A multinomial processing tree model for degradation and redintegration in immediate recall

RICHARD SCHWEICKERT
Purdue University, West Lafayette, Indiana

When items are presented for immediate recall, a verbal trace is formed and degrades quickly, becoming useless after about 2 sec. The span for items such as digits equals the number of items that can be pronounced in the available time. The length of the items affects span by affecting pronunciation rate. Other properties, such as phonological similarity and lexicality, can affect span without affecting pronunciation rate. These properties change the trace's useful lifetime by affecting redintegration. An analogy is drawn between trace reconstruction and repair of errors in speech. When a trace is degraded, one process attempts to form a phoneme string, and another process attempts to form a word. The two processes are autonomous and can be selectively influenced by lexicality and phonological similarity. The resulting processing tree models make simple predictions that depend on whether or not the influenced processes are sequential. The results are illustrated with data from experiments by Besner and Davelaar (1982).

It has been known since at least the time of Oliver Wendell Holmes (1871, p. 101) that a person cannot repeat verbatim a list of unrelated items immediately after presentation, unless the list is short. To this day, we do not know what determines how long the list can be. The number of items in the list is not limited per se, because memory span differs for different materials such as shape names and color names (Brener, 1940). Furthermore, the span for a single material (e.g., digits) can differ when the same individual uses different languages (e.g., Ellis & Hennelley, 1980). These facts challenge Miller's (1956) idea that the span is limited to a certain number of chunks or meaningful units (according to this hypothesis, why should the number of chunks in a list of digits change when the language is changed?).

A leading hypothesis is that short-term memory is not limited to a number of items or chunks but is limited by the time for which a trace endures (Baddeley, Thomson, & Buchanan, 1975; Brown, 1958; Mackworth, 1963; Muter, 1980; Peterson & Peterson, 1959; Schweickert & Boruff, 1986). According to this hypothesis, when items are presented, a trace is formed and begins to degrade, perhaps by decaying over time or by becoming noisy because of interference. The useful lifetime of the trace will depend on what it is used for. Recall will be correct if it is completed before the trace degrades to an unusable state. The trace can be refreshed by rehearsing the material, provided the time required to pronounce the items internally is less than the trace degradation time.

Memory span is often defined as the number of items in a list that, half of the time, can be recalled in order immediately after presentation. Then, the time to pronounce a list of span length equals the time at which the trace has become unusable. Ordinarily, this time is about 2 sec.

A particular version of this hypothesis is given in the model of Schweickert and Boruff (1986). It accounts for differences in memory span for different materials in terms of a race between recall and degradation of the trace. According to the model, a list will be recalled correctly if the time T_r required to emit the list is less than the duration T_v of the verbal trace. Then, the probability of correct recall is $P = P[T_r \leq T_v]$.

The model leads to an equation originally proposed by Baddeley et al. (1975) relating memory span, s, and pronunciation rate, r,

$$s = rt + c, \qquad (1)$$

in which t is the time required to pronounce a list of span length. The intercept c is sometimes interpreted as the number of items recalled with the assistance of long-term knowledge (e.g., Hulme, Maughan, & Brown, 1991), although Schweickert, Guentert, and Hersberger (1990) argue that nonzero intercept values may be caused by an artifact.

Because one way to define the pronunciation rate r is with this equation, the fact that it holds is not surprising, nor is the fact that word length changes the pronunciation rate. What is surprising is that under a wide variety

168

of circumstances, the time t is about the same, approximately 2 sec, regardless of word length (Baddeley et al., 1975; Hulme, Thomson, Muir, & Lawrence, 1984; Mackworth, 1963; Schweickert & Boruff, 1986). The time t can be interpreted as the mean trace duration. The invariance of this time over materials initially suggested that the trace duration is relatively constant and that the limit on memory span could be explained in terms of a limit on the time available for rehearsal or recall. This hypothesis is contradicted by recent experiments, described in the next section, in which the time t was malleable. Reconsideration of Equation 1 shows that a variable affecting memory span can do so by affecting the pronunciation rate r, the trace duration t, the intercept c, or some combination of these. If variables affecting memory span always affected all of the parameters, the equation would provide no resolving power. But the studies mentioned above have shown that word length affects the pronunciation rate r, without affecting the trace duration t. This leads to the question of whether there are variables capable of affecting memory span in another way, by changing t or c without changing r.

Phonological Similarity

Hulme and Tordoff (1989) and, independently, Schweickert et al. (1990) carried out experiments to investigate the locus of the effect of phonological similarity in serial ordered recall. The experiments differed in several respects, but each supported the conclusion that phonological similarity can affect memory span without changing speaking rate. That is, phonologically similar items have a smaller memory span because they have a shorter useful trace duration.

