

The generation effect and the modeling of associations in memory

STEVEN E. CLARK

University of California, Riverside, California

The search of associative memory (SAM) model of Gillund and Shiffrin (1984) was applied to data of two experiments that examined the generation effect (Slamecka & Graf, 1978). Subjects studied a list of related word pairs, in which they either read both words in the pair or generated the right-hand response term using the left-hand stimulus term plus the response word fragment as generation cues. Experiment 1 manipulated encoding condition within subjects and used an incidental learning procedure. Experiment 2 manipulated encoding condition between subjects and used an intentional learning procedure. Memory was tested with recognition, cued recall, and free recall. A higher order association model gave a better and more parsimonious fit to the results than did an item-level association model. The relationship between various versions of SAM and current accounts of the generation effect are discussed, particularly the two-factor theory of Hirshman and Bjork (1988).

Within the framework of list-learning experiments, words that are generated at the time of study are remembered better than are words that are read. This result, termed the *generation effect*, originally demonstrated by Slamecka and Graf (1978), has been replicated under a variety of conditions.

The standard procedure is as follows: subjects study a list of K word pairs, $A_1B_1, A_2B_2, A_3B_3, \dots, A_KB_K$, where the A_i are referred to as stimulus terms, and the B_i are response terms. The words in a pair are usually related (e.g., MORNING-AFTERNOON). In one condition, the subject reads each of the word pairs on the list. In another condition, the subject is presented with the stimulus word, plus a stem or fragment of the response word and must generate the response. For example, the subject might be given MORNING-#FT#RN##N, where the vowels have been deleted.

A number of experiments consistently show that later recall of the response term, given the stimulus as a cue, is better for responses that were generated at study than for responses that were read (Greenwald & Johnson, 1989; Hirshman & Bjork, 1988; Slamecka & Graf, 1978). Also, recognition is better for generated response terms than for read response terms (Greenwald & Johnson, 1989; Slamecka & Graf, 1978). These results are, as Hirshman and Bjork have stated, "remarkably robust."

Theoretical accounts of the generation effect have focused on two kinds of processing: item-specific processing of the generated item, and associative processing between

the cue and the generated item. The distinction between item and associative information also forms a cornerstone in current mathematical models of memory (Gillund & Shiffrin, 1984; Hintzman, 1988; Metcalfe-Eich, 1982; Murdock, 1982; Pike, 1984). This point of intersection suggests that the application of these models to the generation effect would be fruitful, both as a test of the models and in furthering our understanding of the generation effect.

Two versions of the search of associative memory (SAM) model (Gillund & Shiffrin, 1984), which make quite different assumptions regarding the representation of associative information, were compared by fitting them to data of two generation-effect experiments. The original version of SAM is referred to here as the *item-level association* (ILA) model because associations are represented as links between separately stored items in memory. The second version, called the *higher order association* (HOA) model, represents associations as higher order units, separate from the items that make up the association. It is similar to a version of SAM introduced by Shiffrin, Murnane, Gronlund, and Roth (1988). The relative success of these models in fitting data will be informative in understanding the generation effect and the relationship between item-specific and associative information.

One might speculate that mathematical models have been infrequently applied to the generation effect because the endeavor is either too trivial or too complex. This seemingly contradictory statement is explained as follows: With few exceptions, most generation-effect experiments have examined only a few critical test conditions. Current mathematical models are computationally powerful and are capable of giving quite unenlightening exact fits to such experiments, and such an exercise would be trivial. However, across experiments, the pattern of results is more complex, with some results occurring under some conditions but not others. In the experiments reported here,

This research was supported by a grant from the Academic Senate of the University of California and by National Science Foundation Grant DBS 9120911. Thanks go to Tony Lin, Julia Lin, and Christine Wells for assistance in collecting data, and to the UCLA CogFog for insightful discussions and inspiration. Correspondence should be addressed to S. E. Clark, Psychology Department, University of California, Riverside, CA 92521 (e-mail: clark@fido.ucr.edu).

memory is tested under several conditions, so that each experiment presents a modeling challenge.

In the present experiments, subjects studied word pairs in which they either read or generated the right-hand response term. Following this list, subjects were tested for recognition and cued recall or were given a free recall test. For all test conditions, performance was evaluated separately for stimulus and response terms. In addition, recall proportions were calculated for items, without respect to pairs, and for recalling both items in a pair (pair recall). A detailed data set from several test conditions should present a good challenge to the SAM model, which should allow us to distinguish between different versions of the model, as well as different accounts of the generation effect. The SAM model is described in the next section, and various accounts of the generation effect are discussed within the SAM framework. Readers who are less interested in the details of SAM may skip to the section Models for Generation Effects or may skip to Experiment 1.

THE SAM MODEL

Words are represented as separate units called *images*. Memory is represented in terms of retrieval strengths between cues used to probe memory and images stored in memory. The cue-to-image retrieval strength between cue *i* and image *j* is denoted $S(Q_i, I_j)$. It is assumed that subjects probe memory with word cues and context cues. Word cues refer to any words used as cues. Context refers to the background of the experiment in which the word is presented. The SAM model conception of context is analogous to that of Anderson and Bower (1973) and Smith, Glenberg, and Bjork (1978). It includes information such as the physical environment and the subject's internal mental state (mood, for example) and physiological state. As Smith et al. have acknowledged, "context" is a rather vague concept. However, it may be logically necessary, since the subject's task is to remember words that were presented in a particular time and place.

In SAM, recall operates as a two-step process in which images are *accessed* and *recovered*. Recognition is based on *global matching*. The details of the recall and recognition models are given below.

Recall

Access to images in memory operates as a stochastic sampling process, in which the image sampled depends on the retrieval strength of the cues and other images in memory. This relative strength sampling rule for sampling image *i* given cues Q_1, Q_2, \dots, Q_M is given as

$$P_s(I_j|Q_1, Q_2, \dots, Q_M) = \frac{\prod_{i=1}^M S(Q_i, I_j)^{w_i}}{\sum_{j=1}^N \prod_{i=1}^M S(Q_i, I_j)^{w_i}}, \quad (1)$$

where w_i is a weight given to cue *i*. The cue weights sum to 1.0 to instantiate the assumption of limited capacity in

using multiple cues at the time of retrieval. The weights play virtually no role in the present work and were fixed so that weight was distributed evenly over multiple cues.¹ The most important aspect of the sampling rule (for present purposes) is that access is based on *relative strength*. An image will be sampled to the extent that it is *uniquely* connected to the cue set. The multiplicative rule for combining cues ensures that sampling is based on the cue set as a whole, rather than on any one of the cues. Thus, a single weak cue, combined with any number of good cues, will reduce the sampling probability substantially.

After an image is sampled, the cues used to sample it are used to recover the information from that image so that it may be output. The recovery rule is given as

$$P_R(I_j|Q_1, Q_2, \dots, Q_M) = 1 - \exp(-\sum w_i S(Q_i, I_j)). \quad (2)$$

Recovery is based on the *absolute* strength between the cue set and the target image (contrary to sampling, which is based on relative strength). After the image is sampled (accessed), other items in memory do not figure into the probability of recovery.

Both free and cued recall operate as a series of sampling and recovery attempts. Free recall begins by probing with context as a single cue until the first item is successfully recovered. The word recalled is then used in conjunction with context to sample the next item. Recall proceeds in this fashion—using the recalled word as a cue along with context to sample the next item until recall ceases or until any particular word cue fails to produce new items. In this second case, the context cue is again used alone.

Three parameters determine when to change cues and when to stop recall altogether. Both are based on an accumulation of *retrieval failures*. A retrieval failure is defined as any of the following: (1) recovery failure of a sampled image, (2) sampling an image that has already been recalled, or (3) resampling of an image that has been previously sampled, but for which recovery failed. If L_{MAX} consecutive failures accumulate for a given word cue, that cue is abandoned and the context cue is used alone. Recall ceases when K_{MAX} total failures accumulate. For cued recall, memory is probed with context and the word cue, weighted equally. These cues are used until the correct item is sampled and recovered or until Q_{MAX} failures accumulate. Except where noted, for all of the simulations to be reported, the values of these parameters were fixed: $K_{MAX} = 30$, $L_{MAX} = 4$, and $Q_{MAX} = 16$.

