CRAIK, F. I. M. (1986). Specific enhancement effects associated with word generation. *Journal of Memory & Language*, 25, 226-237.
- SCHWARTZ, M. (1971). Subject-generated versus experimenter-supplied mediators in paired-associate learning. *Journal of Experimental Psychology*, 87, 389-395.
- SLAMECKA, N. J., & FEVREISKI, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning & Verbal Behavior*, 22, 153-163.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 592-604.
- SLAMECKA, N. J., & KATSAITI, K. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, 26, 589-607.
- WATKINS, M. J., & SECHLER, E. S. (1988). Generation effect with an incidental memorization procedure. *Journal of Memory & Language*, 27, 537-544.

NOTES

1. Although we know from previous research that the size of the generation effect is insensitive to whether generation succeeds or fails, reviewers of this manuscript asked whether subjects were actually gener-

ating or were merely waiting for feedback to occur. The editor suggested that we conduct a control experiment. Douglas Needham, in our lab, gave subjects 80 unrelated words, all of which were fragmentary during their initial appearance. For one group of 12 subjects, all 80 complete words appeared as feedback; for these subjects, yes/no recognition against 80 new items had a hit rate of .691. For another group of 13 subjects, only 40 of the fragments were followed by feedback; the other 40 remained fragmentary. Because feedback was unpredictable, these subjects could not simply wait for feedback. The hit rate for the 40 words that received feedback was .694, and the hit rate for the 40 that received no feedback was .656 (overall MS_e was 0.0124). Because recognition was about the same when feedback was perfectly predictable as it was when feedback occurred but was unpredictable, we need not worry about whether subjects actually generated at study.

2. We gave typed versions of the pure generation list and the mixed lists to a total of 150 subjects who had participated in other experiments but had time left in their hour. The mean completion rate for the fragments was .974 for the pure list and .975 for the mixed list. Hence the items are easy to generate successfully. The same subjects then freely recalled the words. The subjects with a mixed list recalled more generated than read items (.221 > .123), with mean recall for unmixed generators at .149; for the interaction between item sets and condition, $F(1,148) = 36.9$, $MS_e = 0.0074$, $d = .028$.

3. A total of 38 subjects from other experiments were given a typed version of this list; they completed an average of 93% of the targets correctly.

(Manuscript received October 1, 1990;
revision accepted for publication February 27, 1991.)

Notices and Announcements

Research Conference on Employees, Eldercare, and the Worksite Sponsored by the National Institute on Aging July 1991

The Behavioral and Social Research (BSR) program of the National Institute on Aging held a Research Conference on Employees, Eldercare, and the Worksite in New York City, July 25, 1991. Sponsored in collaboration with the New York Business Group on Health, the meeting brought together scientific researchers, individuals from the private sector, and those involved in the aging network of service providers.

Scientific presentations reviewed what is known about the burdens and contributions of those caregivers, often middle-aged women, who also have a role as paid workers. Representatives from unions and corporations discussed specific programs in the worksite to ease the strains of combining work and family responsibilities. Practical and methodological problems in conducting aging research at the worksite were identified, and recommendations were made for a future agenda for NIA in studying elder care and the workplace.

For additional information contact Dr. Katrina W. Johnson, Behavioral and Social Research, National Institute on Aging, Building 31, Room 5C32, Bethesda, MD 20892 (phone, 301-496-3136; FAX, 301-402-0051).