Lexicality

Memory span is typically greater for words than for nonwords. In the experiment by Schweickert and Boruff (1986), the nonwords were pronounced more slowly than the words, suggesting that the effect of lexicality could simply be a by-product of an effect on pronunciation rate. However, Hulme et al. (1991) found that when words and nonwords were selected so that speech rates were equated, memory span for the words was greater than memory span for the nonwords. They concluded that, for their data, lexicality affected the intercept c in Equation 1, but did not affect r.

Support from an independent experiment was reported by Schweickert, Ansel, and McDaniel (1991). The words and nonwords used as stimuli were closely matched in several ways. Each list of items was presented visually and rehearsed aloud once; during this rehearsal, words and nonwords were spoken at the same rate. Nonetheless, the memory span was longer for words than for nonwords, supporting the hypothesis that the memory trace had a longer useful lifetime for words than for nonwords. The conditions used in the experiment did not provide an indication of whether lexicality affects either the time t or the intercept c in Equation 1.

The results of the experiments just described were interpreted by the investigators to mean that the trace duration can be extended by the reconstruction of partially degraded traces. Reconstruction is easier if the items are words and if they are phonologically dissimilar (Sperling & Speelman, 1970).

Cowan (in press) gives a detailed treatment of the role of trace reconstruction. He proposed that trace reactivation takes place in the pauses between items during spoken recall and that decay of the traces of items takes place during actual speaking. The model in the following section can be thought of as sketching in a few details about what goes on during the reconstruction postulated in Cowan's model. Since the following model is not necessarily tied to the model of Cowan, I use the terms *degradation* and *redintegration* in lieu of his terms *decay* and *reactivation*.

A Model for Trace Degradation and Redintegration

The challenge in modeling short-term memory trace redintegration is that so little is known about the nature of degraded traces that it is premature to specify details about how missing or noisy features are reconstituted. Some knowledge about degraded traces is available; for example, Cowan and Morse (1986) provide a detailed analysis of vowel changes in short-term memory. But until more is known, modeling with a moderate level of detail seems most likely to succeed, and multinomial processing trees are useful structures for this purpose (Batchelder & Riefer, 1986, 1990; Chechile, 1987; Chechile & Meyer, 1976; Hu & Batchelder, 1992; Riefer & Batchelder, 1988). In a multinomial processing tree (e.g., Figure 1), an attempt to recall an item is carried out with a sequence of cognitive processes represented by branches on a path from the root of the tree to one of the terminal nodes. The outcome of the recall attempt depends on which terminal node is reached.

The model in Figure 1 represents the act of immediate recall of a single item in a randomly ordered list. A complete account of immediate recall would model aspects such as serial position effects and memory for the order of the items (see, e.g., Lee & Estes, 1977; Nairne, 1991; Shiffrin & Cook, 1978); a complete account would also
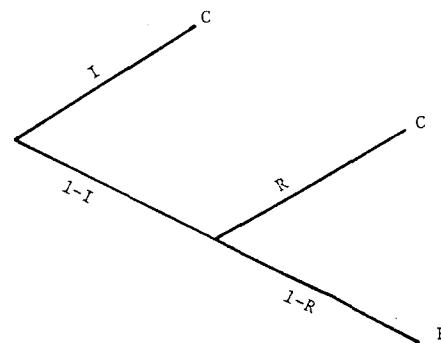

Figure 1. Recall is correct (C) if the trace is intact or reconstructed; otherwise, there is an error (E).

model recall of sentences. These are important but are beyond the scope of this paper.

The model in Figure 1 illustrates a commonly held idea, articulated by Estes (1991), that subjects carry out immediate recall by first attempting direct readout and then guessing if that fails. With probability I, the trace for the item is intact, and the item is recalled unhesitatingly and correctly. With probability $1 - I$, the trace is so degraded as to be ambiguous. In that case, readout cannot be completed, and other slower and more error-prone processes attempt to reconstruct the item on the basis of the remaining information. Reconstruction explains the finding of Standing, Bond, Smith, and Isely (1980) that the smaller the alphabet from which the items are selected, the larger the memory span.

Suppose the reconstruction processes, whatever they may be, are correct with probability R. Then the probability of correctly recalling an item is

$$PC = I + (1 - I)R. \tag{2}$$

An equation of the same form is the basis of Waugh and Norman's (1965) model. The parameter I has essentially the same meaning here as in their model. However, in their model, R would represent the probability the item is recalled from long-term storage. Here, R represents the probability that a degraded trace for the item is correctly reconstructed; this is a long-term memory contribution, but only in the sense that it is done through knowledge of the language. A processing tree model leading to Equation 2 was proposed by Link (1982) for a more general purpose—namely, correcting response times and response choices for guessing.