Recognition

In SAM, recognition decisions are based on a familiarity or strength index, where strength is determined by a global match of the test items to memory. Recognition is based on this global match, rather than on the retrieval of particular images in memory. This match may also be construed as a measure of the familiarity of the test item. The familiarity of cues Q_1, Q_2, \dots, Q_M is given by

$$F(Q_1, Q_2, \dots, Q_M) = \sum_{j=1}^M \prod_{i=1}^M S(Q_i, I_j)^{w_i} \quad (3)$$

which is the denominator of the sampling rule in Equation 2. Familiarity is based on the match of the combined cue set to each of the images in memory (functionally, the items on the list). For recognition of single items, the cue set consists of the test item and context.

There is assumed to be random variability in the system, producing a *distribution* of familiarity for a given test probe type. This produces a signal-detection type of model in which the familiarity distributions for targets and distractors overlap. The variability is instantiated in the model somewhat arbitrarily by assuming a three-point distribution for each cue-to-image strength. Specifically, let x equal $S(Q_i, I_j)$. With random noise in the system, the true value of the cue-to-image strength is x' , where x' equals x , $x(1 + v)$ or $x(1 - v)$, each with equal probability. The value of v is typically fixed at .5.

Parameters

Performance is determined by the strength of cues to images in memory. These strengths, which are determined by rehearsal, preexperimental associations, and type of cue, are given by four parameters. It is assumed that memory is probed with word cues and context cues. For word cues, let b denote an interitem strength between words that are rehearsed together (i.e., words within a study pair), let c be a self-strength between a cue and its corresponding image in memory, and let d be a residual strength between a word cue and the image of a word with which it was not rehearsed (determined primarily by preexperimental association). The context-to-image retrieval strength is denoted a .

Models for Generation Effects

The description of SAM given so far is fairly generic. To apply SAM to the generation effect, certain details must be filled in, and, in doing so, we must commit to some specific assumptions. Earlier, it was noted that two versions of SAM will be fit to data. Several accounts of the generation effect will be explored within the *original*

version of SAM, referred to here as the item-level association (ILA) model, which is described below. Introduction of the higher order association (HOA) model will be deferred until after the ILA model is fit to data.

The Item-Level Association (ILA) Model

Table 1 shows a simple retrieval matrix for two study pairs, A_1B_1 and A_2B_2 , where A_1B_1 is a generate pair (B_1 must be generated), and A_2B_2 is a read pair. The images $A_1, B_1, A_2,$ and B_2 are represented as images on the horizontal axis, and $A_1, B_1, A_2, B_2, X, Y,$ and context are listed as cues on the vertical axis, where X and Y are new items. The cells of the matrix show the parameters for the model to be applied to generation-effect experiments. (The table is structured in the same fashion as the retrieval matrix shown in Raaijmakers and Shiffrin's, 1980, Figure 3.)

Context-to-image strengths may take on three different values: $a(R), a(G)_A,$ and $a(G)_B,$ for read (R) items, and for stimulus (subscript A) and response (subscript B) items in generate (G) pairs. Similarly, for interitem strengths, $b(R)$ denotes the interitem strength between words in a read pair; for generate pairs, $b(G)_{AB}$ is the A -to- B (stimulus-to-response) association strength, and $b(G)_{BA}$ is the B -to- A (response-to-stimulus) association strength. Self-strengths are similarly denoted $c(R), c(G)_A,$ and $c(G)_B.$ It is assumed that the residual strength does not vary with encoding factors, and, thus, d does not vary for read and generate pairs.

It is clear that the full version of this model, with 10 free strength parameters, has a great deal of flexibility. How much of this flexibility is necessary to fit the model to generation-effect data is an important question. A controversy exists regarding how many factors are involved in producing generation effects, with one-factor (Begg, Snider, Foley, & Goddard, 1989; Donaldson & Bass, 1980; Greenwald & Benaji, 1989; McElroy & Slamecka, 1982), two-factor (Hirshman & Bjork, 1988), and three-factor (McDaniel, Waddill, & Einstein, 1988) accounts having been proposed. We can take up this controversy within the framework of the SAM model by examining versions of SAM that may correspond to these different accounts of the generation effect.

Table 1
Retrieval Matrices for Item-Level Association (ILA) and Higher Order Association (HOA) models

	ILA Model				HOA Model						
	Images				Images						
		A_1	B_1	A_2	B_2	Item level				Associative level	
		A_1	B_1	A_2	B_2	A_1	B_1	A_2	B_2	AB_1	AB_2
Cues	A_1	$c(G)_A$	$b(G)_{AB}$	d	d	$c(G)_A$	d	d	d	$b(G)_{A-AB}$	d
	B_1	$b(G)_{BA}$	$c(G)_B$	d	d	d	$c(G)_B$	d	d	$b(G)_{B-AB}$	d
	A_2	d	d	$c(R)$	$b(R)$	d	d	$c(R)$	d	d	$b(R)$
	B_2	d	d	$b(R)$	$c(R)$	d	d	d	$c(R)$	d	$b(R)$
	X	d	d	d	d	d	d	d	d	d	d
Context	Y	d	d	d	d	d	d	d	d	d	d
		$a(G)$	$a(G)$	$a(R)$	$a(R)$	$a(G)$	$a(G)$	$a(R)$	$a(R)$	$a(G)$	$a(G)$

Note— A_1B_1 is a generate pair, A_2B_2 is a read pair. Cues $A_1, B_1, A_2, B_2, X,$ and Y and context are the same for the ILA and HOA models.

Relationship to Current Accounts of the Generation Effect

The framework described above is quite flexible and broad, and several accounts of generation effects can be identified within its domain. Across these accounts, four factors have been identified as contributing to generation effects, including: (1) item-specific, (2) associative, (3) whole-list, and (4) contextual processing. These factors and their relationships within SAM are described below.

Item-specific processing. Response words in A - B pairs show generation facilitation in recognition, and single words presented without additional word cues show facilitation in both recognition (Glisky & Rabinowitz, 1985) and free recall (Gardiner & Hampton, 1985). The fact that word cues are not utilized at retrieval in these cases suggests that the facilitation is due to increases in item-specific processing. In SAM, item-specific processing may be more clearly identified with the self-strength parameter c . However, as the model fits will show later, the self-strengths do not contribute positively to an item's recallability. As a prelude to a discussion that will be expanded later, the kind of item-specific processing that may be operating here may be contextual in nature.

Associative (or relational) processing. The sufficiency of the item-specific account (or any other single-factor account), however, is called into question by results of Hirshman and Bjork (1988), who showed that the pattern of generation effects for both cued and free recall varies with the relatedness of the words in the A - B pair. On the basis of their results, they proposed a two-factor account in which generation of B in A - B pairs involves item-specific facilitation for the B term and also strengthens the A -to- B association. Whether the B -to- A association is strengthened is not specified in the Hirshman-Bjork theory. In their experiments, they looked only at cued and free recall of response terms.

In SAM, item-specific and relational processing most clearly correspond to the self-strength and interitem strengths (the c and b parameters), respectively. However, the b and c parameters do not produce generation effects in the same way that item-specific and relational processing do in the Hirshman and Bjork account of generation effects. In the Hirshman-Bjork account, increases in item-specific processing facilitate cued and free recall, as well as recognition of generated items. Self-strengths contribute to recognition in SAM in the same manner suggested by Hirshman and Bjork. However, for recall, the effect of increasing self-strengths for response items is quite different. As can be seen from Table 1, increasing the self strength for B has no effect on cued recall of B because B is not used as a cue. In SAM, recall probabilities are determined by the cue-to-image strengths of the cues used to probe memory, and B cannot be a cue to recall itself. Note, however, that increases in the self-strength for B will produce interference in recall of A using B as a cue; by strengthening the connection between item B and itself, B will tend to sample its own image more often and will decrease the likelihood of recalling A given B as a cue.

Increases in the A -to- B interitem strength will facilitate recall of B using A as a cue, just as in the Hirshman-Bjork theory. However, in SAM, there is another effect of this increase: Note that the A -to- B interitem strength is part of the sum for the familiarity of the A stimulus term. Thus, increasing $b(G)_{AB}$ not only produces facilitation for recall of B given A as a cue but also facilitates recognition of stimulus terms. That these two components of the generation effect are linked is a unique prediction, and one for which the data are currently equivocal.

Contextual processing. The item-specific/relational two-factor theory, as described by Hirshman and Bjork, can account for a wide range of generation phenomena. However, Greenwald and Johnson (1989) have shown within the SAM model that generation facilitation can be produced by increasing context-to-image connections *without* any changes in item-specific or relational factors. This is certainly true for recall, since the context cue plays an important role in recall. It is unlikely, however, that context manipulations alone will be sufficient to account for the wide range of generation effects. For example, between-list manipulations of context-to-image strengths do not produce changes in recognition performance (see Gillund & Shiffrin, 1984, for details). Increases in context strengths may play a role—however, probably not an exclusive role—in producing generation facilitation.

Whole-list processing. All of the accounts discussed so far assume that the effects of generation are restricted to the word or word pair used in the generation task. However, McDaniel et al. (1988) have shown that the effects of generation extend beyond the word pair. They presented subjects with structured lists of word pairs in which the response words were selected from six different categories (with six words in each category). The generation cue (the A term in the A - B pair) either was a category cue or was uniquely associated with the response term. Generation effects were larger when the generation task utilized category cues associated with several response words than when generation utilized cues uniquely related to only one response item. The version of SAM described above would need to be extended to account for such results. Possible extensions are discussed later.

To test this version of the SAM model, an experiment was conducted that provided data for recognition, cued and free recall. The ILA model gave a reasonable fit to the data overall but failed in certain respects, which motivated the higher order unit model. Since the HOA model was motivated by the results of Experiment 1, the details of the model will be given after presentation of the experiment.

EXPERIMENT 1

The subjects were presented with a single list of 32 word pairs, followed either by cued recall and recognition tests or by free recall. The rationale behind manipulating test conditions in this way is that there is relatively little crosstalk between cued recall and recognition

since the cues are determined by the experimenter. However, this is not the case for free recall, which affects, and is affected by, other test conditions; therefore, free recall subjects were tested separately.

An incidental memory paradigm was used with relatively short presentation times. Long presentation times, which are typically used in generation effect experiments, allow much uncontrolled processing of the words in the read condition. In intentional procedures, the processing certainly goes far beyond reading the words. Long presentation times are in part a consequence of the time it takes to write down responses (and stimulus terms). In Experiment 1, the subjects gave oral responses, allowing a 4-sec/pair presentation rate, rather than the 7- to 14-sec rate previously used.

Watkins and Sechler (1988) have argued that incidental procedures give a "purer" indication of the processes involved in generating words than do intentional procedures in which the comparison condition may involve complex (and unknown) rehearsal strategies. Presumably, with incidental procedures, subjects do little more than read the words when complete words are given, providing a minimal processing baseline with which to compare generation.

Method

Materials. A pool of 48 word pairs was taken primarily (43 pairs) from Palermo and Jenkins (1964) word association norms (5 pairs came from Postman, 1970, norms). The words in each pair were selected so as to be moderately related. For each pair, the response term is usually the fourth or fifth most common response given in the norms, but never the first or second. An example of such a pair is MOUNTAIN-VALLEY, where VALLEY is the fifth most common response (HILL is the most common). A pilot study helped in selecting word pairs for the stimulus set. Word pairs that produced a large number of generation failures were removed from the stimulus set. Also removed were sets of word pairs that were interrelated, resulting in high rates of between-pair cued recall.

Thirty-two of the 48 word pairs were selected randomly for each subject and were presented one pair at a time. For 16 of the pairs, the response word was presented intact; for the other 16 pairs, the vowels of the response term were omitted. The remaining 16 word pairs were used as distractors in the recognition test. For both target and distractor pairs, either the left-hand or the right-hand pair member was tested, but not both. Assignment of words to study and test conditions was randomly determined for each subject.

Procedure. The 32 pairs were presented at a rate of 4 sec/pair. For both generate and read pairs, the subject's task was to read aloud both the stimulus and the response term within 3 sec, after which a beep was presented and the vowels in to-be-generated response terms were filled in, thus presenting the word intact for 1 sec.

Following list presentation, the subjects performed a mental addition task for 2 min. Following this, the subjects were informed regarding the memory test to follow. For cued recall tests, either the left-hand stimulus term (A) or the right-hand response term (B) was presented as a cue, and the subject typed in the appropriate pair member. Cued recall of a response given the stimulus as a cue is denoted $rcl(B|A)$, and cued recall of the stimulus given the response term as a cue is denoted $rcl(A|B)$. Eight of each test type were presented. Of these, half were from generate pairs and half were from read pairs.

For recognition, old stimulus terms (A), response terms (B), or new items corresponding to stimulus (X) and response terms (Y) were presented in random order. A blank line was used to indicate

word position in the pair. The subjects responded "old" or "new" for each test word. Sixteen old and 16 new items were tested. Of the old items, half were stimulus terms and half were response terms. Of these, half were from generate pairs and half were from read pairs.

Recognition and cued recall trials were mixed together in one test phase. A given pair contributed to only one test trial. For example, for the pair DOCTOR-HEALTH, if HEALTH was tested for recognition, neither DOCTOR nor HEALTH could be used on any other test trial. Sixteen recognition and 16 cued recall test trials were given, such that all study pairs were tested.

A second group of subjects performed free recall after list presentation. They were given unlimited time to write down as many words as possible without respect to pairing, but they were asked to indicate paired words if possible.

The assignment of subjects to the recognition/cued recall group or the free recall group was random. Also, the assignment of words to conditions and the order of list presentation and test trials (for the recognition/cued recall group) were determined randomly for each subject.

Apparatus. Stimulus presentation and response collection were controlled by a Zenith 386 computer. Recognition and cued recall responses were collected using a standard keyboard. Free recall responses were collected on paper.

Subjects. Seventy-one introductory psychology students participated in partial fulfillment of a course requirement, 37 in the recognition/cued recall group and 34 in the free recall group. The subjects were run individually for about 20 min each.

Results

Generation performance at study was quite good with 94% correct generation across all subjects (with missing data for 7 subjects due to an equipment failure). Three separate 2×2 analyses of variance (ANOVAs) were computed for recognition, cued recall, and free recall, with pair position (stimulus or response) and encoding condition (generate or read) as within-subject factors. Details of the statistical analysis are given in Appendix A. Recall proportions and recognition d' 's are shown in Table 2 for stimulus and response terms from read and generate pairs. Generation facilitation was shown for both stimulus and response terms for recognition, cued recall, and free recall. In all cases, the amount of facilitation was larger for responses than for stimulus terms, shown by an interaction of encoding condition and pair member (however, the interaction was not significant for cued recall).

The recall results are straightforward and largely consistent with past results. The picture for recognition is more complicated. While several studies have shown no facilitation for recognition of stimulus words (Greenwald & Johnson, 1989, Experiment 1; McDaniel & Waddill, 1990; Slamecka & Graf, 1978), Greenwald and Johnson's Experiment 2 results and the present results do show facilitation.