One approach to modeling with multinomial trees is to estimate the parameters and to test hypotheses about how they differ from group to group in an experimental design. Another approach is to use factorial experiments to analyze the effects of manipulating experimental factors that selectively influence the branch probabilities (Batchelder & Riefer, 1986). This approach is analogous to the analysis of response times with Sternberg's (1969) additive factor method.

Sternberg (1969) proposed that processing is carried out in a series of stages, and if each of two factors selectively influences a different stage, then the combined effect on response time of influencing the two stages would be the sum of the individual effects. Techniques based on selective influence have been very useful in analyzing response times, leading some to propose that they would also be useful for analyzing the percentages of correct responses (Batchelder & Riefer, 1986; Ollman, 1982; Schweickert, 1985).

Although additive effects are expected in reaction time experiments when factors selectively influence different serial stages, Batchelder and Riefer (1986) warned that investigators should not expect additive effects for factors selectively influencing processes in a processing tree. They provide an example in which two such factors produce a crossover interaction, rather than additivity, for percent correct recall.

Although the effects of factors selectively influencing processes in a processing tree may not be simply additive, the effects are systematic, and they are informative about the structure of the processing tree. The tree in Figure 1 predicts that a factor influencing probability I and another factor influencing probability R would have interactive effects on the probability of a correct response. If I is increased by ΔI and R is increased by ΔR, the interaction is predicted to equal their negative product, $-\Delta I \times \Delta R$ (see the Appendix). Because the combined effect of increasing both probabilities is less than the sum of the individual effects, the interaction is said to be underadditive.

The predicted interaction might be small. For example, if I were increased by .10 and R were increased by .20, the interaction of $-.10 \times .20 = -.02$ would be hard to detect. Fortunately, the interactions are not necessarily small. For example, if I were increased by .4 and R were increased by .5, the predicted interaction of $-.20$ would easily be detected. Interactions are important for testing multinomial processing tree models, and predictions based on selective influence are derived in the Appendix for experiments in which one or both of the factors have more than one level.

Applying the Model

Word length and lexicality are factors likely to selectively influence processes in the processing tree in Figure 1. When the moment arrives for recalling a particular item, the trace for that item is more likely to be degraded to the point of ambiguity if the other items in the list are long. This does not mean that the quality of the trace at a particular time, say 1 sec after an item is presented, depends on whether the items are short or long. Rather, the time at which a particular item is about to be recalled depends on whether the other items are short or long. With this reasoning, word length affects the probability $(1 - I)$ of trace degradation. The experiments reported by Hulme et al. (1991) and by Schweickert et al. (1991) were interpreted to mean that lexicality affects the probability of trace reconstruction, R. According to Equation 2, as noted above, these factors should have an underadditive interaction on the probability of correct recall.

From the data available at this time, I am aware of an experiment with relevant data, although the results are, unfortunately, not conclusive. Besner and Davelaar (1982) manipulated the two factors of word length and lexicality in an experiment on immediate ordered recall. The design of the experiment would be ideal for testing the model in Equation 2 except that, for the experimenter's purposes, no more than two levels of any factor were needed, and since the interaction was not of particular interest, the sample size does not provide a powerful test of it. Nonetheless, the experiment illustrates the technique and provides useful, if provisional, conclusions.

Items were either one or three syllables in length. Lexical items were pseudohomophones such as "churtch" and "familly." Nonlexical items were pronounceable controls such as "chitch" and "ramitty." Lists were presented visually in blocks of items with the same number

Table 1
Experiment 2: Percent Correct Recall
in Data of Besner and Davelaar (1982)

	Not Lexical	Lexical
3 syllable items	.578	.824
1 syllable items	.729	.901

of syllables and the same lexical status. Immediately after each list was presented, the subject attempted to recall the items in order. Table 1 gives the probability of recalling an item in its correct serial position.

Each factor had a significant main effect. The interaction was not significant, but, as predicted, the combined effect of lexicality and word length was numerically less than the sum of their individual effects. The simple main effect of lexicality is $.824 - .578 = .246$; the simple main effect of word length was $.729 - .578 = .151$. Their combined effect was $.901 - .578 = .323$, which is .074 less than the sum of their individual effects. The value of this interaction contrast, $-.074$, is close in magnitude to the value predicted by the product of the simple main effects of each factor, $-.037 = -(.824 - .578) \times (.729 - .578)$. The experiment demonstrates that testing the model is feasible, and the result is promising.