A means by which to reconcile these results is not obvious, since the experiments varied in many respects (Greenwald and Johnson do not attempt to reconcile the inconsistent recognition results of their Experiments 1 and 2). However, a few points are worth noting: Greenwald and Johnson obtained facilitation for stimulus terms when recognition was preceded by cued recall; McDaniel and Waddill's subjects were tested only with recognition. One might conclude that the facilitation shown by Greenwald and Johnson was an artifact of sequential

Table 2
Results and Fit of the SAM ILA Model
and HOA Model to Experiment 1

		Data	ILA Model	HOA Model
Recognition				
Read	(<i>A</i> vs. <i>X</i>)	1.338	1.274	1.232
Read	(<i>B</i> vs. <i>Y</i>)	1.243	1.274	1.232
Generate	(<i>A</i> vs. <i>X</i>)	1.979	2.206	2.006
Generate	(<i>B</i> vs. <i>Y</i>)	2.407	2.391	2.306
Cued Recall				
Read	(<i>A B</i>)	0.339	0.318	0.343
Read	(<i>B A</i>)	0.331	0.320	0.343
Generate	(<i>A B</i>)	0.500	0.447	0.549
Generate	(<i>B A</i>)	0.590	0.551	0.606
Free Recall				
Read	<i>A</i>	0.068	0.078	0.068
Read	<i>B</i>	0.070	0.078	0.068
Read	<i>A</i> and <i>B</i>	0.064	0.016	0.068
Read	<i>A</i> or <i>B</i>	0.077	0.140	0.068
Generate	<i>A</i>	0.202	0.249	0.249
Generate	<i>B</i>	0.240	0.275	0.249
Generate	<i>A</i> and <i>B</i>	0.203	0.113	0.249
Generate	<i>A</i> or <i>B</i>	0.245	0.416	0.249

Note—*A* and *B* denote stimulus and response list words; *X* and *Y* denote corresponding nonlist items. Recognition data are given as d' . (*A|B*) and (*B|A*) denote recall of *A* given *B* as a cue and recall of *B* given *A* as a cue.

testing. The present experiment, however, did not use sequential testing. Different test pairs were used in a mixed testing procedure, and it seems quite unlikely that the results were produced by an accumulation of effects over sequential testing.

In addition, our results were obtained using a presentation rate much faster than that used in previous experiments. Due to the short presentation times, subjects in the read condition may do very little but read the two words, thus providing a performance baseline against which generation facilitation appears quite large. Many previous studies have used 7- to 8-sec presentation rates. For read pairs, in particular, it seems likely that subjects would engage in a variety of unknown and uncontrolled behaviors in the 5 or 6 sec that the pair is on the screen after they have finished the reading task.

Application of SAM to Experiment 1

Recognition predictions can be obtained by deriving the means and variances of familiarity distributions given by the strength parameters. Recognition d' for each test condition was calculated in the standard way as the difference between target and distractor means, divided by the standard deviation of the distractor distribution. However, closed-form expressions cannot be obtained for cued or free recall, and predictions can only be obtained by Monte Carlo simulation. The final fits of the model are based on 2,000 simulations. It is not feasible to run 2,000 simulations for all of the exploration of the parameter space; therefore, exploratory simulation runs were based on 200 simulations each.

The need to obtain predictions by simulation also precludes the possibility of an exhaustive search of the pa-

parameter space, or a search driven by computer algorithm (an attempt at the latter consumed 48 h of CPU time on a VAX 8820 with no end in sight). Therefore, predictions were obtained by trying different combinations of parameters, studying the pattern of mispredictions and making adjustments based on properties of the model. This procedure was guided by calculating the sum of the squared errors (*SSE*) comparing model predictions to data (these error values are listed with the parameter values in Table 3). This goodness-of-fit measure must be interpreted with caution, however, since these cannot be viewed as "best" fits. Because of this, the goodness-of-fit measures alone do not provide a sufficient basis for drawing absolute conclusions about the model. Rather, the fits are instructive in revealing properties of the model, and any conclusions must be grounded on arguments concerning the properties of the model.

Two constrained versions of the model were examined, one corresponding to the Greenwald and Johnson (1988) context-only model and the second corresponding to the two-factor account of Hirshman and Bjork, which is based on item-specific and associative information. Both of these models failed, and the full model, which is based on contextual, item-specific, and associative processing.

Context-only model. Although between-list variation in context strength does not affect recognition performance, within-list variation does, so it is reasonable to see how well the context-only model fares in fitting these results. The context-only model was instantiated in SAM by allowing only the context parameters $a(R)$, $a(G)_A$, and $a(G)_B$ to vary. Increasing $a(G)_A$ increased performance for all stimulus terms in generate pairs for all test conditions, and increasing $a(G)_B$ increased performance for response terms in all conditions. Obviously, $a(G)_B > a(G)_A$. However, the value of $a(G)_A$ necessary to produce the magnitude of free recall facilitation for stimulus words produced far too much facilitation for recognition. Also, the assumption that the interitem strengths do not vary underpredicted the amount of facilitation for cued recall. Beyond this, there are numerous demonstrations of generation facilitation in recognition using between-list designs (Greenwald & Johnson, 1988; Slamecka & Graf, 1978); therefore, this model was not considered further.

Hirshman-Bjork theory. The Hirshman-Bjork theory proposes that generation effects are due to strengthening of cue-response associative information and item-specific information for the item that is generated. These assumptions are incorporated into SAM by increasing $b(G)_{AB}$ and $c(G)_B$. This model will produce facilitation for recognition and cued recall of generated items but will not produce facilitation for stimulus terms. Hirshman and Bjork did not examine recall of stimulus terms, so they do not specify whether generation of *B* increases *B*-to-*A* associations or increases item-specific processing for stimulus terms. SAM can produce the cued recall facilitation of stimulus terms by increasing the *B*-to-*A* association [$b(G)_{BA} > b(R)$], which is at most only a minor extension of the Hirshman-Bjork two-factor theory.

It is not clear whether the recognition facilitation for stimulus terms requires increases in item-specific information [represented by an increase in $c(G)_A$ in SAM]; as noted earlier, recognition facilitation is produced as a by-product of increasing the associative strengths. This model is similar to the full SAM model described earlier, but without variation of the context parameters [i.e., $a(R) = a(G)_A = a(G)_B$].

This version of SAM failed because variation in the interitem strengths cannot produce the large generation effect shown in free recall. Increasing the *A*-to-*B* connection will increase the likelihood of recalling *B* using *A* as a cue. However, at that point, recall is at a dead end. The recalled *B* item is used as a cue, and the image that is most likely to be sampled is the *A* item that was used as a cue to recall *B* (which, of course, has already been recalled and is thus a retrieval failure). Because of the pair structure of the list, increasing an interitem connection gains access to only one additional list item. This, of course, will produce substantial facilitation in cued recall, but very little in free recall.

This aspect of the model is consistent with previous experimental results and with the Hirshman-Bjork theory. Hayes-Roth (1977) and Yekovich and Manelis (1980) have shown that as the association between items increases, free recall of those items *decreases*. Consistent with these results, Hirshman and Bjork assume that free recall generation facilitation is *not* due to interitem connections. However, this is where the similarity between their approach and SAM ends on this issue. Hirshman and Bjork assume that free recall is based on item-specific processing, whereas in SAM increases in self-strengths produce interference for free recall (due to self-sampling).

If items are presented singly, rather than in pairs, such that a richer network of associations can be formed, then increasing interitem connections will produce free recall facilitation (Gillund & Shiffrin, 1984). However, it is clear from our simulations that in a list of *pairs*, increasing interitem strengths will not produce the degree of free recall facilitation shown in the data.

Full three-factor model. The high degree of free recall facilitation shown in the data cannot be produced by self-strengths or interitem strengths alone. The only option within the SAM model is to allow context-to-item connections to vary in addition to self- and interitem strengths. The fit of this *three*-factor model is shown in the second column of Table 2, and the parameters used to obtain that fit are listed in the first column of Table 3.

The model was able to produce the overall pattern shown in the data: facilitation for all three testing procedures, for both stimulus and response terms, with greater facilitation for response terms. Free recall performance is slightly overpredicted, cued recall performance is slightly underpredicted, and recognition of cues in generate pairs is slightly overpredicted. The reason for these errors stems from the fact that the level of cued recall performance shown in the data was quite high. In SAM, cued recall performance increases with interitem and context strengths, and both were set quite high. However, both

parameters are constrained by the recognition and free recall results. High context strengths pushed free recall performance slightly too high, and high interitem strengths pushed recognition of stimulus terms too high. To keep recognition of stimulus terms from generate pairs within the ballpark of the data, the self-strength $c(G)_A$ had to be set very low. One interpretation of this low self-strength is that generation of response terms may draw item-specific processing away from stimulus terms.

Closer examination of the free recall predictions reveals the major failure of the model. Probabilities for recalling both items (pair recall of *A* and *B*) or either item (*A* or *B*) in a pair are listed in Table 2. The data show that pair recall was quite close to the unconditional item recall performance. This aspect of the results was not captured by SAM. Table 2 shows the probabilities of recalling both items in a pair as .203 and .064 for generate and read pairs. However, SAM produced probabilities of .113 and .016. Similarly, the data show the probability of recalling either item from a pair (*A* or *B*) to be .24 and .07 for generate and read pairs, as opposed to .416 and .140 for SAM.

The data show that items within a related pair are strongly connected such that access to one item results in access to both. Within the basic version of the SAM model, this high pair recall would be produced by very high interitem strengths. Thus, if item *A* is recalled, it can be used as a cue to recall *B* with high probability. In order for this approach to work, the probability of recalling *B* given *A* must be near 1.0, which is not shown in the cued recall data. Also, the value of *b* necessary to produce this high recall probability overpredicts the level of recognition performance. One means of producing the dependencies shown in the data is to assume that the items within a pair are stored as a single higher order unit. In the next section, this model is described and fit to data.²

A Higher Order Association Model

The SAM model was modified to represent the association between items in a pair as a single unit. This model is referred to as the higher order association (HOA) model. The HOA model has two separate levels of rep-

Table 3
Parameter Values for SAM Model Fits to Experiments 1 and 2

Parameter	ILA Model		HOA Model	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
$a(R)$	0.380	0.900	0.130	0.210
$a(G)_A$	0.760	1.600	0.220	0.300
$a(G)_B$	0.840	1.900	0.260	0.300
$b(R)$	0.880	0.560	0.380	0.350
$b(G)_{AB}$	1.200	0.900	0.500	0.520
$b(G)_{BA}$	0.900	0.850	0.440	0.440
$c(R)$	0.040	0.050	0.550	0.440
$c(G)_A$	0.026	0.026	0.800	0.600
$c(G)_B$	0.180	0.120	0.880	0.750
d	0.026	0.026	0.018	0.180
SSE	0.108	0.139	0.032	0.056

Note— SSE is the sum of squares error, comparing model predictions with data.

resentation, an *item* level and an *associative* level, and, therefore, memory is based on two separate retrieval matrices (as shown in Table 1). The item level corresponds closely to the retrieval matrices shown for the ILA model in Table 1. Each item is stored as a separate image. Self-strengths representing an item's connection to its own image in memory are shown on the diagonal, and all other cells are residuals. There are no interitem associations at the item level. Instead, associations are stored separately as higher order associative units.

Recognition

Recognition operates in the same manner as that in the ILA model. Item familiarity is given by summing cue-to-image strengths over item-level images to obtain a familiarity value. However, because associative information is stored only at the associative level in the HOA model, the associations play no role in recognition.

Recall

Cued and free recall both operate exclusively within the associative level of the system. Recall is not of single images, but rather of associations. Thus, items are always recalled as pairs. The associative matrix shows retrieval strengths between item cues and association images. The strength between an item and its corresponding association takes on an associative strength, and all other strengths are residuals. The main difference in the model is the level at which associations are represented. Cued and free recall operate just as before, as a sampling and recovery process, except that pair-level associations are sampled and recovered, rather than items within pairs.³ For cued recall, the item cue is used with context to sample the association image for the pair. If the association image is sampled, the same cues (context, item cue) are used to recover the image.

For free recall, the recall sequence begins, *and ends*, using context as the only retrieval cue. It doesn't make sense to probe with word cues: word cues can be used only for words that have already been recalled; since the pairs are stored as units, if one word from a pair has been recalled, both have been recalled. Thus, a word cue can only be used to recall a word in a different pair, which presumably should not be effective because the list was structured to minimize between-pair recall. Therefore, the only sensible cues for probing memory are context cues. Sampling and recovery work just as they do in the ILA model, except that the basic unit is the pair, not items within a pair.

Parameters

The parameters are nearly the same as for the ILA model, except for context, which operates at both the item and the associative level. At the item level, context strengths are exactly as before: $a(R)$, $a(G)_A$, and $a(G)_B$ for read items, and stimuli and responses in generate pairs. Context strengths at the associative level take on two strengths, where $a(R)_{AB}$ is the context-to-association strength for read pairs, and $a(G)_{AB}$ is the context-to-

association strength for generate pairs. Rather than have separate free parameters for the context strengths at the associative level, $a(R)_{AB}$ and $a(G)_{AB}$ were fixed as follows: $a(R)_{AB} = a(R)$, and $a(G)_{AB} = \max[a(G)_A, a(G)_B]$.

Other parameters followed closely those of the ILA model. Interitem strengths can take on three values, $b(R)$, $b(G)_{A-AB}$, or $b(G)_{B-AB}$, where the latter two parameters refer to the retrieval strength between *A* and the *AB* unit and between *B* and the *AB* unit, respectively. Self-strengths take on values $c(R)$, $c(G)_A$, and $c(G)_B$ just as before (but only at the item level). One residual strength, d , is used at both the item and the associative level.

Fit of the Higher Order Association Model

The fit of the HOA model is shown in the third column of Table 2. The fit of the model is better than that of the ILA model, especially with respect to free recall. Since pairs are recalled as single units, free recall probabilities for recalling either or both items are exactly the same, which is not far from the data.

The parameters used in obtaining the fit are shown in Table 3. Because free recall is context-driven in the model, the context strengths showed large increases for generate pairs relative to read pairs. Cued recall is primarily dependent on associative strengths, and the parameter values are consistent with the pattern shown in the data. An important aspect of the HOA model is shown in the fit to the recognition data. Because item- and associative-level information are separate, recognition performance is not affected by associative strengths, but only by self-strengths. Consequently, the self-strengths are much higher in the HOA model than in the ILA model, and, moreover, the self-strength for stimulus terms in generate pairs does not decrease in order to compensate for the facilitation produced by increasing interitem strengths.

Summary of the Modeling

Two models were fit to the data of Experiment 1. The first was an extension of the original SAM model. In this (ILA) model, episodic associations are represented as connections or links between items that are stored separately in memory. A second version of the model (HOA), which represents associations as higher order units, was also fit to the data. The fit of the HOA model was better than that of the ILA model, primarily because it accounts for the dependency in recall of items within a pair. The subjects tended to recall either both items in a pair or neither item in a pair.

Beyond the improved fit, there are important conceptual advantages of the HOA model. By separating the associative- and item-level information, the model no longer binds recall and recognition together. In the ILA model, *A*-to-*B* association strengths increased recall probabilities for recalling *B* using *A* as a cue and also increased the familiarity of *A*, thus tying cued recall and recognition performance together. The ILA model made the strong prediction that facilitation in cued recall of responses must be accompanied by facilitation in recognition of stimulus terms. This result was shown. However,

in order to produce the level of cued recall facilitation shown in the data, the *A-to-B* associative strength had to be set so high that it overpredicted the recognition facilitation for stimulus terms. In order to offset the facilitation produced by the interitem strength increase, the self-strength for the stimulus had to be decreased. In short, the ILA model was forced into some contortions in order to produce the pattern of results shown in the data.

The HOA model decouples item-specific and associative information, so that the recognition and cued recall predictions are not linked, and the model is not forced into such contortions. Perhaps more important is that the HOA model provides a framework that is more consistent with the Hirshman–Bjork account of generation effects. From a different angle, one might conclude that the Hirshman–Bjork account is not easily instantiated within a model in which item-specific and associative information are linked together; rather the Hirshman–Bjork account assumes some degree of independence between an association and the items that enter into that association. This is true for SAM, and it is likely to be true for other models as well. In the Hirshman–Bjork account, it seems to be implicitly assumed that associative information plays no role in item recognition. It seems necessary to also assume that item-specific and associative information can be retrieved independently. The HOA model provides an architecture to instantiate this assumption.