Processes for Redintegration

It is worth specifying trace reconstruction in more detail. Suggestions are readily available from models of a closely related activity—namely, monitoring speech and correcting it for errors. According to one prominent theory (Levelt, 1983), when a speaker formulates an intended message, the message passes through working memory, where it is phenomenologically available to the speaker as a string of phonemes in inner speech. The speaker monitors the inner speech for errors in about the same way as he or she would monitor someone else's overt speech. When trouble is detected, a correction is attempted on the basis of all the information the speaker has centrally available. The ability to carry out a correction is limited because of the limited capacity of working memory.

From this point of view, many immediate recall errors are analogous to, if not synonymous with, speech errors. Mechanisms for repairing speech are available for reconstructing short-term memory traces, and vice versa. The major difference is that in immediate recall, the intended message is a list of recently presented items that have already been formulated into language, whereas during speech production, some parts of the intended message have been formulated into language while other parts have not. Consequently, portions of the theory of speech repairs are concerned with such matters as whether or not a paraphrase retains the speaker's intentions; these portions are not too important for immediate recall of items in a randomly prescribed order.

Errors in speech can occur at many levels, including phonetic, lexical, and semantic levels. Models of speech commonly propose processes responsible for each level. The models can be divided into two broad classes, depending on whether the processes proceed autonomously

or communicate continually (see Martin, Weisberg, & Saffran, 1989, for a brief review). Autonomous models for speech production include those of Garrett (1975) and Levelt (1983). Nonautonomous models include those of Dell (1986) and Stemberger (1985), the TRACE model (McClelland & Elman, 1986), and the Boltzmann machine (Hinton & Sejnowski, 1986).

The issue of autonomy has not been resolved empirically in experiments on overt speech, so it is unlikely to be resolved in immediate recall tasks, which inevitably involve some covert speech. But various secondary considerations favor the study of autonomous models. Autonomous models are simpler, because there is less communication. Furthermore, it is unlikely that an experimenter would be able to selectively influence a single level in a nonautonomous model. But, by definition, an autonomous process can be selectively influenced (see, e.g., Cutler, Mehler, Norris, & Segui, 1987, p. 146). Therefore, tests of nonautonomous models are not as straightforward as are tests of autonomous models, thus the autonomous models can be more easily disposed of.

Within the class of autonomous models, there is fairly good agreement that separate processes operate at, roughly, the phonetic, lexical, semantic, and syntactic levels. The agreement is not so good about how the processes are arranged, but if we restrict our consideration to only the phonetic and lexical levels, those most relevant for ordered recall of lists of items, only a few possibilities have been proposed.

The autonomous processes are serial in the models of Garrett (1975) and Levelt (1983); serial processes have also been proposed for the task of phoneme monitoring (see, e.g., Fodor, Bever, & Garrett, 1974). Parallel autonomous processes were proposed for phoneme monitoring by Cutler and Norris (1979). In a model of speech production by Martin et al. (1989, Figure 1), pairs of autonomous processes are in series, and the pairs themselves are executed in parallel. The authors call this arrangement serial, but it has both serial and parallel components.

Because the arrangements are stated so explicitly, the autonomous models are testable by manipulating factors that selectively influence the processes. Processes analogous to, if not the same as, those proposed for repair of speech would be useful for the job of reconstructing a degraded memory trace. As before, if the trace for an item is not degraded, the item is recalled directly from the trace. Otherwise, the item is reconstructed from a degraded trace through two processes. One process attempts to transform the degraded representation into a word or other higher level entity. Such a process is postulated to account for the result, discussed in the introduction, that memory span for words is greater than that for nonwords and that the difference cannot be explained in terms of a difference in the length of the items. The other process attempts to transform the degraded representation into a string of phonemes. At first it may seem superfluous for the system to have a phonemic process in addition to a lexical process. But, to the contrary, some of the most important practical uses for short-term verbal
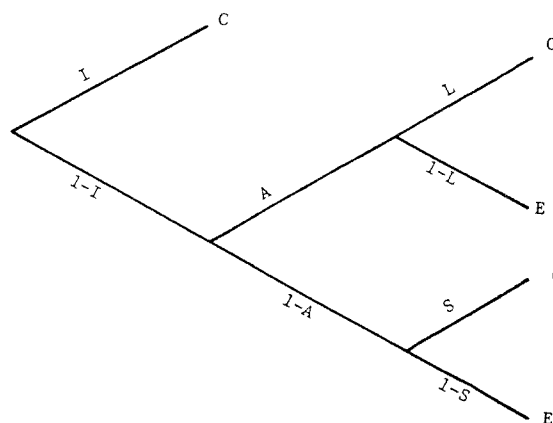
Figure 2. With probability I, the trace is intact and recall is correct. If the trace is too degraded for direct readout, a phoneme string is reconstructed with probability S, and a higher level entity is reconstructed with probability L.
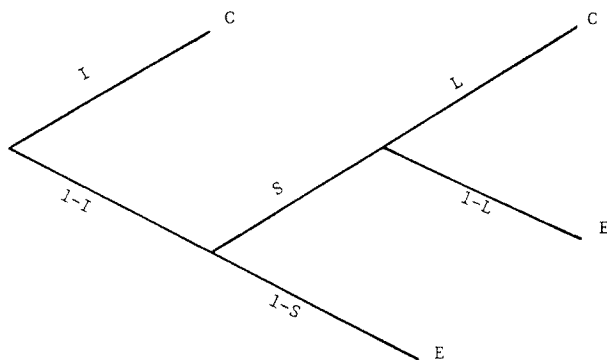

Figure 3. A processing tree in which the phonological process, S, and the lexical process, L, are sequential.

memory pertain to nonwords, such as learning a vocabulary in a foreign language and children learning new words (Baddeley, 1992).

If the material consists of words arranged in meaningful phrases or sentences, then reconstruction may be aided by processes concerned with syntax, semantics, and relations provided by context. These higher level processes may provide enough constraints on the possibilities to change or eliminate lower level effects of phonological similarity, mispronounced words as stimuli, and so on. These processes are beyond the scope of this paper, although of course they are important for the immediate recall of sentences. Readers interested in that topic are referred to the work of Potter and Lombardi (1990).

An arrangement in which autonomous phonemic and lexical processes are not sequential is represented as a tree in Figure 2. In the figure, with probability A, the lexical process determines the response outcome, and with probability $1-A$, the phonemic process determines the outcome. A sequential arrangement of the two processes is represented in the tree in Figure 3.

What do these models predict? Suppose one experimental factor changes the probability L that the lexical pro-

cess produces the correct item, whereas another experimental factor changes the probability S that the phonemic process produces the correct item. In Figure 3, at least one path through the tree contains branches labeled with L and other branches labeled with S. For factors changing these parameter values, an interaction is predicted (see the Appendix). In Figure 2, in which the processes are not sequential, there is no path through the tree containing branches labeled with both an L and an S. This model predicts additive effects of the two factors on percent correct recall (see the Appendix).

An earlier example of a prediction of additivity is in the model of Martin et al. (1989) for the combined effects of semantic and phonological similarity. As noted earlier, these authors call their arrangement of processes "serial," resulting in their assertion that selective influence of serial autonomous processes leads to additive effects. In the terminology used here, the influenced processes in their model (their Figure 1) would not be called serial (or sequential). The terminological difference is not important, but, unfortunately, Martin et al. go on to say that all autonomous models predict additivity, and they interpret all interactions as evidence for nonautonomous models. This conclusion is too strong, because Batchelder and Riefer (1986) have shown that a crossover interaction can result from factors selectively influencing two processes in a processing tree. More details are in the Appendix.

Two factors suitable for investigating the models in Figures 2 and 3 are phonological similarity and lexicality, and these were manipulated in another experiment of Besner and Davelaar (1982). As before, only two levels of each factor were used, so only a limited test of the models can be made. All stimuli were visually displayed nonwords. Pseudohomophones (e.g., "phood") were items that would be pronounced as words and are labeled lexical here. Control items (e.g., "thude") would not be pronounced as words. The control items would be lexical if one or two letters were altered; consequently, the probability of reconstructing an item through lexical knowledge may not be 0, even for the control items. Lists of items with the same level of lexicality and phonological similarity were presented together in blocks. Examples of control items in lists with high phonological similarity are "thude" and "zewd"; examples of control items with low phonological similarity are "thawl" and "zede." Immediately after each list was presented, each subject attempted to recall all the items in order. The probability of correctly recalling each item in its correct serial position was calculated.

It may seem at first that items with high phonological similarity would have a good chance of being reconstructed from a partially degraded trace. If all the items rhyme with *food*, and the initial phoneme for "thude" is available, the entire item, "thude," can be reconstructed. But the items must be recalled in order. If the crucial initial phoneme is not available, there is a very low chance of guessing the item from the remainder. A straightforward calculation shows that this low chance of

Table 2
Experiment 1: Percent Correct Recall
in Data of Besner & Davelaar (1982)

	Not Lexical	Lexical
Phonologically similar	.602	.750
Phonologically dissimilar	.707	.896

guessing when the initial phoneme is lost far outweighs the improved chance of guessing when the later phonemes are lost. Note also that if the initial phoneme of a control item such as "zede" is degraded, very little lexical knowledge would be of use in reconstruction, but detailed knowledge of the features of phonemes in the context of other phonemes would be useful.