One difference between the HOA–SAM model and the Hirshman–Bjork account is that the Hirshman–Bjork account describes two factors underlying generation effects—namely, item-specific and associative processing—and SAM adds an additional factor: increased context-to-image strengths. Context processing is a critical factor in SAM for both the ILA and the HOA models. This raises the question of what one means when one speaks of “item-specific processing.” A very constrained definition of item-specific processing may involve processing of lexical information, such as meaning and phonology, which is logically independent of the context in which the item occurs. On the other hand, in SAM, context strengths represent contextual information that is stored with a specific image in memory. In a broader sense, this contextual processing can also be thought of as item-specific in that it connects a specific item to a specific context. By this broader view of context, the HOA version of SAM is not inconsistent with the Hirshman–Bjork account of generation effects.

Context is a central component of the SAM model account of generation effects. Generation effects for free recall cannot be produced in SAM unless context strengths are increased for generate items, regardless of whether one assumes image-level or higher order associations. Thus, experimental conditions in which contextual processing is likely to be equivalent should selectively eliminate the generation effect for free recall.

Previous experiments have shown that generation effects for free recall are sometimes absent or even negative (Hirshman & Bjork, 1988; Schmidt & Cherry, 1989; Slamecka & Kasaiti, 1987). Many, though not all, of these

failures to show generation facilitation have manipulated encoding conditions between subjects or have used intentional learning procedures, or both. Both of these factors were incorporated into Experiment 2. These manipulations may selectively affect contextual processing. In the intentional learning procedure, subjects are explicitly told that they must remember that the items occurred in the context of the experiment. Knowledge that both generate and read pairs must be remembered should tend to increase and equalize contextual processing. The manipulation of encoding condition between lists may also serve to equate contextual processing, because, within the context of a given list, all of the items are of equal salience, whereas in a mixed list, subjects may attend to generate pairs at the expense of read pairs (see Slamecka & Kasaiti, 1987).

In SAM, this pattern of results would be produced by higher context strengths overall and a much smaller difference in context strengths for read and generate pairs (since subjects know they need to remember both). The prediction for Experiment 2 was that the generation effect would be selectively eliminated for free recall but would remain large for both recognition and cued recall.

EXPERIMENT 2

Method

Materials and Procedure. The materials and the testing procedure were identical to those used in Experiment 1, but with two changes. First, at the outset of the experiment, the subjects were informed that there would be a memory test following presentation of the list. The details of the test were not disclosed until after list presentation. The second difference was that the encoding task was varied between, rather than within, subjects. Thus, a given subject either generated each response for the 32 pairs or read all of the 32 pairs. As in Experiment 1, the subjects in both groups read both words in each pair aloud. These responses were recorded on audiotape.

Subjects. The subjects were 96 different introductory psychology students from the University of California, Riverside, selected from the same pool as used for Experiment 1.

Design. Encoding task (generate or read) and test group (recognition/cued recall or free recall) were manipulated between subjects. The factorial combination of these two variables produced four groups, with 24 subjects per group. Testing of stimulus versus response words was examined as a within-subjects variable.

Results

The generation rate data could not be analyzed due to an equipment malfunction. However, it seems reasonable to assume that the generation error rate was in the same range as in Experiment 1. Recognition, cued recall, and free recall results are given in Table 4. Three 2×2 (encoding condition \times pair member) ANOVAs were performed, one for each type of test. The details of the statistical tests are given in Appendix B. These analyses showed a significant generation effect for recognition of response terms but not for stimulus terms, a significant generation effect for cued recall of stimuli and responses, and a small, nonsignificant generation effect for free recall. Additional analyses showed a significant free recall advantage for responses, but not for stimulus terms.

Experiment 2 contrasts with Experiment 1 in that encoding task (generate vs. read) was manipulated between subjects, and the subjects were informed that a memory test would follow list presentation. It was predicted that under these conditions, generation facilitation should be reduced, particularly for free recall (Schmidt & Cherry, 1989; Slamecka & Kasaiti, 1987). Precisely this pattern of results was shown in Experiment 2. Comparison of Tables 2 and 4 shows that performance for Generate subjects was remarkably close to the results shown for the generate subjects of Experiment 1. Not unexpectedly, however, when the subjects knew that they would be tested for memory of the list, their performance increased for the read group (relative to the performance of the read group of Experiment 1). The largest increases were for free recall, as well as for recognition of stimulus terms. Recognition and free recall of stimulus terms, which showed large generation facilitation in Experiment 1, showed very small statistically unreliable facilitation in Experiment 2.

Thus, the changes in procedures in Experiment 2 had almost no effect on the performance for words in generate pairs, and the generation effect was selectively eliminated for free recall (of stimulus terms in particular) and for recognition of stimulus terms. One goal of Experiment 2 was to minimize generation facilitation for free recall, while maintaining facilitation for other conditions. This goal was largely achieved, although generation effects were not entirely eliminated for free recall. How this factor plays into the fit to Experiment 2 is described next.

Fit of SAM to Experiment 2

ILA Model. The fit of the ILA model is shown in column 2 of Table 4, and the parameters are listed in Table 3.

In Experiment 1, the fit of the model was driven in large part by variations in the context strengths. The purpose of Experiment 2 was to selectively eliminate the generation effect for free recall. Within the SAM framework, this should require that context strengths not differ very much between generate and read conditions. Also, the fact that the subjects knew about the memory test suggests that they would have made a more concerted effort not only to perform the read or generate task but also to engage in rehearsal processes that would connect the list words to the context of the experiment. Thus, the context strengths should have been higher overall for Experiment 2 than for Experiment 1. Both of these predictions were borne out in the fit.

The fit is relatively good, with the exception of the pair recall probabilities for free recall. This shortcoming of the fit is parameter-independent. The ILA model cannot produce this pattern of results with any combination of parameter values. Thus, the modified model (HOA model), which represents word pairs as higher order units, was fit to the data.

Higher Order Association (HOA) Model. The fit of the HOA model is shown in column 3 of Table 4, and the parameters are listed in Table 3. The overall fit of the HOA model is better than that of the ILA model, primarily by virtue of fitting the pair recall probabilities for free recall. Again, the HOA model predicts pair recall to equal item recall, which is close to the data. Also, as in the ILA model, the context parameters for the HOA model were higher for the fit to Experiment 2 than for the fit to Experiment 1, consistent with the predictions.

GENERAL DISCUSSION

The generation effect was examined within the framework of a well-specified model—namely, the SAM model of Gillund and Shiffrin (1984). This exploration revealed important aspects about SAM and has important implications for accounts of the generation effect.

SAM was fit to two experiments in which subjects were presented with related word pairs. Memory was tested with recognition, cued recall, and free recall. The test conditions were designed so as to minimize crosstalk between tests. The results were straightforward: Experiment 1 used an incidental learning procedure and manipulated encoding task within list. The results showed generation facilitation for all three test conditions, for both stimulus and response terms, with greater facilitation for generated responses than for the stimulus terms used in the generation task. In Experiment 2, the subjects were informed of the memory test following list presentation, and encoding task was manipulated between subjects. The pattern of generation effects was changed very little relative to Experiment 1, with the exception that the generation effect was greatly reduced for free recall (particularly for stimulus terms) and was eliminated for recognition of stimulus terms.

In both experiments, free recall was analyzed in terms of words recalled and in terms of pair recall, in which both

Table 4
Results and Fit of SAM Model to Experiment 2

		Data	ILA Model	HOA Model
Recognition				
Read	(<i>A</i> vs. <i>X</i>)	1.751	1.623	1.589
Read	(<i>B</i> vs. <i>Y</i>)	1.429	1.623	1.589
Generate	(<i>A</i> vs. <i>X</i>)	1.941	2.050	1.923
Generate	(<i>B</i> vs. <i>Y</i>)	2.187	2.299	2.198
Cued Recall				
Read	(<i>A B</i>)	0.40	0.360	0.402
Read	(<i>B A</i>)	0.38	0.359	0.409
Generate	(<i>A B</i>)	0.53	0.532	0.521
Generate	(<i>B A</i>)	0.60	0.598	0.580
Free Recall				
Read	<i>A</i>	0.171	0.136	0.157
Read	<i>B</i>	0.168	0.135	0.157
Read	<i>A</i> and <i>B</i>	0.147	0.019	0.157
Read	<i>A</i> or <i>B</i>	0.191	0.252	0.157
Generate	<i>A</i>	0.206	0.157	0.224
Generate	<i>B</i>	0.241	0.185	0.224
Generate	<i>A</i> and <i>B</i>	0.195	0.028	0.224
Generate	<i>A</i> or <i>B</i>	0.251	0.314	0.224

Note—*A* and *B* denote stimulus and response list words; *X* and *Y* denote corresponding nonlist items. (*A|B*) and (*B|A*) denote recall of *A* given *B* as a cue and recall of *B* given *A* as a cue.

words in a pair were recalled. In both experiments, for generated and read items, pair recall was almost identical to item recall. If one word from a pair was recalled, it was likely that both words would be recalled. This result was important in testing between variations of models within the SAM framework.

Representation of Associative Information

Two variants of SAM were fit to the data of these two experiments. In the ILA model, associations are represented at the item level, as connections between cues and words in memory. In the HOA model, an association is represented as a single unit, separate from the items that enter into the association. The ILA model was not able to produce the pattern of item and pair recall shown in the data. The subjects tended to recall pairs as pairs. The ILA model produces both items in a pair by first recalling one and then using the recalled item to recall the next. Thus, recall of the second item in the pair operates like cued recall. Recall of the second item in a pair was near unity, and cued recall varied between .30 and .60; therefore, the ILA model could not fit the pair recall data.

Relationship Between Associative and Item-Specific Information

The relationship between item-specific and associative information depends upon the representation of associations. In the HOA model, associations are stored as separate units from the items, allowing independent access to item and associative information. Thus, recognition of individual words is based on item-specific information and is not affected by any associations involving that word. This contrasts with the ILA model in which items and associations are represented at the same level within the system. In the ILA model, it was necessary to assume that item-specific information for stimulus terms decreased in order to balance the facilitation produced as a by-product of strengthening the *A-to-B* association. The HOA model does not require this tradeoff.

Contextual Facilitation

The fit of SAM to generation effect data draws on the full power of the model, suggesting that generation effects are produced by changes in item-specific, interitem, and contextual information. The proposal that item-specific and interitem information are involved in generation effects is not new; however, theoretical analyses with SAM indicate that free recall facilitation requires that contextual information is also increased for generate pairs. In Experiment 1, generation facilitation in free recall was quite large. Items from generate pairs were recalled roughly three times as often as items from read pairs. Varying self-strengths and interitem strengths could not produce this magnitude of facilitation. This was true for both the ILA and the HOA versions of the model.

Other Models

One might question whether this dependency on context as a retrieval cue is peculiar to the SAM model. This

dependency in SAM became apparent only after extensive theoretical work involving thousands of simulations. To properly address this question for other models will require the same degree of theoretical rigor. However, we can approach the question in a preliminary fashion for a class of models similar to SAM called *global memory models*. These models include convolution-association distributed memory models, such as TODAM (Murdoch, 1982) and CHARM (Metcalf-Eich, 1982), Pike's (1984) Matrix Model, and Hintzman's (1988) multiple-trace model.

One might rephrase the question: In these models, what is the effect of increasing item-specific and associative information on free recall of items studied as pairs? Retrieval in all of these models is completely cue-dependent. Thus, they are similar to SAM in that item-specific information for the item to be recalled does not play a role in gaining access to that item. In SAM, the recovery process is driven by the same cues used to gain access to (sample) the item. So, for SAM, the item-specific self-strength does not contribute to postaccess recovery processes. In the other global memory models, the accessed information is errorful, and this noise must be filtered prior to output. Additional item-specific information in the trace may be helpful in this postaccess sharpening of the output. It remains to be seen whether this source of facilitation would be sufficient to produce the large increase in performance shown in the data.

The associative information in these models is likely to produce results similar to results produced by SAM. Associative information will allow recall of one item in a pair using the other item as a cue. However, this is not the source of facilitation in free recall. Subjects gain access to more pairs. The question then is, what kind of cue is likely to gain access to pairs? It is unlikely that items from other pairs will be useful. This would require strong between-pair associations, which would lead to poor cued recall performance. Again, a likely candidate for this cue is a contextual cue.

Relationship Between Item-Specific and Contextual Processing

In SAM, context-to-image strengths take on the role of item-specific processing in the Hirshman-Bjork theory. This raises the question of what is meant by "item-specific." Contextual processing may be considered to be item-specific in the sense that it does not involve other items. It may be useful to think of two kinds of item-specific processing: one that connects a particular item to context, and one that involves processing of an item's features independent of context and other items. An example of the latter would include the spelling or sound of the item.

Multifactor Accounts of the Generation Effect

Both versions of SAM that were applied to the data bear a strong similarity to Hirshman and Bjork's (1988) two-factor account of the generation effect. However, the HOA model provided a better framework for instan-

tiating the assumptions of the Hirshman–Bjork model than did the ILA model. Specifically, the HOA model provides a framework for keeping item-specific and associative information separate at retrieval.

Extensions and Limitations

Associative strength. The Hirshman–Bjork theory was proposed in response to their results showing that preexperimental associative strength produced different patterns of generation effects for cued and free recall. Some of their word pairs were first associates, and others were third associates. They showed that free recall performance was lower for first associates than for third associates, and that this disadvantage for first associates increased as a result of the generation task. This would suggest that free recall performance decreased as the associative strength of items within pairs increased. A quite different pattern was shown for cued recall. Not unexpectedly, cued recall performance increased with the associative strength between items in pairs. Hirshman and Bjork argue that the strengthening of within-pair associations *inhibits* free recall.

How might SAM account for these results? In the ILA model, strengthening associations has a very small positive effect on free recall; in the HOA model, associative strength has no effect on free recall. Thus, in SAM, variation in the associative strength itself cannot produce the deficit. Episodic increases in associative strength may involve storing less of other kinds of information (e.g., contextual information). However, it is less clear how, within the SAM framework, to account for the disadvantage in free recall due to preexperimental associative strength. One possibility suggested by Hirshman (1988) is that weaker associations may receive more processing at study than stronger associations.

Whole-list factors. The three-factor account of the generation effect within the SAM framework assumes that the facilitation is produced by within-pair factors only. However, McDaniel et al. (1988) have shown that generation effects extend beyond the pair when categorically structured lists are used. In all of the applications here, it was assumed that residual strengths (the d parameter in SAM) did not vary. To account for these whole-list results, it seems clear that this constraint on the model would need to be relaxed. Thus, generation effects would be produced for items outside the pair that are related to the cue used in the generation task.

Concluding Remarks

The goal of this research was to apply a well-specified mathematical model to an important result in the memory literature, the generation effect. These theoretical analyses indicate that the generation effects shown after study of related word pairs are produced by increases in three kinds of information: item-specific, associative, and contextual. Furthermore, these analyses indicate that the word pairs are represented as higher order associative units, separate from the components that enter into that association. This representation is required so as to allow

independent access to item-specific and associative information. While these conclusions are drawn specifically within the framework of the SAM model, the results will likely hold for other models as well.

These theoretical analyses demonstrate the importance of the generation effect for models of memory. What is important is not that generation produces facilitation but rather that generation produces a *particular* pattern of facilitation across testing conditions. Further investigation of generation effects in paradigms that allow examination of the effect for different retrieval tasks should lead to additional insights for models of memory.