The data of Besner and Davelaar (1982) are listed in Table 2. The main effect of each factor was significant, but the interaction was not. This pattern is predicted by the model in Figure 2, in which the lexical and phonological processes are not sequential. The alternative sequential arrangement in Figure 3 predicts an interaction (see Appendix). Note again that an experiment specifically designed to test the tree models would have more factor levels and greater power.

Besner and Davelaar (1982) also put forth an interesting hypothesis. They proposed that the phonological code used to maintain information in short-term memory is different from the code used for lexical access from printed information. Their hypothesis was supported by a dissociation of the effects of articulatory suppression. Those results will not be discussed here, because articulatory suppression seems to produce wide-ranging effects rather than selective influence.

In conclusion, experiments on memory span suggest that immediate recall can be increased by an item or two if degraded traces are easier to reconstruct. Reconstruction can be explained in terms of the autonomous processes postulated in some models for tasks such as the repair of utterances and phoneme monitoring. These processes include one influenced by lexicality and another one influenced by phonological similarity. Multinomial processing trees provide a natural way to represent the outcomes of the processing. The structure of the tree can be found if there are factors that selectively influence the probabilities associated with the branches in the tree.

REFERENCES

BADDELEY, A. D. (1992). Working memory. *Science*, **255**, 556-559.

BADDELEY, A. D., THOMSON, N., & BUCHANAN, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **14**, 575-589.

BATCHELDER, W. H., & RIEFER, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical & Statistical Psychology*, **39**, 129-149.

BATCHELDER, W. H., & RIEFER, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, **97**, 548-564.

BESNER, D., & DAVELAAR, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology*, **36**, 701-711.

BRENER, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, **26**, 467-482.

BROWN, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10**, 12-21.

CHECHILE, R. (1987). Trace susceptibility theory. *Journal of Experimental Psychology: General*, **110**, 203-222.

CHECHILE, R., & MEYER, D. L. (1976). A Baysean procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology*, **13**, 269-295.

COWAN, N. (in press). Verbal memory span and the timing of spoken recall. *Journal of Memory & Language*.

COWAN, N., & MORSE, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, **79**, 500-507.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.

CUTLER, A., & NORRIS, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

DELL, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, **93**, 283-321.

ELLIS, N. C., & HENNELLEY, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, **71**, 43-52.

ESTES, W. K. (1991). On types of item coding and source of recall in short-term memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 155-174). Hillsdale, NJ: Erlbaum.

FODOR, J. A., BEVER, T. G., & GARRETT, M. F. (1974). *The psychology of language*. New York: McGraw-Hill.

GARRETT, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133-177). New York: Academic Press.

HINTON, G. E., & SEJNOWSKI, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 282-317). Cambridge, MA: MIT Press.

HOLMES, O. W. (1871). *Mechanism in thought and morals*. Boston: Osgood.

HU, X., & BATCHELDER, W. H. (1992). *The statistical analysis of general processing tree models with the EM algorithm* (Tech. Rep. No. MBS 92-06). Irvine: University of California, Irvine Research Unit in Mathematical Behavioral Sciences.

HULME, C., MAUGHAN, S., & BROWN, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory & Language*, **30**, 685-701.

HULME, C., THOMSON, N., MUIR, C., & LAWRENCE, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, **38**, 241-253.

HULME, C., & TORDOFF, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, **47**, 72-87.

LEE, C. L., & ESTES, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning & Verbal Behavior*, **16**, 395-418.

LEVELT, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, **14**, 41-104.

LINK, S. W. (1982). Correcting response measures for guessing and partial information. *Psychological Bulletin*, **92**, 469-486.

MACKWORTH, J. F. (1963). The duration of the visual image. *Canadian Journal of Psychology*, **17**, 62-81.

MARTIN, N., WEISBERG, R. W., & SAFFRAN, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory & Language*, **28**, 462-485.

McCLELLAND, J. L., & ELMAN, J. L. (1986). Interactive processes in speech perception: The TRACE model. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), *Parallel distributed processing: Exploration in the microstructure of cognition, Vol. 2:*

Psychological and biological models (pp. 58-121). Cambridge, MA: MIT Press.

MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.

MUTER, P. (1980). Very rapid forgetting. *Memory & Cognition*, **8**, 174-179.

NAIRNE, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition*, **19**, 332-340.

OLLMAN, R. (1982, August). *Additive factors and the speed-accuracy tradeoff*. Paper presented at the meeting of the Society for Mathematical Psychology, Princeton, NJ.

PETERSON, L. R., & PETERSON, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, **58**, 193-198.

POTTER, M. C., & LOMBARDI, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory & Language*, **29**, 633-654.

RIEFER, D. M., & BATCHELDER, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, **95**, 318-339.

SCHWEICKERT, R. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin*, **97**, 530-546.

SCHWEICKERT, R., ANSEL, B., & McDANIEL, M. A. (1991, November). *Lexicality and generation effects on memory span*. Paper presented at the 32nd Annual Meeting of the Psychonomic Society, San Francisco.

SCHWEICKERT, R., & BORUFF, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 419-425.

SCHWEICKERT, R., GUENTERT, L., & HERSBERGER, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, **1**, 74-77.

SHIFFRIN, R. M., & COOK, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning & Verbal Behavior*, **17**, 189-218.

SPERLING, G., & SPEELMAN, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 152-202). New York: Academic Press.

STANDING, L., BOND, B., SMITH, P., & ISELY, C. (1980). Is the immediate memory span determined by subvocalization rate? *British Journal of Psychology*, **71**, 525-539.

STEMBERGER, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language*. (Vol. 1). London: Erlbaum.

STERNBERG, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, **57**, 421-457.

WAUGH, N. C., & NORMAN, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89-104.

APPENDIX
Factors Selectively Influencing Processes in Multinomial Processing Trees

Each node in a multinomial processing tree represents a process used in attempts to recall an item. Each process has a finite number of possible outcomes; for instance, one outcome of a process might be that a remnant of a memory trace was successfully identified as a word, whereas another outcome might be the failure to do this. Each outcome is represented by a branch starting at the node representing the process. It is assumed here that two branches start at each node; in principle, there could be more. Each terminal node of the tree represents the result of a recall attempt. For our purposes, each terminal node corresponds to either the correct or incorrect recall of an item. The attempt to recall a particular item on a particular trial is represented by a path from the root of the tree to one of the terminal nodes. When a node is reached in the recall attempt, a certain

process, say process a, is executed; it is successful with probability p_a, and fails with probability $1-p_a$. The branch corresponding to the outcome of the process, success or failure, is included in the path, and a new node is reached. The probability that a path is followed is the product of the probabilities associated with the branches on the path. The probability of correct recall of the item is the sum of the probabilities of the paths ending at terminal nodes associated with correct recall.

A process may be executed as the result of more than one chain of events. That is, for a given parameter p, there may be more than one node whose outgoing branches have the probabilities p and $1-p$. Suppose this does not happen in such a way that some path from the root of the tree to a terminal node encounters the same parameter more than once; that is, the product of probabilities on the branches of a path through the tree has no parameter with a power greater than 1.

The processing trees considered here have binomial nodes—that is, a node a is the starting node for exactly two branches, and the parameter p_a appears only on these two branches, once as p_a and once in the expression $1-p_a$. An experimental factor is said to selectively influence process a if changing the level of the factor changes the value p_a and of $1-p_a$ but does not change the probability associated with any other branch.

The pair of branches starting at a node contributes nothing to the probability calculation if the two branches end in terminal nodes of the tree with the same outcome, o. Such pairs of branches can be removed, so their starting node becomes a terminal node of the tree labeled with the outcome o. This procedure can be repeated until in the tree remaining, each parameter occurs on at least one path associated with correct recall and on at least one path associated with incorrect recall. Changing the level of a factor selectively influencing the probability of a branch will then produce both a change in the probability of correct recall and a change of equal magnitude but opposite sign in the probability of incorrect recall.

Suppose the levels of two factors are changed. In a factorial experiment, if the combined effect of changing the levels of both factors is greater than the sum of the effects of changing them separately, the interaction is said to be overadditive. Additivity and underadditivity are defined analogously.

Given two parameters p and q, it follows from the assumptions that the probability of some paths through the tree may involve only the parameter p, some may involve only q, some may involve both p and q, and some may involve neither. Then the probability of a correct recall can be written as

$$PC = Ap + B(1 - p) + Cq + D(1 - q) + Epq + F(1 - p)q$$
$$+ Gp(1 - q) + H(1 - p)(1 - q) + I, \qquad (A1)$$

in which none of the expressions A, B, . . ., I depend on p or q.