REFERENCES

- ANDERSON, J. R., & BOWER, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
- BEGG, I., SNIDER, A., FOLEY, F., & GODDARD, R. (1989). The generation effect is no artifact: Generating makes words more distinctive. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 977-989.
- CLARK, S. E., & SHIFFRIN, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition*, *15*, 367-378.
- DONALDSON, W., & BASS, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning & Verbal Behavior*, *19*, 26-35.
- 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.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- GLISKY, E., & RABINOWITZ, J. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 193-205.
- GREENWALD, A. G., & BENAJI, M. R. (1989). The self as a memory system: Powerful, but ordinary. *Journal of Personality & Social Psychology*, *57*, 41-54.
- GREENWALD, A. G., & JOHNSON, M. M. S. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, *17*, 673-681.
- HAYES-ROTH, B. (1977). Evolution of cognitive structures and processes. *Psychological Review*, *84*, 260-278.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.
- HIRSHMAN, E. (1988). The expectation-violation effect: Paradoxical effects of semantic relatedness. *Journal of Memory & Language*, *27*, 40-58.
- 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.
- MCDANIEL, M. A., & WADDILL, P. J. (1990). Generation effects for context words: Implications for item-specific and multifactor 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.
- MCLEROY, 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.
- METCALFE-EICH, J. (1982). A composite holographic associative recall model. *Psychological Review*, *89*, 157-173.
- MURDOCK, B. B., JR. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626.
- MURDOCK, B. B., JR., & OGILVIE, J. C. (1968). Binomial variability in short-term memory. *Psychological Bulletin*, *70*, 256-260.
- PALERMO, D. S., & JENKINS, J. J. (1964). *Word association norms: Grade school through college*. Minneapolis: University of Minnesota Press.

- PIKE, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review*, **91**, 281-294.
- POSTMAN, L. (1970). The California norms: Association as a function of word frequency. In L. Postman & G. Kepele (Eds.) *Norms of word association* (pp. 241-320). New York: Academic Press.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207-262). New York: Academic Press.
- SCHMIDT, S. R., & CHERRY, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition*, **17**, 359-369.
- SHIFFRIN, R. M., MURNANE, K., GRONLUND, S. D., & ROTH, M. (1988). On units of storage and retrieval. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane-Flower Symposium on Cognition* (pp. 25-68). Hillsdale, NJ: Erlbaum.
- 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., & KASAITI, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, **26**, 589-607.
- SMITH, S. M., GLENBERG, A. M., & BJORK, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, **6**, 342-353.
- WATKINS, M. J., & SECHLER, E. S. (1988). Generation effect with an incidental memorization procedure. *Journal of Memory & Language*, **27**, 537-544.
- YEKOVICH, F. R., & MANELIS, L. (1980). Accessing integrated and non-integrated propositional structures in memory. *Memory & Cognition*, **8**, 133-140.

NOTES

1. The weights were fixed in similar fashion to previous applications (Clark & Shiffrin, 1987). When context is used alone as a cue, it is given a weight of 1.0. It is assumed that when context is used in conjunction with any number of other cues, the context cue is given a weight of .5, and the remaining weight is divided among the other cues. In the present application, at most one other (word) cue is used, so the word cue and the context cue were both fixed at .5.

2. The second problem could be minimized by assuming that some interitem strengths and context strengths are correlated. Thus, some interitem strengths are very high and some very low, producing a moderate level of cued recall. Since high context and interitem strengths occur together, the items that are accessed in free recall will tend to have high interitem strengths, thus producing a high level of dependency.

3. It is assumed here that recovery is also based at the pair level. An alternative assumption would be that pairs are sampled, but the items are "unpacked" as items. This assumption would produce high levels of pair recall but would also allow for some variation in the recall of items within the pair. This variation of the model, however, would greatly complicate matters, so it was not explored here.

APPENDIX A

Statistical Analyses for Experiment 1

Recognition

Hit rates and false-alarm rates were calculated for each subject for each condition. Nonlist words presented on the left (X) and right (Y) were considered separately. Recognition d' was calculated for each condition. (Note—Recognition d' is undefined when the hit rate or false-alarm rate is equal to 1 or zero. To calculate d' , these hit and false-alarm rates were estimated using the procedures described by Murdock & Ogilvie, 1968). Recognition d' was then submitted to a 2×2 ANOVA, which showed no main effect of test position [$F(1,36) = 1.25, MS_e = 0.826, p > .27$], but an overall generation advantage [$F(1,36) = 69.05, MS_e = 0.436, p < .0001$]. The interaction was also statistically reliable [$F(1,36) = 4.75, MS_e = 0.534, p < .04$]. Separate analyses showed that the generation facilitation was reli-

able for both stimulus terms [$t(36) = 3.63, p < .001$] and response terms [$t(36) = 7.99, p < .0001$].

Cued Recall

A 2×2 ANOVA showed better recall for both stimuli ($A|B$) and responses ($B|A$) when the responses were generated at study [$F(1,36) = 23.66, MS_e = .069, p < 0.0001$]. There was no difference in recall of response and stimulus terms ($A|B$ vs. $B|A$) [$F(1,36) < 1$] and no interaction [$F(1,36) = 1.58, MS_e = 0.053, p > .25$]. Separate analyses showed that the generation facilitation was reliable for both stimulus terms [$t(36) = 2.52, p < .02$] and response terms [$t(36) = 5.25, p < .0001$].

Free recall

Recall proportions were calculated by items without respect to pairs (item recall) and by pairs (pair recall). For item recall, the proportion of items correctly recalled was calculated for stimuli and responses for generate and read pairs and submitted to a 2×2 ANOVA, which showed that, overall, responses were recalled more often than stimulus terms [$F(1,33) = 11.88, MS_e = 0.273, p < .005$], and words from generate pairs were recalled more often than words from read pairs [$F(1,33) = 42.6, MS_e = 4.70, p < .0005$]. The interaction was also statistically significant [$F(1,33) = 9.13, MS_e = 0.291, p < .005$], indicating that generation effects were larger for responses than for stimulus terms. Separate analyses showed that the generation facilitation was reliable for both stimulus [$t(33) = 5.65, p < .0001$] and response terms [$t(33) = 7.01, p < .0001$].

For pair recall, the number of pairs for which both words were recalled was calculated for generate and read pairs. Pair recall probabilities were significantly higher for generate pairs (.20) than for read pairs (.06) [$t(33) = 6.06, p < .001$]. Pair recall performance was close to item recall performance, indicating that pairs were generally recalled as pairs.

APPENDIX B

Statistical Analyses for Experiment 2

Recognition

Hit rates and false-alarm rates were calculated for stimulus (A or X) and response (B or Y) terms for generate and read subjects. These hit rate and false-alarm rate scores were used to compute a d' score for each subject. A 2×2 ANOVA performed on the d' scores showed better recognition for generate than for read subjects [$F(1,46) = 5.02, MS_e = 1.074, p < .03$]. There was no overall difference in performance for stimulus and response terms ($F < 1$). However, the interaction of these factors was reliable [$F(1,46) = 5.38, MS_e = 0.361, p < .03$]. Separate analyses showed that generation superiority was reliable for response terms [$t(94) = 4.39, p < .001$], but the .19 generation advantage was not reliable for stimulus terms [$t(94) = 1.09, p > .10$].

Cued Recall

Proportion of correct recalls was computed for stimulus [recall of A given B as a cue, $rcf(A|B)$] and response [recall of B given A as a cue, $rcf(B|A)$] terms for generate and read subjects and was submitted to a 2×2 ANOVA. Generate subjects showed better performance than read subjects [$F(1,46) = 12.60, MS_e = 0.058, p < .001$]. There was no effect of stimulus versus response terms ($F < 1$) and no interaction [$F(1,46) = 1.61, MS_e = 0.023, p > .20$]. Separate analyses showed that the generation facilitation was significant for recall of stimuli [$t(94) = 3.50, p < .001$] and responses [$t(94) = 4.94, p < .001$].

Free Recall

Proportion of words recalled was computed for stimulus and response terms for generate and read subjects. The ANOVA showed a small, statistically unreliable advantage for generate subjects over read subjects [$F(1,46) = 3.36, MS_e = 0.021, p < .08$]. Response terms were recalled better than stimulus terms [$F(1,46) = 7.09, MS_e = 0.001, p < .02$]. The interaction of encoding condition with pair position (stimulus or response) was also significant [$F(1,46) = 9.55, MS_e = 0.001, p < .005$]. Sep-

arate analyses showed that generation facilitation was reliable for recall of responses [$t(94) = 3.29, p < .01$], but not for stimulus terms [$t(94) = 1.728, p > .10$]. Pair recall was only slightly better for generate pairs than for read pairs, and the .05 difference was not significant [$t(46) = 1.612, p > .05$].

(Manuscript received May 14, 1993;
revision accepted for publication February 14, 1994.)