The combined effects of selectively influencing the processing at two different nodes a and b depends on how the nodes are arranged in the processing tree. Consider, for example, the effects of changing the parameters A and L in Figure 2. The probabilities of correct recall are given in the top part of Table A1 for three values of A and four values of L. The probability of correct recall increases monotonically with L at each level of A, because a branch on a path leading to correct recall is labeled with L. The effects of changing A are more complicated. For low values of L, the probability decreases with A, whereas for high values of L, the probability increases. The change in sign is produced because one path to correct recall in Figure 2 goes through a branch labeled A, whereas another path goes through a path labeled $1 - A$.

Table A1
Effects of Changing Parameters A and L in the Tree in Figure 3

A	L			
	.20	.40	.60	.80
	Probability of Correct Recall			
.25	.54	.58	.62	.66
.50	.48	.56	.64	.72
.75	.42	.54	.66	.78
	Interaction Contrasts			
.25	—	—	—	—
.50	—	.04	.08	.12
.75	—	.08	.16	.24

The interaction contrast for cell (i,j), in which factor 1 is at level i and factor 2 is at level j, is $PC_{ij} - PC_{1j} - PC_{i1} + PC_{11}$. For example, the interaction contrast for the cell in row 3 and column 4 is $.78 - .66 - .42 + .54 = .24$. The interaction contrasts are in the lower part of Table A1. They are monotonically increasing in both the rows and columns.

In general, the effects of selectively influencing two processes, a and b, in a processing tree meeting the assumptions leading to Equation A1 can be summarized as follows. The levels of the factors can be ordered so that for a fixed level of one of the factors, percent correct is monotonic with changes in the level of the other factor. As the levels of one factor (e.g., factor 1) increase, the relation between percent correct and the other factor may change, but not more than once, from monotonically increasing to monotonically decreasing. If the interaction contrasts are not all zero, they will all have the same sign and will either monotonically increase or decrease in the rows and columns, with no change in the direction of monotonicity. The interactions can be classified in three ways, depending on how the influenced processes are arranged in the tree: (1) The factors will have additive effects if and only if no path from the root to a terminal node contains a branch with an expression involving p and another branch with an expression involving q; (2) if the factors have overadditive effects, then at least one path from the root to a terminal node includes branches labeled with p_a and p_b, or branches labeled with $1 - p_a$ and $1 - p_b$; and (3) if the factors have underadditive effects, then at least one path from the root to a terminal node includes branches labeled with p_a and $1 - p_b$, or labeled with $1 - p_a$ and p_b.

The statements above can be established through straightforward calculations of the interactions. To calculate the effect on the probability of correct recall of a change Δp in a parame-

ter p, note that paths contributing to the probability of correct recall may include p, or $1 - p$, or neither, and higher powers of p do not occur. Consequently, probability of correct recall as a function of p can be written,

$$PC(p) = Jp + K(1 - p) + L,$$

in which J, K, and L are real numbers. When p is increased by Δp, the corresponding probability is

$$PC(p + \Delta p) = J(p + \Delta p) + K(1 - p - \Delta p) + L,$$

and the increase in probability of correct recall is

$$PC(p + \Delta p) - PC(p) = (J - K)\Delta p.$$

This can be calculated from

$$PC(p+\Delta p) - PC(p) = \frac{\partial PC(p)}{\partial p} \Delta p. \qquad (A2)$$

The interaction contrast for a factor increasing p by Δp and another factor increasing q by Δq can be derived by repeated application of Equation A2.

$$PC(p+\Delta p, q+\Delta q) - PC(p, q+\Delta q) - PC(p+\Delta p, q) + PC(p, q)$$

$$= \frac{\partial PC(p, q+\Delta q)}{\partial p} \Delta p - \frac{\partial PC(p, q)}{\partial p} \Delta p$$

$$= \frac{\partial}{\partial p} [PC(p, q+\Delta q) - PC(p, q)] \Delta p$$

$$= \frac{\partial}{\partial p} \frac{\partial PC(p, q)}{\partial q} \Delta q \Delta p$$

$$= \Delta p \Delta q \frac{\partial^2 PC(p, q)}{\partial p \, \partial q}.$$

The interaction contrast will have the same sign as the second derivative in the equation above. This sign depends only on the arrangement of the nodes in the processing tree and not on the values of the probabilities themselves.

To use the formula above to predict the interaction contrast when I and R are changed in Equation 2, note that the partial derivative of PC with respect to R is $1 - I$ and that the derivative of this with respect to I is -1. Thus, the interaction contrast is predicted to be $-\Delta I \Delta R$.